Guiding principles for natural product drug discovery

Author
Camp, David, Davis, Rohan, A. Evans-Illidge, Elizabeth, Quinn, Ronald

Published
2012

Journal Title
Future Medicinal Chemistry

DOI
https://doi.org/10.4155/fmc.12.55

Copyright Statement
Copyright 2012 Future Science. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. For full bibliographic citation, please refer to the version available at www.future-science.com.

Downloaded from
http://hdl.handle.net/10072/48359
Guiding principles for natural product drug discovery

David Camp,1,* Rohan A. Davis,1 Elizabeth A. Evans-Illidge2 and Ronald J. Quinn1

1 Eskitis Institute, Griffith University, Brisbane, Queensland 4111, Australia
2 Australian Institute of Marine Science, Townsville, Queensland, 4810, Australia
Abstract. Natural products (NPs) have historically been a fertile source of new drugs for the pharmaceutical industry. However, this once popular approach has waned considerably over the past two decades as the high-throughput screening (HTS) of mega-libraries comprised mainly of molecules with non-natural (synthetic) motifs has unfolded. Contemporary HTS libraries contain molecules compliant with physicochemical profiles considered essential for downstream development. Until recently, there was no strategy that aligned NP screening with the same physicochemical profiles. An approach based on Log P has addressed these concerns and, together with advances in isolation, afforded NP leads in timelines compatible with pure compound screening. Concomitant progress related to access of biological resources has provided long-awaited legal certainty to further facilitate NP drug discovery.

Key terms

Convention on Biological Diversity (CBD): an international treaty that addresses environmental and biodiversity conservation as well as associated trade, development and intellectual property rights. It was opened for signature in 1992 and entered into force in 1993. The CBD has three objectives: the conservation of biodiversity, the sustainable use of its components, and the equitable sharing of benefits arising from the use of genetic resources.

Nagoya Protocol: a legally binding protocol to the CBD. It provides further specific and legally enforceable instruction on the practical application and implementation of provisions pertaining to access and benefit sharing and the use of traditional knowledge. The Nagoya Protocol is expected to be ratified in 2012.
**Biodiscovery:** the search for products and processes derived from biodiversity and inherent biochemical systems which can be developed into new applications in a range of industry sectors that includes pharmaceuticals and agrochemicals. The term biodiscovery is preferred to bioprospecting, to avoid connotations of unsustainable exploitation of resources and to better capture the nature of research and discovery.

**Chemistry space:** many different descriptors are used to represent molecules *in silico* [1]. Descriptors, *inter alia*, can be used to describe molecular and physicochemical properties (*e.g.* molecular weight, rotatable bonds, water solubility, *etc.*), or be calculated from a molecule’s topology to describe properties like polar surface area. Chemistry space is a term that is often used in place of “multi-dimensional descriptor space” [2] and, conceptually, is the space occupied by all energetically stable stoichiometric combinations of electrons, atomic nuclei and topologies in discrete molecules [3]. Drug-like chemistry space is defined here by a particular set of descriptors that limits total chemistry space to carbon-based small molecules that share certain characteristics with other molecules that act as drugs. The number of drug-like molecules has been estimated to lie between $10^{18} – 10^{200}$ compounds [1].

**Log P:** defined as the ratio of the non-ionized form of a drug that is distributed between *n*-octanol and water phases at equilibrium. Higher values are indicative of greater lipophilicity.

**PAINS:** literally pan-assay interfering compounds [4,5]; these are frequent hitters or otherwise promiscuous compounds containing substructures that often escape *in silico* filters designed to identify reactive functional groups in small molecules. Because PAINS can also contain protein-reactive substructures, molecules with such moieties typically display a trend of nonspecific biological activity as data from more assays are analyzed.
**Prefractionated library**: a generic term that is typically used to describe any fractionation of crude extracts before primary screening. Prefractionation of crude extracts into fractions containing a few to many components is achieved by column chromatography, high-performance liquid chromatography (HPLC), or liquid–liquid partitioning. The simplest prefractionation protocols can reduce a crude extract to one fraction (e.g. the enrichment of alkaloids in a sample using cation exchange) while more involved procedures have been reported to afford up to 200 fractions per crude extract.

**Mass-directed isolation**: the process of isolating compounds from some type of biota (e.g., marine invertebrates, plants, microbes, etc.) that is based solely on mass spectrometric (MS) data. In relation to biodiscovery, this involves the initial identification of MS ions in a bioactive extract or fraction that is predicted to correspond to the active compound(s). Compounds are isolated following large-scale extraction of biota by chromatography and MS analysis of the fractions. Biological evaluation is then undertaken to confirm activity and potency.

**Bioassay-guided fractionation**: the process of isolating biologically active compounds from some type of biota (e.g. marine invertebrates, plants, microbes, etc.) that is based solely on bioactivity. Biological screening initially identifies a bioactive extract or fraction from a NP-based library. Large-scale extraction of the relevant biota is then undertaken, followed by iterative rounds of chromatography withd biological testing conducted after each fractionation step. Only those fractions showing bioactivity undergo subsequent separation, and this is a process requiring many cycles until the compound(s) responsible for initial screening activity are isolated.

**Dereplication**: the process by which the chemical and biological characteristics of the unknown compounds are compared with the chemical and biological characteristics of known compounds from databases to eliminate those that have been identified
previously. Some NPs are discovered more frequently than others during screening programs. As the number of described NPs increases, so does the probability of rediscovering known compounds.

About 34% of current marketed drugs can trace their origins back to an unmodified natural product (NP) or a semi-synthetic analogue [6]. While this metric reflects the historic role NPs have played in the pharmaceutical industry, it does not capture the reality that this once popular approach has waned considerably since the early 1980s. The decline can be traced back to a paradigm shift that coalesced advances in molecular biology, combinatorial chemistry and high-throughput screening (HTS). It was expected that combinatorial chemistry would deliver massive numbers of novel chemical motifs that would ultimately result in the filing of increasing numbers of new chemical entities (NCEs). At the same time biodiverse collections became harder to legally acquire and use for commercial research due to lack of certainty and clarity over access and benefit sharing (ABS) requirements [7,8].

The synergies between these new technologies, and the changing landscape for biodiscovery, led many pharmaceutical companies to either cut back or disband their NP drug discovery programs and shift resources into the new paradigm. The situation worsened for NP extract screening as the speed of discovery became de rigour, although this later morphed into the “fail fast, fail cheap” mantra, as combinatorial libraries did not live up to the high expectations placed on them.

To maintain the forward momentum of projects performed in industry that utilized both pure compound library and NP extract screening, hits from pure compound libraries were often progressed to lead identification (hit-to-lead) in 6 months while it took 1-2 full-time employees (FTE) in this time frame to complete a reasonable number
of bioassay-guided fractionations (e.g. 20) to arrive at the hit stage. The concomitant prosecution of both pure compound and extract screening highlighted several shortcomings of bioassay-guided fractionation when this *modus operandi* was simply translated to the new paradigm.

First, the screening group had to maintain cells or a quantity of protein for the NP group that delivered fewer and fewer fractions as the active component(s) were gradually purified. This is wasteful in terms of reagents being added to empty wells of microtiter plates and also in terms of time that screening staff could use to focus on a new assay. Second, structural elucidation of the actives was required following an often laborious isolation which could add to the frustration, particularly if the molecule violated most, if not all, lead- and drug-like physicochemical properties, or contained a substructure that was a medicinal chemistry “dead end”. Third, if project reviews are carried out after 6 months, a not unreasonable period following initial screening then, more often than not, the same amount of data is not available for all isolations. Fourth, acquisition of secondary and counter screening data during hit identification, and early pharmacokinetic data for the hit-to-lead phase, potentially means that recollection of the original biota (microbial biota can be re-cultured) may be necessary so that the active component could be reisolated in larger quantities unless the molecule was synthetically tractable (and medicinal chemistry resources were available to attempt a synthesis).

As a result, NP drug discovery was increasingly marginalized and viewed as a last resort by project teams; only to be considered when screening the compound library did not yield any leads. Put simply, classical bioassay-guided fractionation so engrained in NP drug discovery was not competitive with the HTS timelines of pure compound libraries. Consequently, the quest for NP drug discovery over the past 20 years has been increasingly taken up by smaller niche companies offering a point of difference and the
“true believers” from publicly funded research organisations, like the National Cancer Institute (U.S.A.), and various academic groups.

Yet, despite this somewhat melancholy backdrop, a recent review by Newman and Cragg has shown that NPs and their derivatives continue to make a significant contribution to the pharmaceutical industry [6]. Thus, from 1981-1987, NPs contributed substantially to the total number and percentage of all new drug approvals per annum. After this, though, the number of drug approvals based on NPs and their semi-synthetic analogues began to drift downwards. To some extent, however, NPs are a microcosm of the pharmaceutical industry and the decreasing number of approvals each year is merely reflecting a general trend. Fortunately, the value of NPs is captured when analyzed as a percentage of all new drug approvals (notwithstanding the fact this metric resembles the fluctuations in the stock market). Hence, after a somewhat steady period between 1981-1986, a definite decline is observed from 1987-1991 where the percentage of new approvals drops to 20%. This is followed by a short rally over the next 3 years where the average jumps to 40% before a crash occurs in 1995 (29%) which turns into a definite bear market that reaches a nadir of 12.2% in 1997. A bullish, though somewhat erratic, run has since played out reaching a peak in 2010 where NPs and their derivatives accounted for 50% of all new drug approvals. Clearly, NPs are still furnishing leads to the pharmaceutical industry, albeit in lower total numbers than pre-1987. The challenge, as we see it, is to better integrate NPs into contemporary drug discovery so that both the total number and percentage contributions can increase.

**Guiding principles for natural product drug discovery**

To achieve this goal, we believe that there are two broad principles that together can underpin the discovery of NP drugs. Further development/refinement of these
principles over time will not only deliver better leads for the pharmaceutical industry but will also make the screening of NP libraries more competitive and truly complementary to pure compound libraries.

Perhaps the easiest of these guiding principles to implement is to simply learn from the mistakes that initially plagued the combinatorial chemistry paradigm and develop NP workstreams that: 1) address physicochemical profiling prior to screening and; 2) make better use of technology to constantly shorten discovery timelines.

It should be noted that the early period of the combinatorial chemistry paradigm was not without controversy [9]. The focus on quantity and speed unfortunately meant that combinatorial chemistry and HTS was oversold as a panacea for the dearth of NCEs. When NCEs did not emerge from the pipeline, the focus rightly shifted away from throughput and towards quality. The readjustment was driven, in part, by a better understanding of the types of molecules that should be screened and the end point that resulted in a drug [10,11].

Lipinski et al.’s seminal analysis revealed that many combinatorial libraries contained molecules that did not echo specific physicochemical properties common to 90% of orally active drugs that advanced to phase II clinical trials. This disconnect led the authors to summarize their findings via the now well-known “Rule of 5” (Ro5) [12], so-called because the first four parameters are all multiples of five. In essence, the Ro5 is a set of guidelines that articulates the ranges of four key properties assumed to be a good indicator of oral absorption, i.e., MW < 500; calculated Log P (cLog P) < 5; number of H-bond donors (HBD) < 5; and number of H-bond acceptors (HBA) < 10. In this sense, the Ro5 does not define drug-like chemical space per se but is rather a predictor for oral bioavailability [13]. Potential bioavailability issues occur if there are two or more Ro5 violations.
It is politic to remember that the Ro5 also contains an important caveat, the so called 5th rule; namely “compound classes that are substrates for biological transporters are exceptions to the rule” [12]. “Pleading the 5th” has been used by proponents of NPs to justify the inclusion of what many in the pharmaceutical industry may consider to be dubious molecules into screening libraries. While we would support the inclusion of NPs into screening sets, we would also argue that it should be for the right reasons and not because their pedigree automatically affords them some special status under the “5th amendment”. For example, some more structurally complex NPs like paclitaxel (Taxol®) 1 or rapamycin (Rapamune®) 2, most likely take full advantage of their molecular architecture and spatial arrangement of functional groups, rather than an active transport mechanism, to make a mockery of the Ro5 [14,15]. Thus, it is not the fact they are NPs automatically captured by the 5th rule that makes them and other structurally complex NPs like trabectedin (Yondelis®) 3 good drugs existing outside of “Lipinski space”, but some of their other features/properties. The real issue is to understand what these are and adjust a NP drug discovery program accordingly.

Figure 1. Chemical structures for paclitaxel (Taxol®) 1, rapamycin (Rapamune®) 2, and trabectedin (Yondelis®) 3.
In this respect, while Lipinski’s pivotal investigation was important in guiding the process toward the end point for orally available drugs, it was another two years before direction concerning the physicochemical profiles of molecules that comprised the actual screening set (i.e. chemical starting points for drug discovery) was reported by Teague and co-workers [16]. “Lead-like” molecules are smaller and more polar than drug-like molecules and were proposed to address the generally observed trend within the pharmaceutical industry that the initial hit from HTS tends to increase in mass as chemical moieties are introduced to improve selectivity, potency, and bioavailability (or any combination thereof). For this reason it makes sense to start below the drug-like profile so that additional mass and functionality can be added *en route* to the NCE without it potentially suffering from “molecular obesity” [13].

Exactly how far below the drug-like profile a chemical starting point should reside is open to interpretation (Table 1). Oprea *et al.*, for example, defined a lead-like molecule employing a property-based analysis [17]. However, Hann and Oprea also advocated screening libraries of molecules with “reduced complexity” that had even more stringent physicochemical profiles than lead-like molecules [18]. Significantly, although the actual guidelines being implemented may differ, the physicochemical profiling of compound libraries that now occurs in the pharmaceutical industry has become an intrinsic part of the design and selection process.
Table 1. Physicochemical profiles for reduced complexity, lead-like and drug-like molecules.

<table>
<thead>
<tr>
<th>Physicochemical parameter</th>
<th>Reduced complexity</th>
<th>Lead-like</th>
<th>Drug-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>molecular weight</td>
<td>≤ 350</td>
<td>≤ 460</td>
<td>≤ 500</td>
</tr>
<tr>
<td>H-bond donors</td>
<td>≤ 3</td>
<td>≤ 5</td>
<td>≤ 5</td>
</tr>
<tr>
<td>H-bond acceptors</td>
<td>≤ 8</td>
<td>≤ 9</td>
<td>≤ 10</td>
</tr>
<tr>
<td>cLog P</td>
<td>≤ 2.2</td>
<td>-4.6 – 4.2</td>
<td>≤ 5</td>
</tr>
<tr>
<td>rotatable bonds</td>
<td>≤ 6</td>
<td>≤ 10</td>
<td>–</td>
</tr>
<tr>
<td>heavy atoms</td>
<td>≤ 22</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>number of rings</td>
<td>–</td>
<td>≤ 4</td>
<td>–</td>
</tr>
<tr>
<td>Log $S_w^a$</td>
<td>–</td>
<td>≥ -5</td>
<td>–</td>
</tr>
</tbody>
</table>

A short time after the concepts of lead- and drug-likeness had gained traction, Leeson and Davis critically examined a wider range of physicochemical properties to determine which may actually be the most important for drug development [10]. Their hypothesis that the physicochemical parameters that remained consistent over the relatively short history of modern drug development were more likely to also be the most important was simple yet profound. A comparison of the physicochemical properties of 864 drug approvals up to 1982, and 329 between 1983-2002, showed that the median $cLog$ P, per cent polar surface area (% PSA) and the number of HBDs remained the same while other physicochemical properties like MW, HBA, rotatable bonds and the number of rings increased in a statistically significant manner [10]. Of the three most constant physicochemical properties, lipophilicity, as measured by $cLog$ P, was found to vary the least prompting Leeson to rate it as the most important [10].
This was not completely unexpected as the role of Log P in influencing drug potency, pharmacokinetics and toxicity had been established for many years [19-22] and has even prompted Ganesan to dub it the “Lord of the Rules” for drug discovery and development [23].

The second guiding principle is somewhat more ethereal in the sense that its execution is not generally able to be carried out by those involved in the science, but rather by high level policy makers from national governments. However, given the Convention on Biological Diversity (CBD) has fundamentally changed biodiscovery, it is absolutely essential that anyone contemplating NP drug discovery stays abreast of any developments and remains cognisant of emerging national and international legislation that permits ABS of genetic resources.

Clearly, NP drug discovery would not be possible without access to the world’s biodiversity of which greater than 80% of terrestrial biodiversity is estimated to be spread across a mere 17 countries [24]. The distribution of the world’s genetic resources is important as, historically, biodiversity has been approximately inversely proportional to a country’s wealth and scientific capacity to not only undertake biodiscovery, but also commercialize outcomes [25]. A dichotomous paradigm has ensued whereby the developing world became a supplier of biodiversity (and sometimes associated traditional knowledge) that the developed world translated into commercial outcomes [26]. By the late 1980s many of the governments of developing countries and non-government organizations effectively lobbied that the situation was inequitable [27-29]. Interestingly, the global shift from NPs to combinatorial chemistry as a source of molecular diversity for screening programs coincided with the advent of the term “biopiracy” during the 1980s.
The disconnect between the wealth of genetic resources in the majority of megadiverse countries and their ability to translate this into a commercial outcome was central to international negotiations that resulted in a new treaty - the CBD - which was opened for signature at the first Earth Summit in Rio de Janeiro in 1992, and entered into force in 1993. The CBD has three clear objectives: the conservation of biodiversity, the sustainable use of its components, and the equitable sharing of benefits arising from the use of genetic resources [101]. In this respect the CBD is much more than an environmental treaty to address the alarming rate of global biodiversity loss; it is a hybrid that also deals with trade, development, and intellectual property rights [30].

How these two guiding principles have been addressed by research groups from academia and industry to improve the underlying process, and by international law makers in creating the opportunity in the first place, will be discussed in more detail in their respective sections below.

Guiding principle 1: Addressing physicochemical properties in natural product screening and improving the speed that hits, leads and drugs are isolated.

Screening of pure natural product libraries

There have been several approaches to address the deficiencies associated with screening NP extracts in an effort to facilitate downstream isolation of active components. One has entailed isolation and structural elucidation of as many pure NPs as possible [31-33]. This strategy has several obvious advantages as the isolated compounds can be treated the same as any other compound in a library allowing immediate assessment of the compound’s potential thus eliminating the time delay between identification of a hit extract and isolation of the active NP. With a structure in-
hand, it is also possible to exclude compounds from the screening set that do not conform to prescribed physicochemical parameters or have chemical alerts associated with substructures.

Unfortunately, it is simply not yet feasible to isolate every minor component in an extract library due to the underlying logistics, cost and technology deficit. Consequently, the limited throughput associated with purification restricts the ability to build a library representative of all NP chemical space. Ignoring minor components would lead to an inevitable loss of chemical diversity. As a result, approaches based on pure NPs alone are neither comprehensive nor generic, particularly if a biased subset (e.g. alkaloids) is being pursued.

By way of example, it has previously been demonstrated that it is possible to prepare a screening library of highly diverse NPs that are drug-like in their physicochemical parameters [33]. However, this is still a limited subset of all NP diversity and fails to deliver a comprehensive coverage of NP chemical space. A more complete exposure of NPs can only be achieved via a comprehensive sampling of the chemical diversity found in nature through inclusion of minor components found in extracts using HTS.

*Screening crude extract libraries*

Compound diversity is greatest in crude extracts and, as a consequence, their screening addresses the issue of minor components that may not be isolated using the pure NP strategy. Crude extract libraries dovetail nicely with HTS and, indeed, were reported as the progenitors of pure compound screening in Pfizer [34]. We also note that we performed our first HTS on crude extract libraries in a collaboration with Astra in 1994 [35,36]. In addition to their high chemical diversity, crude extracts can be
prepared quickly and inexpensively (like combinatorial chemistry libraries). However, crude extracts also have several drawbacks. It has been reported that the chemical complexity of crude extracts can lead to minor metabolites being undetected [37], particularly if they are masked by other components or if NP pan assay interfering compounds (PAINS) [4,5] are present. Generally speaking, the presence of NP PAINS like large polyphenolic tannins are notorious for interfering with many assay formats and screening technologies [38-40] and can result in a significant number of false positives that may be time consuming to dereplicate. Ideally, NP PAINS would be removed from an extract prior to screening so that detection of potentially potent minor components is facilitated.

Furthermore, the isolation of the active component can be time- and resource-intensive, and can result in chemically unattractive compounds being frequently isolated. Despite an attractive veneer, the HTS of crude extract libraries does not adequately address the issues of isolating lead- or drug-like molecules in a timely manner.

*Screening prefractionated extract libraries*

Numerous prefractionation strategies have been developed to address issues associated with the HTS of crude extracts in an effort to maintain and detect minor components, improve the quality of screening data and reduce timelines associated with downstream bioassay-guided fractionation. Approaches are broadly scoped with reports ranging from the preparation of 4-200 fractions per sample. Table 2 summarizes published strategies for the preparation of prefractionated HTS-ready libraries.
Table 2. Summary of prefractionation protocols for NP screening.

<table>
<thead>
<tr>
<th>Group</th>
<th>Year</th>
<th>Sample origin</th>
<th>Pretreatment strategy</th>
<th>No. pretreated extracts</th>
<th>No. fractions per pretreated extract&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. HPLC fractions per sample</th>
<th>Plate format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequoia Sciences [38]</td>
<td>2002</td>
<td>plant</td>
<td>Silica PAG</td>
<td>6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>40</td>
<td>200</td>
<td>96/384</td>
</tr>
<tr>
<td>bioLeads GmbH [32]</td>
<td>2002</td>
<td>microbial</td>
<td>not specified</td>
<td>2</td>
<td>varied (1-8)</td>
<td>10&lt;sup&gt;c&lt;/sup&gt;</td>
<td>not specified</td>
</tr>
<tr>
<td>Merlion [39]</td>
<td>2007</td>
<td>microbial</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>4&lt;sup&gt;a, d&lt;/sup&gt;</td>
<td>96</td>
</tr>
<tr>
<td>University of Utah&lt;sup&gt;e&lt;/sup&gt; [41]</td>
<td>2008</td>
<td>marine inv.</td>
<td>HP20SS</td>
<td>5&lt;sup&gt;1&lt;/sup&gt;</td>
<td>20</td>
<td>80</td>
<td>96</td>
</tr>
<tr>
<td>Wyeth [37]</td>
<td>2008</td>
<td>microbial&lt;sup&gt;g&lt;/sup&gt;</td>
<td>HP20/XAD7</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>96</td>
</tr>
<tr>
<td>St Jude Children’s Research Hospital&lt;sup&gt;e&lt;/sup&gt; [40]</td>
<td>2010</td>
<td>plant</td>
<td>PAG</td>
<td>1</td>
<td>24</td>
<td>24</td>
<td>384</td>
</tr>
<tr>
<td>Griffith University [42]</td>
<td>2012</td>
<td>plant</td>
<td>PAG&lt;sup&gt;h&lt;/sup&gt;</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td></td>
<td>marine inv.</td>
<td>Oasis HLB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Performed on C<sub>18</sub> HPLC except for bioLeads where details were not specified

<sup>b</sup> Six extracts originally (5 organic and 1 aqueous); hexane rich extract discarded

<sup>c</sup> Nine fractions from the culture, 1 fraction from the mycelia

<sup>d</sup> Crude extract went straight to HPLC where six fractions were obtained; the first (solvent front) and final fractions were discarded

<sup>e</sup> Organization of corresponding author provided in cases where two or more groups contributed to the publication

<sup>f</sup> Five extracts originally (4 organic and 1 aqueous); aqueous extract discarded

<sup>g</sup> Crude extracts from at least 2 cultures were pooled prior to pretreatment

<sup>h</sup> PAG used on plant samples only
It is interesting to note that MerLion and Wyeth developed similar reversed-phase C$_{18}$ high-performance liquid chromatography (HPLC) methods to address the HTS of extracts derived from microbial samples [37,39]. Importantly, the downstream HTS data reported by both groups vindicates their decision to screen prefractionated libraries compared with crude extracts. In each case, around 80% of the activity was found only in the fractions; the crude extract accounted for approximately 10% of the total activity [37,39].

All groups that worked with plant samples employed polyamide gel (PAG) solid phase extraction (SPE) to remove tannins [38,40,42]. While marine invertebrates do not contain tannins, they can have a high salt content which may dilute compounds of interest and reduce the likelihood of detection. Salts can also suppress ionization in downstream liquid chromatography-mass spectrometry (LC-MS) analysis making identification of the active component more challenging [41]. Salts have effectively been removed from marine invertebrates via SPE through a HP-20SS (polystyrene) adsorbent [41] or by employing a prewash on a HPLC column [42] just prior to fraction collection.

Examples of automated prefractionation following extract screening have also been reported. A recent example by Crews et al. has outlined how HTS of (mainly) crude extracts derived from marine invertebrates, plants and microbial specimens can be prosecuted [43]. Salts and lipophilic compounds were ostensibly removed from the marine organisms via water and hexane extractions, respectively, while tannins from plant samples were excluded by PAG. After screening the ensuing library, hit extracts were subjected to an automated purification and fractionation process that was followed by dereplication and structure elucidation of the active components. However, the data
from Merlion and Wyeth (obtained from 11 and 9 HTS campaigns, respectively) plainly
showed that hit rates were higher when fractions were screened compared with crude
extracts and suggests that screening a prefractionated library would be more efficient.
Indeed, lipophilic false positives like hexylcinnamaldehyde (cLog P ~ 5) were isolated
using this process [43]. Detergent-like compounds and fatty acids/lipids are known to
form aggregates and interfere in certain HTS assays [44,45]. The presence of high Log
P components in the screening library is due to the fact that crude extracts invariably
contain lipophilic material unless some attempt is made to remove them. Moreover,
hexane extraction does not remove all lipophilic molecules [42]. Ultimately, screening
libraries based on crude extracts will be compromised in some way by the presence of
interfering high Log P constituents. Intriguingly, the screening of hexane extracts
(which potentially exacerbates the issues associated with crude extracts containing
lipophilic material) was also reported [43]. In one particular example described by the
authors at length, a common diterpene found in sponges, *i.e.* spongia-13(16),14-dien-
19-oic acid (cLog P = 6.2), was isolated.

The prefractionated approaches listed in Table 2 all attempt to eliminate
interfering, nuisance and spectator compounds that have deleterious effects on an assay,
are frequent hitters, or generally dilute potential molecules of real interest. But what
molecules should be *retained* when viewed through the lens of drug discovery?
Recalling Lipinski’s Ro5, Teague’s original lead-like concept, Oprea’s reduced
complexity sets, Leeson’s analysis of the most important physicochemical properties,
and that Log P is considered “Lord of the Rules” - all of which were developed in
response to issues that originally troubled HTS of combinatorial chemistry libraries - it
follows that construction of a NP screening library, not only devoid of NP PAINS, but
also compliant with established lead- and drug-like parameters, particularly Log P, will
enrich extracts (and subsequent fractions) with compounds that have a much higher
probability of progressing from hits to leads. Clearly, front-loading the screening set
will obviate the necessity to “build-in” functionality that accomplishes a more lead- or
drug-like profile at later lead identification/optimization stages.

Having said this, we do add what we consider to be an important caveat, namely,
the definition of a lead in NP drug discovery should be modified to account for large
MW compounds that can potentially fail up to three Ro5 parameters, i.e. MW, HBD and
HBA, but nevertheless satisfy the most important property, Log P. This captures leads
such as paclitaxel 1, rapamycin 2 and trabectedin 3, which were ultimately drugs in
their own right (Figure 1). This would also include molecules that required only minor
modification such as such as pneumocandin B, 4, which was converted to the antifungal
caspofungin (Cancidas®) 5 (Figure 2). Likewise, we would also contend that NPs
compliant with lead- or drug-like profiles, but nevertheless containing unattractive
structural features such as the five contiguous rings in camptothecin 6, that ultimately
afforded topotecan (Hycamtin®) 7 and irinotecan (Camptosar®) 8, should not
immediately be discounted by the cognitive bias of medicinal chemists.
Returning to the issue of delivering NP leads compliant with lead- and drug-like physicochemical properties in a timely manner; there are two critical issues related to the NP drug discovery workstream that must be addressed before efficiency gains and improved outcomes are forthcoming. First, the screening set should be front-loaded with components having lead- and/or drug-like physicochemical profiles. The profound and positive impact of physicochemical profiling is now reflected in libraries comprised of more lead- and drug-like compounds that, in turn, has translated into more leads from second generation libraries progressing to the clinic [46,47]. We consider the screening of NP libraries that employ similar physicochemical filtering to be absolutely essential if NPs leads are to be taken up in a more widespread manner by industry once again.
Even though the structure (and by logical extension, the concentration) of the individual components in a microtiter plate well containing a NP extract or fraction are generally not known before isolation is attempted, we would contend that an important goal of NP drug discovery is to afford new and novel leads. To our way of thinking, the structure is of secondary importance to ensuring the components are well positioned in lead- and drug-like chemical space in the context of screening extracts and prefractionated libraries. As a consequence, compounds that are isolated downstream from screening will intrinsically adhere to physicochemical profiles that will justify a potentially arduous isolation. Second, if NP drug discovery is to ever compete with timelines for the HTS of pure compounds, then the rate at which NPs are isolated requires vast improvement.

Next generation prefractionated extract libraries filtered on Log P

The correlation between lipophilicity and retention time on a C\textsubscript{18} reversed phase HPLC column has been exploited to develop an alternative method to determine Log P [48]. We have also observed a relationship between lipophilicity, as measured by a compound’s cLog P, and retention time on a C\textsubscript{18} reversed phase HPLC column while undertaking bioassay-guided fractionation projects. Indeed, our earlier biodiscovery efforts employing bioassay-guided fractionation following screening of 140,000 crude extracts (35,000 macro biota extracted with 2 solvents and screened in duplicate) were marred by pursuing late eluting, highly bioactive components that were ultimately found to be useless as NP leads. Once this was realized, and understood in terms of lead- and drug-like physicochemical properties, the process was modified accordingly whereby the crude extracts were re-injected onto a reversed phase C\textsubscript{18} HPLC column and analyzed [49].
Over time, we recognized that molecules isolated from later eluting fractions consistently had both cLog P and Log P values > 5 and, as a result, these fractions were excluded from further analysis. While the fractionation process was a step in the right direction and helped steer isolation projects toward lead- and drug-like compounds that addressed Log P, it did not facilitate timely identification of actives. Clearly, a method that pre-filtered extracts on Log P and allowed some preliminary structural data to be obtained was required. Such a process would obviate the need to analyze extracts post-screening and concomitantly hasten the identification of potential hit compounds.

The lessons learned from fractionating extracts after screening guided our efforts toward developing a generic approach that could be applied to both major and minor constituents within crude or semi-purified extracts. Our current methodology [42] relies on initially preparing crude extracts that are subsequently passed through an SPE cartridge containing Oasis® HLB, which is a copolymer of divinylbenzene and N-vinylpyrolidone. During exhaustive testing of adsorbents, we found that this particular matrix was superior in retaining high Log P components that, in the past, were present in the screening set. The process effectively filters on Log P and allows the earlier eluting components that may contain a highly desirable basic nitrogen atom to be captured. This observation also suggests that screening sets derived from Diaion™ HP-20SS [41] or from crude extracts, particularly hexanes, [43] would contain molecules with Log P > 5 that typically manifest as frequent hitters in biological assays.

In a further refinement, based on Wyeth’s [37] and Merlion’s [39] reported HTS data of their prefractionated libraries, we next aimed to concomitantly improve the quality of screening data and increase the speed of isolation. Thus, the optimized extracts were fractionated on reversed phase C_{18} HPLC to generate a prefractionated library [42]. Eleven fractions were collected from 18,453 extracts to generate a
screening set of 202,983 data points. The HPLC separation is significantly improved following the removal of the highly lipophilic constituents in the previous step. Salts from marine organisms are eliminated using a prewash on the HPLC column prior to fractionation.

This methodology permits MW, structural data and retest of active pure compounds to be acquired in a highly efficient manner directly following primary screening, the point at which little to no chemical information is generally known about bioactive constituents. Thus, confirmation is obtained by LC-MS analysis of the active fraction to further resolve the individual constituents and obtain retention time, ultraviolet (UV) spectra and, importantly, MW information for each component. In some cases it is also possible to obtain preliminary structural data via nuclear magnetic resonance (NMR) spectroscopy. Rescreening of the individual constituents allows the precise compound(s) to be identified and affords some knowledge of potency and novelty. The MS and NMR data can be used for dereplication and also as a trigger to pursue an isolation project. This allows NP drugs that satisfy the Log P criteria, but may fail other lead- and drug-like parameters, like Taxol® 1, Rapamune® 2 and Yondelis® 3, to still be isolated on the basis of potency, new/novel structural motifs or ligand efficiency [50] if the supporting data warrants further effort.

The combined MW, HPLC and other spectroscopic data also facilitates scale-up isolation of the active components without recourse to bioassay-guided fractionation. Here, isolation of the actives from the original biota is guided by the chromatographic retention time of the active in the prefractionated library. Ultimately, mass-directed isolation can be used to home in on the desired molecule(s) in a highly efficient manner.

The first three HTS campaigns that used the prefractionated library were run against malaria [51-54], human African trypanosomiasis (HAT) [55-58], and Yersinia
outer protein E (YOPE) regulation [59]. Sixty compounds were isolated from the malaria screen, 58 from HAT, and 30 from YOPE where activity ranged from good (IC$_{50}$ < 10 µM; 48 malaria, 30 HAT, 7 YOPE) to moderate (IC$_{50}$ 10 – 50 µM; 4 malaria, 10 HAT, 10 YOPE) to poor (IC$_{50}$ > 50 µM; 8 malaria, 18 HAT, 13 YOPE). Twenty-two compounds resulting from these three HTS campaigns have been published, and their chemical structures are shown in Figures 3-5. For the malaria and HAT projects, the focus was on isolating compounds with Log P < 5 and MW < 500. Molecular weight was not as critical as activity for the YOPE project and, as a consequence, bioactives with higher MWs were isolated (Table 3). We have previously reported [42] that the 118 compounds from the malaria and HAT projects were enriched in lead- and drug-like profiles compared with the *Dictionary of Natural Products* [60]. In some cases, the NP lead has progressed into lead identification and optimization projects. So far, synthetic follow-up has been published for YOPE [61].

Interestingly, there were two compounds from the malaria project and seven compounds from HAT that were not complaint with Log P < 5. Although the total number of non-compliant molecules was small, the result was nevertheless puzzling considering the process employed Log P as the primary filter. However, a more thoughtful analysis would suggest that the presence of trifluoroacetic acid (a strong acid) employed to filter on Log P would lead to ionization of NPs containing basic moieties which, in turn, can lead to misleading results as Log P calculations specifically predict the partitioning of neutral (*i.e.* un-ionized) species between n-octanol and water. As a consequence, we believe that the distribution coefficient (Log D) is a more appropriate measure as it considers the distribution of both ionized and un-ionized species at a given pH. Indeed, it has been proposed that Log D at pH 5.5 (Log D$_{5.5}$), which is the pH of the small intestine where oral drug absorption occurs, is a better
description of the lipophilic nature of drug-like molecules under physiological conditions [62]. Importantly, no compounds from the malaria project and only one compound from the HAT project (unpublished structure) exceeded the cut-off value when Log D$_{5.5}$ > 5 was used to determine lipophilicity. Although the HAT-active compound was found in the last fraction, this result could also indicate that the algorithm used to determine Log D$_{5.5}$ was not accurate for this class of compounds. The only molecule in Table 3 that violates Log P (the HAT active compound 19) becomes complaint when Log D$_{5.5}$ is employed as a measure of lipophilicity.

Figure 3. Published structures of natural products isolated from marine invertebrates and plants that are active against the malaria parasite.
Figure 4. Published structures of natural products isolated from marine invertebrates and plants that are active against the HAT parasite.

Figure 5. Published structures of natural products isolated from marine invertebrates and plants that inhibit YOPE.
Table 3. Physicochemical profiling of isolated natural products 9-30.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Physicochemical parameters&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MW</td>
</tr>
<tr>
<td><strong>Antimalarials</strong></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>257</td>
</tr>
<tr>
<td>10</td>
<td>241</td>
</tr>
<tr>
<td>11</td>
<td>294</td>
</tr>
<tr>
<td>12</td>
<td>327</td>
</tr>
<tr>
<td>13</td>
<td>350</td>
</tr>
<tr>
<td>14</td>
<td>790</td>
</tr>
<tr>
<td>15</td>
<td>747</td>
</tr>
<tr>
<td><strong>Antitrypanosomals</strong></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>269</td>
</tr>
<tr>
<td>17</td>
<td>283</td>
</tr>
<tr>
<td>18</td>
<td>356</td>
</tr>
<tr>
<td>19</td>
<td>368</td>
</tr>
<tr>
<td>20</td>
<td>344</td>
</tr>
<tr>
<td>21</td>
<td>312</td>
</tr>
<tr>
<td>22</td>
<td>310</td>
</tr>
<tr>
<td>23</td>
<td>473</td>
</tr>
<tr>
<td>24</td>
<td>470</td>
</tr>
<tr>
<td>25</td>
<td>898</td>
</tr>
<tr>
<td><strong>YOPE actives</strong></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>914</td>
</tr>
<tr>
<td>27</td>
<td>928</td>
</tr>
<tr>
<td>28</td>
<td>522</td>
</tr>
<tr>
<td>29</td>
<td>437</td>
</tr>
<tr>
<td>30</td>
<td>928</td>
</tr>
</tbody>
</table>

<sup>a</sup> All <i>in silico</i> physicochemical properties were calculated using ChemAxon’s Instant JChem software version 3.0.4 (2009).
As with many NP screening programs, the structure of each constituent in an extract or fraction is generally not known prior to screening. The process described here addresses Log P and MW so that the components adhere to contemporary concepts of lead- and drug-likeness. It cannot, however, predefine structures. In some cases, although isolated compounds may exist in lead- and drug-like space, they would not be considered for further development, *e.g.* 16-18 which are reminiscent of DNA intercalators. NP screening must therefore also be able to deliver two or three lead series in the same way screening of pure compound libraries does so that a project team has options to pursue if one series is chemically unattractive. Whilst singletons are not immediately discounted, they can be considered too “high risk” and de-prioritized. This is understandable given the cost to progress a lead through the drug discovery pipeline [47,63,64].

Fortunately, strategies can be easily implemented to deliver lead series that provide insights into structure-activity relationships (SAR). One tactic employs the compilation of a chemical knowledge base during the initial bioassay- or mass-directed isolation process to facilitate isolation of analogues. Here, UV profiling in conjunction with $^1$H NMR spectroscopy permits other fractions derived from the same extract, or extracts from closely related species in the same genus or family, to be scanned for matching scaffolds or substructures. Subsequent scale-up chromatography can then afford additional compounds of interest. This approach was used to isolate pseudoceramines A 26, B 28 and spermatinamine 27 from the marine sponge *Pseudoceratina* sp. in the YOPE project [59]. While compounds 26-28 accounted for the bioactivity in the hit fractions, further large-scale extraction and isolation enabled
the related bromotyrosine analogues, pseudoceramines C 29 and D 30 to be purified and screened.

Another approach enables a lead series to be expanded through simple derivatization (e.g. methylation and acetylation reactions at a primary or secondary amine, or hydroxyl group) on the active compound. Other functional groups like Ar-Br, -COOH, -CHO, and -NH₂ in the hit molecule lend themselves to Suzuki-Miyaura couplings [65], reductive aminations, or the formation of amides/carbamates. By way of example, NPs 31-33 (Figure 6) were identified following screening against the glucocorticoid receptor [66]. Compounds 31-33 displayed IC₅₀ values of 0.9, 27, and 35 μM, respectively. Subsequent methylation and acetylation of the most active compound, endiandrin A (31), using standard conditions yielded two more analogues for the series; diacetyl-O-endiandrin A (34) and dimethyl-O-endiandrin A (35).

Figure 6. Chemical structures for endiandrin A 31, nectandrin B 32, (-)-dihydroguaiaretic acid 33, diacetyl-O-endiandrin A 34, and dimethyl-O-endiandrin A 35.

Finally, there is no doubt that some NPs leads can be challenging to synthesize. Nevertheless, the arsenal of chemo- and stereoselective reactions now available to organic chemists versed in the art has resulted in some spectacular synthesis. The remarkable story of discodermolide, a polyketide isolated from a Caribbean sponge
having ten times the potency of paclitaxel (Taxol®) 1 for tubulin polymerization and microtubule stabilization, from initial synthetic efforts undertaken by various academic groups to the synthesis of 60 g of pure compound by Novartis chemists [67-71] was highlighted in a review by Koehn and Carter [72]. This is testimony to the increasing power of modern synthetic methodologies. Another example is eribulin (Halaven®) 37, an antitumor drug based on a substructure found in the sponge NP, halichondrin B 36 (Figure 7), and manufactured by total synthesis to afford one of the more complex small molecule drugs on the market to date. Indeed, as Koehn and Carter [72] comment, “Success such as these make the compelling case that virtually no crucially important compound is beyond the reach for clinical evaluation.”

Figure 7. Chemical structures for halichondrin B 36 and eribulin 37.

Fermentation is also an option for obtaining multi-gram quantities of the drug directly as is the case with the microbial NP, romidepsin, a histone deacetylase inhibitor [6]. Precursors en route to the actual drug may also be obtained from fermentation to deliver semi-synthetic drugs [73]. Trabectedin (Yondelis®) 3 is a good example. The NP was originally isolated from the marine tunicate Ecteinascidia turbinata. Wild harvest of the tunicate could not deliver a sustainable source of the compound so the bacterium Pseudomonas fluorescens was used to afford the related safracin compounds [74]. A semi-synthetic procedure based on conversion of microbially produced cyanosafracin B was then developed to produce kilogram quantities of trabectedin [75]. Allowing
microbials to produce scaffolds for further derivatization also facilitates the search for more potent and selective compounds compared with the NP. The semisynthetic cholesterol-lowering statin, simvastatin, for example, is more potent than the fungal NP, lovastatin [73].

Guiding principle 2: Addressing access to genetic resources

The Convention on Biological Diversity

Prior to the CBD, the world’s biodiversity was viewed as the common heritage of mankind. The advent of the CBD has significantly and irrevocably altered biodiscovery practices by providing a legal instrument that gives countries sovereignty over their biodiversity and genetic resources and the right to control access [76]. In return, signatories of the CBD have the responsibility to conserve and use their biodiversity in a sustainable manner, and via Article 15, must facilitate access to genetic resources and ensure equitable sharing of benefits arising from their use [101].

Whilst the CBD provided a clear policy and legal basis for source countries to regulate access and benefit sharing (ABS), practical implementation of the third objective (benefit sharing) and Article 15, in particular, has proven difficult. This was due to several factors. First, there was the not insignificant matter of identifying a negotiating party. Was it an individual, a community, government agency, or the national government? Second, agreement had to be reached on a common understanding of likely benefits such as the quantity and likelihood of monetary profits, and the value of non-monetary benefits such as support for conservation and capacity building. Third, the “value” of the players’ respective contributions had to be considered; custodianship and conservation of the resource from the provider versus
innovation and investment costs from those developing genetic resources [77]. The initial result in some instances was, at best, excessive negotiation and transaction costs and, at worst, legal uncertainty - both disincentives to the industry investment required to drive biodiscovery programs [26]. The CBD quite unintentionally spawned an era where access to nature for biodiscovery had actually become more difficult. At the same time, combinatorial chemistry was being pitched as a safer, more viable alternative for the emerging HTS paradigm. It is little wonder then that the CBD failed to deliver the anticipated bioprospecting bonanza to the developing world [78].

While implementation of international treaties like the CBD requires each country to develop their own domestic laws and regulatory frameworks, the CBD provided sufficient flexibility and policy guidance to enable individual institutions to develop CBD compliant agreements ahead of schedule.

The Bonn Guidelines

In an effort to facilitate national legislation based on the CBD, the Bonn Guidelines were drafted to provide helpful and specific, though voluntary, guidance on measures related to implementation of ABS [103]. While some countries pursued measures to promote biodiscovery and implement the Bonn Guidelines [79-81],[103] the governments of many developing countries felt that their voluntary nature didn’t go far enough to combat biopiracy. Indeed, in 2004, the CBD and Bonn Guidelines were themselves the recipient of the notorious Captain Hook Biopiracy Award, for “creating the illusion that equitable benefit sharing is being addressed while facilitating biopiracy…” [102]. After pressure from many developing countries in a range of international forums, including the 2nd World Summit on Sustainable Development and
the 57th meeting of the UN General Assembly, the CBD was handed a mandate to
develop a new international regime for access and benefit sharing [26],[104].

335  *The Nagoya Protocol*

The resulting Nagoya Protocol is a legally binding protocol to the CBD [104].

As a consensus outcome following 6 years of intense, robust and often volatile
negotiations over contentious issues and diametrically opposed views, the Nagoya
Protocol is indeed a remarkable achievement. It was presented to (and adopted by) the
10th Meeting of the CBD’s Conference of the Parties in 2010 and, as at November 2011,
had collected 64 signatories. It is expected to come into force in 2012.

The Nagoya Protocol is a major step forward in that it provides agreed
principles to deliver legal certainty and transparency for the entire value-chain of
biodiscovery, from original providers of the genetic resources, through to all stages in
the ensuing research and development, and the commercialisation process. Importantly,
the protocol includes provisions for clear demonstration of compliance, such as
certificates of compliance, and the introduction of checkpoints to transparently monitor
compliance. This is an important point as the potential for legal provenance and
compliance with the Nagoya Protocol may become a criterion for patentability in the
future [82].

The challenge ahead is for parties to avoid developing onerous requirements that
increase transaction costs and undermine legal certainty. For example, certificates of
compliance must be practical and enforceable [83,84] and checkpoints for on-going
compliance should make use of existing regulatory procedures. Simple and accessible
web-based databases and visualisation tools have been proposed as suitable methods for
transparent compliance monitoring [85], and these hold great promise.
In the case of developed, megadiverse countries like Australia, a recent analysis of the existing and developing domestic regulatory framework for ABS has found that it is already largely compliant with the Nagoya Protocol [86]. Moreover, government authorities in Australia are extremely mindful of the need to avoid further regulatory burden and instead look to existing mechanisms for passive check-points, such as through funding bodies [87]. Taking a balanced approach that considers both the position of the owner of genetic resources and that of industry will facilitate biodiscovery. By addressing the legitimate concerns of developing nations with megadiversity through the CBD, Bonn Guidelines and Nagoya Protocols, NP drug discovery is now poised to re-emerge in a landscape of international and national legal certainty.

**Future perspective**

The efficient screening and identification of small molecule modulators of biological systems using HTS and high-content screening (HCS) will be employed for many years to come. It is therefore paramount that NP researchers deliver new and innovative approaches that dovetail with automated screening in order for NP drug discovery to increase its current efficiency in identifying new drugs leads for downstream development.

Innovation in delivering lead molecules is the key message here as many of the technical challenges relating to isolation and structural elucidation of bioactive NPs have been addressed [88,89]. Many NP PAINS, for instance, can be removed prior to screening while prefractionation to afford a screen-friendly set is readily accomplished by HPLC. Moreover, it is becoming increasingly commonplace that the structure of many active component(s) in a fraction can be determined by LC-MS and NMR
spectroscopy. This not only allows new and novel molecules to be distinguished from previously identified compounds but also facilitates prioritization and allocation of resources to isolation projects. The isolation of a pure compound in milligram quantities can be achieved in a matter of hours using mass-directed isolation. Subsequent structural elucidation by NMR spectroscopy is likewise completed in the same time frame by skilled NP chemists. The entire procedure of going from an active fraction to a defined molecule can be a matter of days rather than months [90].

Other improvements to NP drug discovery that borrow from efficiency gains first employed in the HTS of combinatorial chemistry libraries are likewise possible. For example, a drawback of preparing libraries in plate-based formats is the need to expose an entire plate’s contents to the potentially deleterious effects of the open laboratory when only one fraction may be required for retesting. Dimethyl sulfoxide (DMSO) is notoriously hygroscopic [91] and it is common knowledge among compound management groups that the ingress of adventitious moisture can result in precipitation of samples or an increased rate of degradation over successive freeze-thaw cycles [92,93]. A modified procedure that addresses the issue of water uptake and limited sample volume (80 µL) per 384 well plate will go some way to further dovetailing NP drug discovery with current paradigms. Microtubes, for example, permit cherry picking of individual fractions and subsequent transfer to assay-ready plates via piercing of presplit septa [94,95].

In terms of chemistry, innovative new leads may be acquired from previously untapped sources of biota including a broader range of plant species, marine organisms and microbes [96]. The marine environment is proving to be a particularly exciting source of potential new leads [89,97].
Ultimately though, NP drug discovery will only advance in quantum leaps when these incremental improvements are coupled with a process that fundamentally aligns itself with the now recognized principles of drug discovery. The optimization of extracts compliant with Log P that can then be fractionated to yield a screening set of low chemical complexity per fraction while retaining chemical diversity space is just one of potentially many ways this challenge can be addressed [42]. Coupling libraries biased toward molecules with Log P values attractive for downstream development together with mass-directed isolation (and incorporation of a MW filter) can simultaneously address concerns about molecular obesity and the inherent slowness of working with NPs [13,98].

Although not directly relevant to discovering lead compounds for medicinal chemistry, finding small-molecule probes that selectively interrogate physiological processes is also an important application for NPs, and potentially a critical step in showing that a particular site on a protein or process is “druggable” [99]. The methodology described here can be modified to efficiently deliver NP probes (chemical tools) of interest to the academic community for chemical biology or chemical genetics projects. However, it is critical that the lessons learned from the early days of HTS are not forgotten even though probe compounds can explore chemistry space outside that typically used to define drugs and leads. If probe libraries are comprised of NP PAINS, non-lead-like molecules and aggregators, then the literature will be replete with artefact data [5]. As Rishton puts it, “a good lead is a good tool is a good probe” while “a non-lead is a non-tool is a non-probe” [100].

Finally, we perceive that policy and legal impediments surrounding biodiscovery have been largely settled in many countries and are heading in the right direction in others. By way of example, legislation enacted by Australian Governments addresses
In summary, the major impediments to discovery programs focused on NPs, i.e. the issues of screening libraries suitable for downstream drug discovery and development, and the policy problem of legal certainty and compliance, are being resolved. The stage is now set for a renaissance of NPs to impact positively on drug discovery and emerging interfacial sciences like chemical biology.

**Executive summary**

*Setting the scene*

- Slightly over 1,100 small molecule NCEs were introduced to the clinic from January 1981 to December 2010. Of these, 34% were NPs, or a semi-synthetic analogue.
- While this metric reflects the fact that NPs have historically been a fertile source of new drugs for the pharmaceutical industry, this once popular approach has waned considerably over the past two decades as a paradigm shift toward the HTS of mega libraries comprised mainly of molecules with non-natural (synthetic) motifs has unfolded.
- This paradigm shift coincided, *inter alia*, with the introduction of the CBD, which initially, albeit unintentionally, stymied access to biodiversity for NP drug discovery programs on account of policy and legal hurdles.
- The classical NPs drug discovery approach employing bioassay-guided fractionation has not been able to deliver compounds with physicochemical profiles considered
essential for development into an orally administered drug in the same timelines as
the HTS of pure compound libraries.

Addressing physicochemical properties in natural product screening and improving the
speed that hits, leads and drugs are isolated.

- The physicochemical profiling of compound libraries that occurs in the
  pharmaceutical industry has become an intrinsic part of the design and selection
  process.
- One physicochemical descriptor has received particular attention; Log P, the
  partition coefficient of a compound between n-octanol and water, which is a
  measure of a compound’s lipophilicity.
- In one telling investigation, Leeson and Davis identified that the lipophilicity of
drugs, as measured by cLog P, has remained constant over time suggesting that
departure from a narrow window can adversely affect the efficacy of a potential
therapeutic.

Next generation prefractionated extract libraries filtered on Log P

- NP workstreams that can align screening sets with lead- and drug-like
  physicochemical profiles facilitate the downstream triage of biological and chemical
  data.
- A NP drug discovery strategy based on Log P has proven its utility against three
  anti-infective screens and holds much promise for timely isolation of NP leads at a
  hitherto unachievable rate.

Addressing access to genetic resources

- Alongside demonstrated adaptation of NP screening sets for HTS, there has been
  resolution of the legal and policy impediments to biodiscovery. The Nagoya
Protocol has established detailed benchmarks in international law for access and benefit sharing.

- The scene is now set for a renaissance in biodiscovery by combining the molecular diversity inherent in nature with the discovery potential of HTS.

References

3. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery*, 6(11), 881-890 (2007). *A very good overview that outlines how drug-like concepts have caught the attention of the pharmaceutical and how they may be used to stymie innovation.*
10. Leeson PD, Davis AM. Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem.*, 47(25), 6338-6348 (2004). **Physicochemical profiling is undoubtedly important in modern drug discovery, but what makes this article different from the rest is the way the issue was approached.**


33. Quinn RJ, Carroll AR, Pham NB et al. Developing a Drug-like Natural Product Library. J. Nat. Prod., 71(3), 464-468 (2008).* One of the earliest articles to articulate the importance for natural products to undertake physicochemical profiling in line with current drug discovery programs.


88. Harrigan GG, Goetz GH. Chemical and Biological Integrity in Natural Products Screening. *Comb. Chem. High Throughput Screen*., 8(6), 529-534 (2005).


**Websites**


The Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. 2011. http://www.cbd.int/abs/text/