Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Chemical Biology

The impact of natural products upon modern drug discovery

A Ganesan

In the period 1970–2006, a total of 24 unique natural products were discovered that led to an approved drug. We analyze these successful leads in terms of drug-like properties, and show that they can be divided into two equal subsets. The first falls in the 'Lipinski universe' and complies with the Rule of Five. The second is a 'parallel universe' that violates the rules. Nevertheless, the latter compounds remain largely compliant in terms of log *P* and H-bond donors, highlighting the importance of these two metrics in predicting bioavailability. Natural products are often cited as an exception to Lipinski's rules. We believe this is because nature has learned to maintain low hydrophobicity and intermolecular H-bond donating potential when it needs to make biologically active compounds with high molecular weight and large numbers of rotatable bonds. In addition, natural products are more likely than purely synthetic compounds to resemble biosynthetic intermediates or endogenous metabolites, and hence take advantage of active transport mechanisms. Interestingly, the natural product leads in the Lipinski and parallel universe had an identical success rate (50%) in delivering an oral drug.

Address
School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom

Corresponding author: Ganesan, A (ganesan@soton.ac.uk)

Current Opinion in Chemical Biology 2008, **12**:306–317

This review comes from a themed issue on
Molecular diversity
Edited by Gerry Maggiora and Christopher Hulme

1367-5931/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.cbpa.2008.03.016

Introduction

Nature has evolved over time to produce a bewildering diversity of secondary metabolites. Based on empirical observations and folklore, natural product extracts were the first, and for a long time, the only medicines available to mankind. Although crude extracts remain the primary healthcare for a majority of the world's population, they are largely supplanted by active pharmaceutical ingredients in the Western world. Furthermore, the dependence upon natural products is no longer obligatory and many drugs are purely synthetic small molecules or manufactured biologics such as vaccines, antibodies, and recombinant proteins. Given these alternatives, there needs to be a rationale for the continued exploration of natural products as leads, and two major arguments can be put forward:

- *Premise 1*: Natural products interrogate a different area of chemical space than synthetic compounds.
  If this were untrue, it would be more profitable to concentrate on more readily accessible synthetic compounds. However, there *are* significant differences in the molecular architecture produced by nature when compared to the synthetic molecules of medicinal chemistry [1••,2••,3•,4••,5]. Although both aim to produce biologically active matter, biosynthesis operates under a different set of constraints and guiding principles than the synthetic organic chemist (Table 1). In nature, a very parsimonious set of building blocks is utilized, whereas we have access to tens of thousands of commercially available chemicals. As a consequence, we achieve numbers by repeating a reliable sequence of reactions over and over again while changing the input. Nature, on the contrary, diversifies by taking its limited building blocks and partitioning them into a multitude of pathways. Further differences occur in the type of synthetic transformation performed. Nature is oxophilic, and has developed enzymes that exquisitely accomplish site-selective C–H activation [6••,7•] to introduce oxygen and discriminate between numerous functional groups at different oxidation levels. Meanwhile, medicinal chemistry concentrates on nitrogen and often includes ancillary atoms such as sulfur and halogens that are relatively rare in nature. Finally, the chiral enzymes of biosynthesis usually yield the product as a single stereoisomer. Although medicinal chemists are themselves chiral and target chiral enzymes or receptors, they prefer to work in 'flatland' with molecules low in stereochemical features.
- *Premise 2*: Natural products are amenable to further improvement.
  If this were untrue, the natural product extracts would suffice, or the purified natural product would become the final drug without modification. Although this can often be the case, it runs counter to the drug discovery paradigm where initial leads are subjected to extensive medicinal chemistry campaigns before a candidate is selected. *A priori*, natural products should undergo the same iterative cycle of improvement, as their evolutionary reason for existence is not for use as a therapeutic agent. Thus, one can expect that the natural product can be further improved, whether in terms of efficacy and selectivity for the target or achieving optimal pharmacokinetic and pharmacodynamic properties. For

Table 1

Some fundamental differences between biosynthesis and synthesis

	Biosynthesis	Synthesis
Building blocks	Few	Many
Strategy	Branching of intermediate	Alteration of building block
Scaffold diversity	High	Low
Functional group tolerance	High	Low
Novel motifs	Common	Rare
C–H activation	Common, site-specific	Rare
Stereocontrol	Easy, enantioselective	Difficult, case-by-case basis

example, the opium alkaloid morphine is an important drug that is obtained solely from nature and continues to be used in both extract and pure form. At the same time, morphine has encouraged the discovery of many semisynthetic and fully synthetic compounds based on the same pharmacophore that are successful second-generation opioid drugs.

The molecular architecture of drug-like matter

Biological space is modest in size — the human genome is on the order of 3×10^4 genes of which only a fraction is targeted by current therapeutics [8,9•]. Meanwhile, chemical space is infinite, and there are an estimated [10] 10^{60} organic compounds with a molecular weight cutoff of 500. Our imperfect understanding of which areas of chemical space are best suited to interact with biological space is the major bottleneck of drug discovery. In recent years, there were various attempts at narrowing this gap by statistical analyses to define descriptors for small-molecule drug-like space. The most famous, Lipinski's 'Rule of Five' [11••] predicts passive oral absorption based on log *P*, molecular weight, and H-bond donors and acceptors. Subsequently, others have considered the importance of parameters such as the number of rotatable bonds, polar surface area (PSA) [12••], and ligand efficiency [13,14••].

The success of these rules lies in the ease with which the metrics are calculated, and the careful choice of dataset leading to their derivation. For example, Lipinski restricted his study to compounds reaching Phase II clinical trials, with the assumption that failures because of poor permeability would have dropped out at an earlier stage of the discovery process. In addition, the emphasis was on small molecules, with peptides or nucleotides deliberately excluded. Meanwhile, it is useful to keep in mind that the 'rules' are in fact guidelines [15•] and that 20% of all oral drugs violate at least one rule.

Lipinski has noted [16••] that many natural products remain bioavailable despite violating the Rule of Five. Computational comparisons (for example [3•]) of natural product datasets compared to synthetic compounds or drugs have not uncovered a reason for this unexpected behavior. Although the 'average' natural product differs from the 'average' synthetic drug in terms of elemental composition and stereochemical complexity, both show similar values for Lipinski parameters. In part, this must be because of the smoothing out effect when calculating averages with large datasets. Furthermore, Lipinski's analysis was based on Phase II candidates, whereas the natural product datasets were not filtered in a similar way. We should really concentrate on the subset of *successful* natural products that went into development, just as Lipinski did not look at *all* compounds in medicinal chemistry programs. Surprisingly, this type of analysis has not been previously reported, and is the subject of the next section.

The molecular architecture of successful natural products

How many marketed drugs have a natural product origin, or are based on a pharmacophore first identified in a natural product? This question is easy to answer, thanks to the excellent and comprehensive surveys by Newman at the National Cancer Institute (NCI). The most recent survey [17••], covering the period 1981–June 2006, lists a total of 1184 new chemical entities (NCEs) receiving approval. Of these, 52% have a natural product connection, 18% are biologics, and 30% purely synthetic. The question we would like to ask is the reverse: How many *unique* natural products led to a successful drug launch? Systematic data on this is lacking, and I have attempted to answer this question by metaanalysis of the 1981–2006 timeslice covered in Newman's review. In doing so, a number of filters were applied to all the 1184 NCEs reported by Newman, as follows:

1. Drugs that were inspired by a natural product lead discovered pre-1970 were discarded. This is an arbitrary decision to emphasize newer natural products that were the result of modern screening campaigns rather than second or later generation drugs based on classic natural products.
2. Where there are two or more natural product leads of closely related structure, I have selected the first compound disclosed in the literature. For example, the statin class of cholesterol-lowering drugs grew from the natural products mevinolin and compactin, which differ only by a methyl group.
3. One of Newman's categories is 'ND' or natural product derived. This literally refers to the starting material used in the drug's preparation and is not synonymous with a lead. For instance, semisynthetic steroid hormones are usually manufactured by multi-step routes from plant steroids but the latter did not serve as an inspiration for the discovery of the former.

4. Drugs that are based on an understanding of human physiology or endogenous ligands are excluded. Thus, the origin of CNS drugs can often be traced back to human neurotransmitters rather than a secondary metabolite isolated from another organism.
5. Drugs that can be rationally predicted by a mechanistic understanding of the target's action are excluded. In the antiviral area, many drugs are nucleoside analogs or transition state inhibitors of viral proteases. In both cases, drugs that work by such mechanisms did not need the discovery of a natural product lead.

Some of the filters may be contentious, but I believe the net result is a reasonable first-pass approximation. It tells us how many unique natural product chemotypes discovered in 1970 or later led to a marketed drug in 1981–2006. The reader might want to pause and take a stab at the answer before reading on. When I have polled academic and industrial chemists, a pretty broad range of values has ensued!

A full list of the natural product linked drugs in Newman's dataset is provided in the supplementary information, subdivided into categories with light annotation. Applying the above filters yields a set of 24 natural products that were the starting point for marketed drugs in the 25-year period 1981–2006 (Table 2, structures in Figure 1). Of these, 19 were isolated from soil microorganisms, actinomycetes in particular, while the remaining 5 were of plant origin. Marine natural products are conspicuously absent, as their systematic exploration became widespread only recently. The majority of these successful natural products were discovered by the pharmaceutical industry through high-throughput screening methods, with others coming from research institutes specializing in natural product chemistry. Overall, nearly half of these leads were discovered in Japan, a testament to the country's leading expertise in natural products. The third column from the right illustrates the arduous journey from lead discovery to approved drug, with a lag of over a decade being typical. It also shows that a single compound can be the lead for multiple drugs. The second last column gives the route of administration for the approved drugs, while the final column is an indicator of the success of natural product derived drugs, with 17 entries among the top 500 drugs of 2006. Ten of the 24 leads were antimicrobial agents (3 for intracellular targets, 7 for cell wall or membrane targets) while 14 were against targets in man (11 intracellular, 3 membrane or extracellular).

In terms of structure, these 24 leads are predominantly of polyketide, peptide, or terpenoid origin. Alkaloids are absent, and a possible explanation is that this class was among the first to be examined as their basic properties aid isolation, and highly biologically active members were already heavily exploited pre-1970. Although a cursory glance at Figure 1 might indicate a random set of structures, they can be classified into two broad categories (Tables 3 and 4) in terms of chemical space. The tables list values for molecular descriptors used in Lipinski's rules, as well as Veber's number of rotatable bonds and PSA, and Hopkins's ligand efficiency. All values that fall beyond the cutoff are shaded (MW $>$ 500, $C\log P > 5$, Hd $>$ 5, Ha $>$ 10, Rot $>$ 10, PSA $>$ 140, HA $>$ 35). In addition, I have noted the number of stereogenic centers as a measure of architectural complexity and a predictor of tractability for medicinal chemistry efforts. A generous St value $>$5 has arbitrarily been set as one that would discourage synthesis of analogues.

Exactly half of the 24 natural products lie in what can be called the 'Lipinski universe' (Table 3). The 'Rule of Five' is violated only once (spergualin has too many H-bond donors) in these compounds, and many of them are closer to the tighter constraints placed on lead-like space [18••] rather than a drug. The average molecular weight is only 319, although the natural product origin is betrayed in the degree of stereochemical complexity, with an average of 4 stereogenic centres present. The other half (Table 4) displays very different molecular properties. We might describe them as a 'parallel universe' to Lipinski space. The existence of these two separate clusters can be visualized graphically, as shown in the molecular weight distribution (Figure 2). With the exception of pseudomonic acid and lipstatin, all of the natural products in the parallel universe have at least two Lipinski 'alerts'. This is accompanied by high values for rotatable bonds, PSA, heavy atoms, and stereogenic centers. Of the 12 natural product leads in the Lipinski universe, 6 (50%) led to orally administered drugs, while the 12 leads in the parallel universe had the identical outcome of 6 (50%) oral drugs. Thus, for these natural products, compliance or otherwise of Lipinski rules is not a reliable predictor of oral bioavailability. The subset of orally administered drugs, just like the total, shows a bimodal distribution when plotted against molecular weight of the lead (Figure 2).

Rules for successful natural products

Tens of thousands of biologically active natural products were discovered in the period 1970–2006. Yet, only 24 of these had the 'right stuff' that resulted in an approved drug [19•]. On the basis of the data in the preceding section, we can devise some guiding principles that will help in assessing the worth of natural product leads (or indeed synthetic compounds as well) as potential therapeutic agents.

log *P* is the lord of the rules

Although natural products in the 'parallel universe' (Table 4) may appear to break all the rules, they are remarkably compliant with regard to log *P*. This under-

Table 2

The 24 natural products discovered since 1970 that led to an approved drug in 1981–2006

Lead, year, and structural class	Origin	Discoverer	Drug, year	Route	Ranking
Validamycin, 1970 Oligosaccharide	Actinomycete	Takeda (JAP)	Acarbose, 1990 Voglibose, 1994	po po	357
Midecamycin, 1971 Macrolide	Actinomycete	Meiji (JAP)	Miocamycin, 1985	po	
Pseudomonic acid, 1971 Polyketide	Bacteria	Beecham (UK)	Mupirocin, 1995	top	436
Taxol, 1971 Diterpene	Plant	Res Triangle Inst/NIH (USA)	Paclitaxel, 1993 Docetaxel, 1995	iv iv	81 123
Cephamycin C, 1971 β-lactam	Actinomycete	Lilly (USA)	Moxalactam, 1982 Cefotetan, 1984 Cefbuperazone, 1985	iv iv iv	
Coformycin, 1974 Nucleoside	Actinomycete	Inst Microbial Chem (JAP)	Pentostatin, 1992	iv	
Echinocandin B, 1974 Cyclopeptide	Fungus	Ciba-Geigy (SWI)	Caspofungin, 2001 Micafungin, 2002 Anidulafungin, 2006	iv iv iv	293
Mizoribine, 1974 Nucleoside	Fungus	Toyo (JAP)	Mizoribine, 1984	po	
Rapamycin, 1974 Polyketide	Actinomycete	Ayerst (CAN)	Sirolimus, 1999 Everolimus, 2004 Zotarolimus, 2005	po po po	434
Compactin, 1975 Polyketide	Fungus	Sankyo (JAP)	Lovastatin, 1984 Simvastatin, 1988 Pravastatin, 1989 Fluvastatin, 1994 Atorvastatin, 1997 Cerivastatin, 1997 Pitavastatin, 2003 Rosuvastatin, 2003	po po po po po po po po	264 2 41 195 1 71
Cyclosporine A, 1975 Cyclopeptide	Fungus	Sandoz (SWI)	Ciclosporin, 1983	po	122
Lipstatin, 1975 Polyketide	Actinomycete	Roche (SWI)	Orlistat, 1987	po	277
Bestatin, 1976 Peptide	Actinomycete	Inst Microbial Chem (JAP)	Ubenimex, 1987	po	
Thienamycin, 1976 β-lactam	Actinomycete	Merck (USA)	Imipenem, 1985 Meropenem, 1994 Panipenem, 1994 Faropenem, 1997 Biapenem, 2002 Ertapenem, 2002 Doripenem, 2005	iv iv iv po iv iv iv	247 231
Artemisinin, 1977 Sesquiterpene	Plant	Qinghaosu Res Grp (PRC)	Artemisinin, 1987 Artemether, 1987 Artenusate, 1987 Arteether, 2000	po po po po	
Forskolin, 1977 Diterpene	Plant	Hoechst (IND)	Colforsin, 1999	iv	
Plaunotol, 1977 Diterpene	Plant	Sankyo (JAP)	Plaunotol, 1987	po	
Avermectin B_1a, 1979 Polyketide	Actinomycete	Kitastato Inst (JAP)/Merck (USA)	Ivermectin, 1987	po	
SQ26,180, 1981 β-lactam	Actinomycete	Squibb (USA)	Aztreonam, 1984 Carumonam, 1988	iv iv	

Table 2 (*Continued*)

Lead, year, and structural class	Origin	Discoverer	Drug, year	Route	Ranking
Spergualin, 1981 Peptide	Bacteria	Inst Microbial Chem (JAP)	Gusperimus, 1994	iv	
Arglabin, 1982 Sesquiterpene	Plant	Inst Phytochem (USSR)	Arglabin, 1999	po	
FK506, 1984 Polyketide	Actinomycete	Fujisawa (JAP)	Tacrolimus, 1993	po	103
Daptomycin, 1986 Cyclodepsipeptide	Actinomycete	Lilly (USA)	Daptomycin, 2003	iv	
Calicheamicin γ_1, 1988 Polyketide	Actinomycete	Lederle (USA)	Gemtuzumab, 2000	iv	

The second last column lists the route of administration (po = oral, iv = intravenous, top = topical) and the final column gives the ranking among the global top 500 drugs of 2006, according to IMS Health.

lines the central importance of log *P* in drug discovery. Although an increase in log *P* can often yield a higher affinity for the target, it tends to be outweighed [20••] by pharmacokinetic liabilities such as solubility, permeability, plasma protein binding, metabolic turnover, and toxicity. The single most important lesson from natural products lies in their ability to maintain low log *P* regardless of other characteristics. In the Lipinski universe (Table 3), average log *P* is 0, while in the parallel universe (Table 4) it has only risen to 2.2 despite an average molecular weight of 917. It is thus possible to operate in non-Lipinski space with high molecular weight and large numbers of H-bond acceptors and PSA, provided lipophilicity is not compromised. To do so requires the presence of polar functional groups, and this is compatible with biosynthetic pathways which are extremely chemoselective and regioselective. Making such compounds is a lot more challenging for medicinal chemists, and is likely to involve long routes with protection and deprotection schemes for specific functional groups. Consequently, when compounds with higher molecular weight are made synthetically, log *P* usually suffers. To quote Lipinski [16••], '. . .if you look at companies that are selling compounds they usually quote Rule of Five compliance rates and typically what you find is that the parameter that is most difficult to control combinatorially is lipophilicity'.

A study [21•] by AstraZeneca indicates that lipophilic compounds are the most likely to be discontinued during development. More recent papers from AstraZeneca [22•,23••] track the evolution of molecular properties of oral drugs and drug discovery programs over time at several pharmaceutical companies. These show that the molecular weight of compounds made by medicinal chemists has risen, perhaps because of the complexity of modern targets now that the 'low-bearing' fruit in drug discovery has been harvested. At the same time, this increase has not been accompanied by a corresponding increase in log *P*, showing that medicinal chemists are consciously aware of the importance of avoiding high lipophilicity.

If you cannot be passive, be active

The rules for assessing druglikeness are based upon passive absorption through the lipid membrane, and are no longer applicable for carrier-mediated or active transport. Such processes may be more common than historically believed [24••]. A recent estimate suggests 758 transporters in the human genome, with the substrate tolerances unknown for most of them. Many synthetic orally bioavailable drugs may have a component of active transport, and such mechanisms may account for the anomalously high bioavailability of natural product drugs that violate the rules. Because biosynthesis pathways have common features (Table 1), a foreign natural product is more likely than a foreign synthetic molecule to be similar to endogenous ligands or metabolites and accepted as a substrate by transporters. Of course, once such natural products get into the cell, there is a separate issue in their recognition as a xenobiotic, being susceptible to clearance by active efflux pumps.

It is tempting to speculate that the body has evolved two parallel strategies for avoiding high molecular weight xenobiotics. Those with high log *P* are 'influx-limited', because of poor solubility, distribution, and propensity for first-pass metabolism. Thus, the defense mechanism is to avoid such compounds reaching the site of action in the first place. On the contrary, those compounds with low log *P* are 'efflux-limited'. They reach target cells upon which they can be absorbed by active transport. The defense mechanism relies on similarly active efflux for clearance. For such compounds, efforts to rationally minimize efflux [25•] or improve passive transport should be an important stage of the drug development process.

Not all H-bonds are equal

The energy penalty of a H-bond donor that needs to interact with bulk water is higher than that of a H-bond acceptor that is reacting in a reversible manner. This is implicit in the Lipinski rules, as the cutoff for H-bond donors is half of that for H-bond acceptors. Nature follows the same logic, as only 4/12 compounds in Table 4 exceed

Figure 1

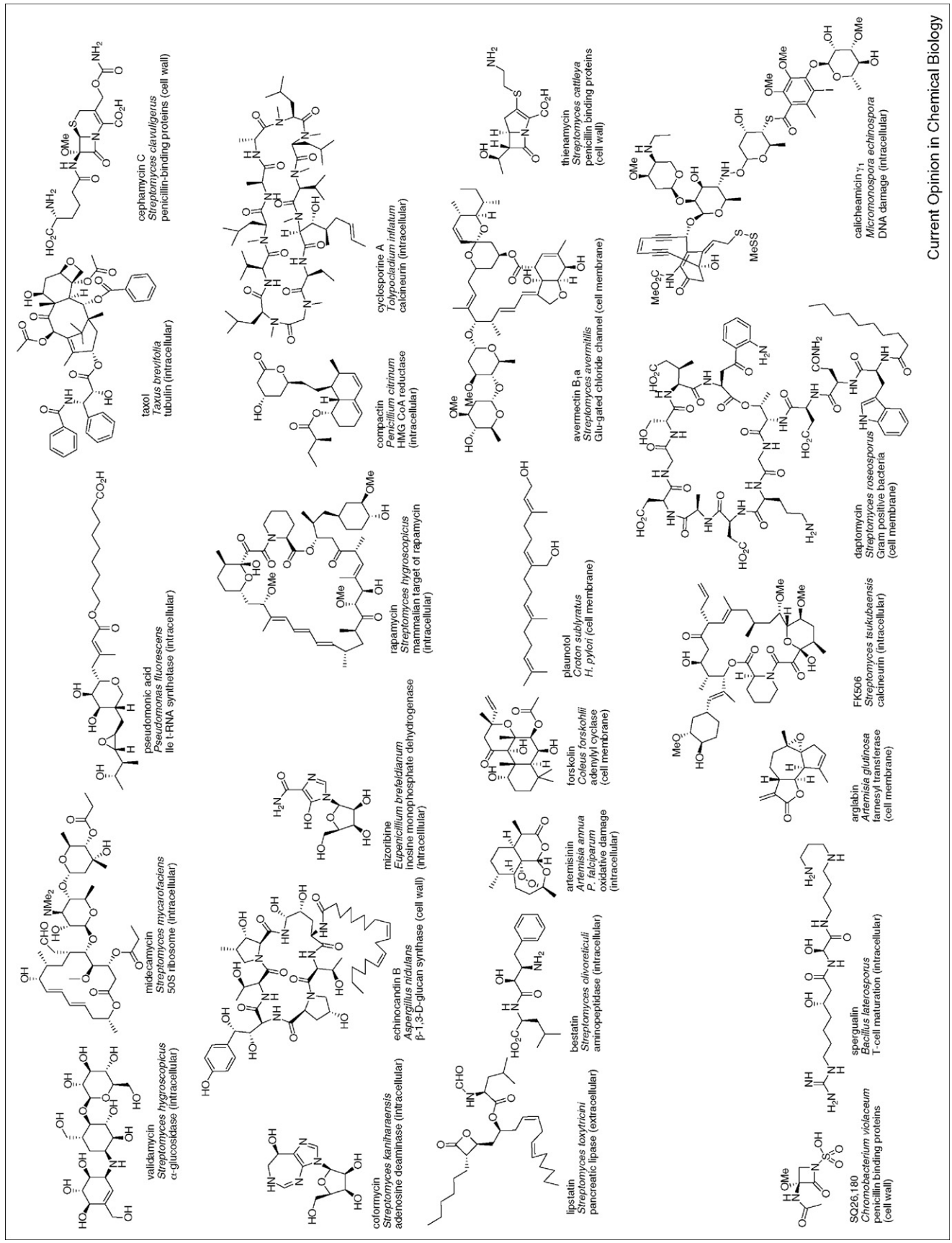

validamycin
Streptomyces hygroscopicus
α-glucosidase (intracellular)
colomycin
Streptomyces kaniharaensis
adenosine deaminase (intracellular)
lipstatin
Streptomyces toxytricini
pancreatic lipase (extracellular)
SQ26,180
Chromobacterium violaceum
penicillin binding proteins
(cell wall)
midecamycin
Streptomyces mycarofaciens
50S ribosome (intracellular)
echinocandin B
Aspergillus nidulans
β-1,3-D-glucan synthase (cell wall)
spergualin
Bacillus laterosporus
T-cell maturation (intracellular)
pseudomonic acid
Pseudomonas fluorescens
Ile t-RNA synthetase (intracellular)
mizoribine
Eupenicillium brefeldianum
inosine monophosphate dehydrogenase
(intracellular)
artemisinin
Artemisia annua
P. falciparum
oxidative damage
(intracellular)
bestatin
Streptomyces olivoreticuli
aminopeptidase (intracellular)
forskolin
Coleus forskohlii
adenylyl cyclase
(cell membrane)
rapamycin
Streptomyces hygroscopicus
mammalian target of rapamycin
(intracellular)
plaunotol
Croton sublyratus
H. pylori (cell membrane)
arglabin
Artemisia glutinosa
farnesyl transferase
(cell membrane)
taxol
Taxus brevifolia
tubulin (intracellular)
compactin
Penicillium citrinum
HMG CoA reductase
(intracellular)
FK506
Streptomyces tsukubaensis
calcineurin (intracellular)
cephamycin C
Streptomyces clavuligerus
penicillin-binding proteins (cell wall)
cyclosporine A
Tolypocladium inflatum
calcineurin (intracellular)
avermectin B1a
Streptomyces avermitilis
Glu-gated chloride channel (cell membrane)
daptomycin
Streptomyces roseosporus
Gram positive bacteria
(cell membrane)
thienamycin
Streptomyces cattleya
penicillin binding proteins
(cell wall)
calicheamicin γ1
Micromonospora echinospora
DNA damage (intracellular)
Current Opinion in Chemical Biology

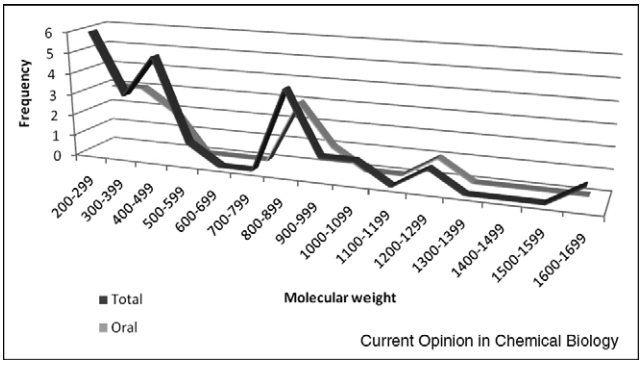
Table 3

The 'Lipinski universe' of successful natural product leads

NP	Formula	MW	log *P*	Hd	Ha	Rot	PSA	HA	St
Cephamycin C	$C_{16}H_{21}N_3O_9S$	431	-4.3	4	10	11	186	29	3
Coformycin	$C_{11}H_{16}N_4O_5$	284	-1.9	5	9	2	132	20	5
Mizoribine	$C_9H_{13}N_3O_6$	259	-1.4	5	8	3	151	18	4
Compactin	$C_{23}H_{34}O_5$	391	4.0	1	5	7	73	28	7
Bestatin	$C_{16}H_{24}N_2O_4$	308	-0.9	4	5	8	113	22	3
Artemisinin	$C_{15}H_{22}O_5$	282	3.0	0	5	0	54	20	7
Forskolin	$C_{22}H_{34}O_7$	411	1.4	3	7	3	113	29	8
Plaunotol	$C_{20}H_{34}O_2$	307	2.8	2	2	11	41	22	0
Thienamycin	$C_{11}H_{16}N_2O_4S$	272	-0.9	3	5	5	104	18	3
SQ26,180	$C_6H_{10}N_2O_6S$	238	-2.2	2	6	3	113	15	1
Spergualin	$C_{17}H_{37}N_7O_4$	404	-1.4	7	9	18	190	28	1
Arglabin	$C_{15}H_{18}O_3$	246	1.4	0	3	0	39	18	5
Average	$C_{15}H_{23}N_2O_5$	319	0.0	3	6	6	109	22	4

NP, natural product; MW, molecular weight; log *P*, *C* log *P*; Hd, H-bond donors; Ha, H-bond acceptors; Rot, number of rotatable bonds; PSA, polar surface area; HA, heavy atom count of nonhydrogen atoms; St, stereogenic centers. Values generated by the PubChem database.

Figure 2



The distribution of molecular weights for the natural products in Tables 3 and 4, indicating a clear bimodal separation. Both the total number and those administered orally are shown.

5 H-bond donors, whereas 10/12 have more than 10 H-bond acceptors. In the AstraZeneca analysis [22•], H-bond donors and PSA were the other parameters besides log *P* that remained constant when comparing recently launched oral drugs against older ones. An important mechanism by which natural products can escape the Lipinski constraints is by judicious positioning of H-bond donors and acceptors to form *intramolecular* H-bonds. With such intramolecular H-boding, there is less of a desolvation penalty and a reduction in PSA, leading to significantly higher permeability than would be predicted. Once at the target, there can be a conformational rearrangement with the bioactive conformation being quite different than that in solution. For medicinal chemists, accurately designing intramolecular H-bonds in a synthetic series is a nontrivial task, and usually rationalized after the fact by experimental observations. Of all the popular metrics, the counting of H-bond donors and acceptors is the least precise. It does not take into account the actual H-bond strength, which is dependent on the functional group and influenced by neighbouring group effects

Natural products go all the way

Among the 24 leads in Table 2, 17 progressed to an approved drug with no modification. Another six natural products were modified by semisynthesis and it is only the β-lactam SQ26,180 that was replaced by a synthetic analog. Contrast this with synthetic compounds, where it would be unprecedented for an initial hit to become the development candidate. Does this mean that natural products are sufficiently evolved for immediate use as a therapeutic agent? Not necessarily, as it is more likely that their complexity deters conventional medicinal chemistry campaigns. Even in the Lipinski universe (Table 3), the majority of compounds contain multiple stereogenic centers and oxygenated functional groups, both of which remain challenges that considerably decrease synthetic accessibility. In the parallel universe (Table 4), these difficulties are compounded further to the extent that they are largely unsurmountable. The best that can be achieved are simple transformations of the natural product itself rather than radical alteration of the scaffold. With synthetic drugs too, reducing the complexity of the initial lead is not an easy task. A survey [26] of drugs launched in 2000 concluded that drugs tend to look like existing drugs, or are closely related in structure to the initial lead. It remains to be seen if the currently fashionable fragment-based [27•] 'bottom-up' approaches are sufficiently general and powerful to avoid this issue.

Ideally, we should be able to identify a minimal pharmacophore for every natural product, with a view to their replacement by semisynthetic or fully synthetic congeners with superior therapeutic properties. In practice, the synthetic bottleneck described above is a sufficient hindrance that it is rarely accomplished. Of the 24 leads in Table 2, only 2 led to wholly synthetic second-generation drugs (Figure 3) in the 1981–2006 timeframe, and both were aided by synthetic tractability. One is compactin, which readily lends itself to simplification as the structure naturally falls into two independent domains linked by a

Figure1 legend continued Structures of natural products discovered since 1970 that led to an approved drug in 1981–2006. For each natural product, the producing organism, mechanism of action, and localization of the target are indicated.

Table 4

The 'parallel universe' of successful natural product leads

NP	Formula	MW	log *P*	Hd	Ha	Rot	PSA	HA	St
Validamycin	$C_{20}H_{35}NO_{13}$	498	-5.2	12	14	7	253	34	14
Midecamycin	$C_{41}H_{67}NO_{15}$	814	2.1	3	16	14	206	57	9
Pseudomonic acid	$C_{26}H_{44}O_9$	501	2.5	4	9	17	146	35	8
Taxol	$C_{47}H_{51}NO_{14}$	854	3.0	4	14	14	221	62	11
Echinocandin B	$C_{52}H_{81}N_7O_{16}$	1060	1.8	14	16	20	368	75	15
Rapamycin	$C_{51}H_{79}NO_{13}$	914	4.3	3	13	6	195	65	15
Cyclosporine A	$C_{62}H_{111}N_{11}O_{12}$	1203	5.2	5	12	15	279	85	12
Lipstatin	$C_{29}H_{49}NO_5$	492	7.5	1	5	21	82	35	5
Avermectin B_{1a}	$C_{48}H_{72}O_{14}$	873	2.3	3	14	8	170	62	20
FK506	$C_{44}H_{69}NO_{12}$	804	3.3	3	12	7	178	57	14
Daptomycin	$C_{72}H_{101}N_{17}O_{26}$	1621	-3.7	22	29	35	702	115	13
Calicheamicin γ_1	$C_{55}H_{74}IN_3O_{21}S_4$	1368	3.2	8	23	24	308	84	19
Average	$C_{46}H_{70}N_4O_{14}$	917	2.2	7	15	16	259	64	13

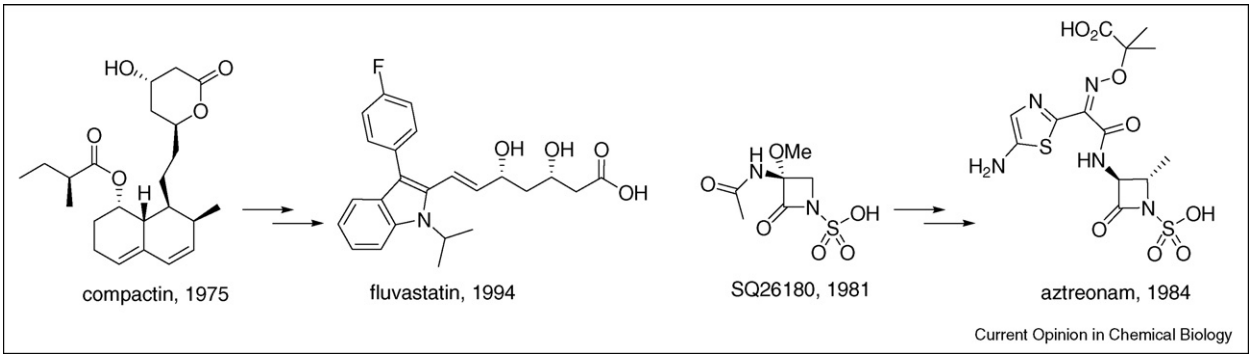
NP, natural product; MW, molecular weight; log *P*, *C* log *P*; Hd, H-bond donors; Ha, H-bond acceptors; Rot, number of rotatable bonds; PSA, polar surface area; HA, heavy atom count of nonhydrogen atoms; St, stereogenic centers. Values generated by the PubChem database.

spacer. The lactone (or more accurately the hydroxy acid to which it is converted *in vivo*) binds to the enzyme active site and once this was realized, the decalin could be replaced by other hydrophobic synthetic fragments. The second success story is the even simpler monobactam SQ26,180 which is lead-like rather than drug-like. Although the methoxy group was a liability *in vivo*, recognition of the key features in the monobactam rapidly led to fully synthetic monobactam sulfonic acids such as aztreonam.

Two personal examples serve to illustrate the challenges and intensity of resources needed to tackle such projects. Pfizer funded a PhD student to undertake the total synthesis of okaramine alkaloids, which have interesting structural features as well as potent anti-insecticidal activity. Our initial target was okaramine C (Figure 4), the simplest in the family with high biological activity, and our modular disconnection was aimed at independently varying each half of the alkaloid to identify structure–activity relationships (SAR). Along the way, an unexpected aza-Claisen rearrangement led to the synthesis [28] of okaramine J instead. Conditions to avoid the rearrangement later enabled the preparation of okaramine C, and it seemed the total syntheses would set the stage for the more exciting exploration of unnatural analogs. However, the syntheses were challenging enough to occupy all of the student's time, and the project was not of sufficiently high priority at Pfizer to be continued. Meanwhile, a proposal to prepare okaramine analogs would also be unlikely to receive favor with funding agencies, as there would be little novel chemistry involved. Concurrent with our work, other groups have reported the total synthesis of okaramines, but none have described biological data or the preparation of analogs. Perhaps, this was not their intention and they were interested only in total synthesis as an end in itself, or they may have faced similar logistic constraints to us. Whatever the case, we have no further information on this fascinating class of alkaloids beyond what was reported in their initial isolation.

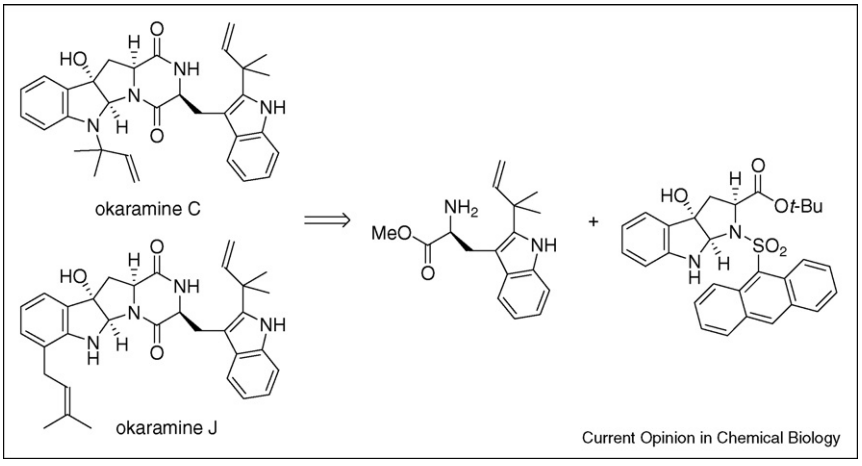
More successful are our ongoing attempts to understand and improve the properties of cyclopeptide and cyclodepsipeptide histone deacetylase (HDAC) inhibitors (Figure 5). We have achieved the total syntheses of spiruchostatin A [29] and azumamide E [30••]. Azumamide E is only the second example of an HDAC inhibitor with a carboxylic acid zinc-binding warhead with submicromolar potency against the enzyme target, while an unnatural hydroxamic acid analog had a nearly 20-fold gain in potency. Although the total syntheses took time (approximately three man-years), the knowledge gained was then readily translatable to analog synthesis. For example, the related HDAC inhibitor FK228 is in advanced Phase II clinical trials as an anticancer agent,

Figure 3



Successful transformation of a natural product lead to a fully synthetic drug.

Figure 4

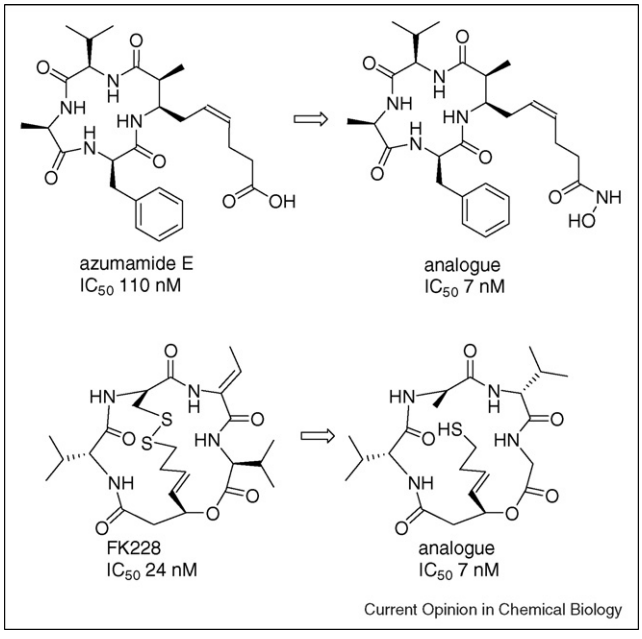


The modular total synthesis of okaramine C and okaramine J from two building blocks.

but no analogs or SAR were known. By total synthesis [31••], we have remedied this situation, and systematically investigated the relative importance of the functional groups embedded within this complex natural product.

The margin between success with the HDAC inhibitors and failure with the okaramines largely comes down to two factors. Firstly, a robust and practical synthetic route is necessary. All too often, the organic synthesis community views natural product targets as a race to be first or an opportunity to showcase methodology rather than design a route suitable for analog generation. Secondly, there needs to be the will to prepare and test analogs, accompanied by ample long-term funding to achieve this in practice, as such programs often consume more than a decade of man-years.

Figure 5



From natural product macrocyclic histone deacetylase (HDAC) inhibitors to unnatural analogs. IC_{50} values against the total HDACs from HeLa cell nuclear extracts are indicated.

Patience is rewarded

The majority of the 24 compounds in Table 1 became the marketed drug without any modification. In other words, the lead that was available on day 1 became the final product. This does not mean, however, that the process is smooth or rapid. From lead disclosure to approval took 10 or more years for 17/24, and of these 4 exceeded the 20-year mark. The last entry in Table 2 dates back to 1988, and no natural product lead isolated thereafter made it to an approved drug by mid-2006. These delays could be because of the need to optimize formulation and ADME/T properties, as well as issues with ensuring natural product supply on a manufacturing scale. Another reason is biology, as many of these natural products were ahead of their time and discovered before their molecular mechanism of action was elucidated and validated as a therapeutic target.

Break the rules

It is obvious that a drug should satisfy an unmet medical need, and this is equally true for synthetics and natural products. A natural product that is 'first-in-class' and novel in its mechanism of action is more likely to be pushed into development, and less likely to have competition from simpler small molecules. Many of the natural products in Table 2 are truly revolutionary modern medicines with a tremendous impact upon healthcare. For example, the

immunosuppressants cyclosporine, rapamycin, and FK506 have made organ transplantation a feasible option while the avermectins have largely eradicated river blindness in Africa. Similarly, the cholesterol-lowering drugs based on inhibition of HMG CoA reductase are widely consumed daily and play an important role in lowering the risk of cardiovascular disease.

To reiterate, the most important criterion for drug development is that the new agent satisfies an unmet need. If it does so, it will be progressed regardless of whether it fits our preconceived notions of what a small molecule orally available agent should look like. Furthermore, it may not fit the usual dogma of reversible interaction with the target, or the complete avoidance of potential toxicophores. For example, Table 2 includes β-lactam antibiotics, arguably the most successful antibacterial agents, which work by an irreversible mechanism. Meanwhile, calicheamicin is an enediyne antibiotic that is extremely cytotoxic but suitable as a therapeutic when targeted to specific cells by conjugation with a monoclonal antibody. The trioxane artemisinin that works by oxidative damage and the membrane pore-forming plaunotol are further examples of irreversible antimicrobial drugs. Finally, the sesquiterpene arglabin contains both an epoxide and an α,β-unsaturated Michael acceptor, two functional groups that would ring warning bells in medicinal chemistry. One further point worth mentioning is the dogma of a well defined molecular target in modern drug discovery. While this is certainly preferable, there are successful drugs in Table 2 for which the mechanism of action remains poorly understood.

Is the glass half empty or half full?

The 35-year period 1970–2006 witnessed 24 natural products leads culminating in an approved drug — is this a reasonable rate of return? The answer will depend on whether one is a proponent of natural products or not. Naysayers will declare that 20-odd drugs is a poor return for the amount of resource expended globally for the past 30 years in natural product drug discovery. The pro-natural products community will point to the same number as a success, as natural product screening delivered nearly one *unique* chemotype per year that was successfully translated to a drug.

Given the long timelines of drug discovery, we will see similar numbers in the near future, as this will reflect natural product leads discovered in the 1990s. For example, since Newman's endpoint of mid-2006, natural products such as ω-conotoxin, ecteinascidin, and epothilone have all led to approved drugs and there are numerous others [32••] in the development pipeline. The continued health of natural product drug discovery in the medium term is more worrying. Owing to the downsizing or abandonment of such efforts in the pharmaceutical industry, natural product screening has been globally reduced in recent years. This irreversible shift in focus can only mean that fewer biologically active natural products will be discovered or serve as the starting point for future medicines. At the same time, industry continues to identify and validate new therapeutic targets. The decline of natural product screening is particularly disheartening as it has successfully served [33••] to identify leads for such novel targets in the past. In the long term, there will hopefully be a renaissance of natural product drug discovery, fuelled by the continued improvements in the process of screening natural product extracts. In addition, the future developments in engineered biosynthesis should enable us to obtain not only natural products but also libraries of 'unnatural' natural products [34•].

These events occur against the backdrop of an evermore challenging landscape for drug discovery. In 2007, the FDA approved [35] only 17 new molecular entities (NMEs). Given that one was radiolabeled ammonia and another hydroxyethyl starch, we can further reduce the number to 15 small molecules, which are predominantly second-generation improvements of earlier drugs. With such pressures, one can only hope that natural products that have served as an important source of drugs in the past will not be overlooked in 21st century drug discovery.

Acknowledgements

The data in Tables 2–4 were first presented at the Zing conference, Small Molecule Drug Discovery: From Early-stage to the Clinic, Antigua, 17–20 January 2008. I am especially grateful to the following for their helpful comments and discussions: Drs Lilian Alcaraz, Andrew Davis and Paul Leeson (all at AstraZeneca), Miles Congreve (Astex), János Fischer (Gedeon Richter), Paul Fish (Pfizer), Chris Larson (Exelixis), Emanuele Perola (Vertex) and John Proudfoot (Boehringer-Ingelheim).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cbpa.2008.03.016.

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

1. Bartlett PA, Entzeroth M (Eds): *Exploiting Chemical Diversity for* •• *Drug Discovery*. RSC Publishing; 2006.
A monograph covering many aspects of molecular diversity and its applications in drug discovery. Of particular relevance are the chapters in Sections 2 and 4.

2. Boldi AM (Ed): *Combinatorial Synthesis of Natural Product-based* •• *Libraries*. CRC Press; 2006.
More specialized than [1••], this timely book concentrates on the fruitful interface between natural products and combinatorial chemistry, with individual chapters written by experts on the field.

3. Grabowski K, Schneider G: **Properties and architecture of** • **natural products revisited**. *Curr Chem Biol* 2007, **1**:115-127.
An updated version of the author's earlier work on molecular architecture.

A collection of synthetic bioactive compounds is compared against collections of pure natural products and semisynthetics. The natural product set was much higher in scaffold diversity, stereochemical features, and oxygen atoms and less likely to contain aromatic fragments.

4. •• Ortholand JY, Ganesan A: **Natural products and combinatorial chemistry — back to the future**. *Curr Opin Chem Biol* 2004, **8**:271-280.
This review describes the historical relationship between combinatorial chemistry and natural products as sources for lead discovery, the reasons why the latter fell out of favor in industry, and closes with probable future developments and strategies for natural product prospecting.

5. Ganesan A: **Natural products as a hunting ground for combinatorial chemistry**. *Curr Opin Biotechnol* 2004, **15**:584-590.

6. •• Chen MS, White MC: **A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis**. *Science* 2007, **318**:783-787.
Laboratory methods for selective C–H oxidation have long been a holy grail for synthetic organic chemistry. Here, White demonstrates reaction conditions that preferentially oxidize tertiary C–H bonds with retention of configuration with relatively high tolerance of other functional groups.

7. • Hartwig JF, Cook KS, Hapke M, Incarvito CD, Fan Y, Webster CE, Hall MB: **Rhodium boryl complexes in the catalytic, terminal functionalization of alkanes**. *J Am Chem Soc* 2005, **127**:2538-2552.
Complementary to [6••], the borylation developed by Hartwig and others achieves the selective C–H activation at the terminal methyl group of alkanes.

8. Russ AP, Lampert S: **The druggable genome: an update**. *Drug Discov Today* 2005, **10**:1607-1610.

9. • Overington JP, Al-Lazikani B, Hopkins AL: **How many drug targets are there?** *Nat Rev Drug Discov* 2006, **5**:993-996.
An update of their earlier paper on estimating the size of the druggable genome.

10. Bohacek RS, McMartin C, Guida WC: **The art and practice of structure-based drug design: a molecular modeling perspective**. *Med Res Rev* 1996, **16**:3-50.

11. •• Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings**. *Adv Drug Deliv Rev* 1997, **23**:3-25.
The seminal paper by Lipinski describing the origins of the Rule of Five and their implementation at Pfizer.

12. •• Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD: **Molecular properties that influence the oral bioavailability of drug candidates**. *J Med Chem* 2002, **45**:2615-2623.
Analysis of over 1100 drug candidates at SmithKlineBeecham, leading to the importance of rotatable bonds and polar surface area as descriptors of drug-like space.

13. Hopkins AL, Groom CR, Alex A: **Ligand efficiency: a useful metric for lead selection**. *Drug Discov Today* 2004, **9**:430-431.

14. •• Abad-Zapatero C, Metz JT: **Ligand efficiency indices as guideposts for drug discovery**. *Drug Discov Today* 2005, **10**:464-469.
An overview of ligand efficiency and its applications from Abbott, providing a fuller account than Hopkins' initial publication [13].

15. • Zhang MQ, Wilkinson B: **Drug discovery beyond the 'rule-of-five'**. *Curr Opin Biotechnol* 2007, **18**:478-488.
A discussion of the important drug-like space that exists outside compounds that comply with all of Lipinski's rules.

16. •• Lipinski CA: **Chris Lipinski discusses life and chemistry after the Rule of Five**. *Drug Discov Today* 2003, **8**:12-16.
A thought-provoking interview with Lipinski where he discusses various aspects of the rules and their exceptions.

17. •• Newman DJ, Cragg GM: **Natural products as sources of new drugs over the last 25 years**. *J Nat Prod* 2007, **70**:461-477.
The most recent of Newman's authoritative surveys, and obligatory reading for anyone interested in the topic.

18. •• Leach AR, Hann MM, Burrows JN, Griffen EJ: **Fragment screening: an introduction**. *Mol Biosyst* 2006, **2**:429-446.
An overview of the concept of lead-like versus drug-like space, followed by a discussion of methodologies used in fragment screening.

19. • Chin YW, Balunas MJ, Chai HB, Kinghorn AD: **Drug discovery from natural sources**. *AAPS J* 2006, **8**:E239-E253.
A survey of drug discovery from natural products, with vignettes of many of the agents discussed in Table 2.

20. •• Gleeson MP: **Generation of a set of simple, interpretable ADMET rules of thumb**. *J Med Chem* 2008, **51**:817-834.
Statistical analysis of ~30 000 GlaxoSmithKline compounds. The conclusion is that increases in log *P* or molecular weight almost always have a negative impact upon ADMET properties.

21. • Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD: **A comparison of physiochemical property profiles of development and marketed oral drugs**. *J Med Chem* 2003, **46**:1250-1256.
An interesting analysis of the similarities and differences between oral compounds that fail in development and marketed drugs.

22. • Leeson PD, Davis AM: **Time-related differences in the physical property profiles of oral drugs**. *J Med Chem* 2004, **47**:6338-6348.
A comparison of oral drugs launched before 1983 versus those from the period 1983–2002. Values for lipophilicity, polar surface area, and H-bond donors remain the same suggesting they are the most important determinants of oral bioavailability.

23. •• Leeson PD, Springthorpe B: **The influence of drug-like concepts on decision-making in medicinal chemistry**. *Nat Rev Drug Discov* 2007, **6**:881-890.
The medicinal chemistry programs at AstraZeneca, GlaxoSmithKline, Merck, and Pfizer are compared and contrasted. Both global trends in medicinal chemistry and those idiosyncratic to company culture are identified.

24. •• Dobson PD, Kell DB: **Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule?** *Nat Rev Drug Discov* 2008, **7**:205-220.
An excellent review of the present understanding of membrane transporters, and the role they play in drug delivery.

25. • Raub TJ: **P-glycoprotein recognition of substrates and circumvention through rational drug design**. *Mol Pharm* 2006, **3**:3-25.
A summary of the recognition features of P-glycoprotein, and examples of medicinal chemistry programs where minimizing efflux was an important objective.

26. Proudfoot JR: **Drugs, leads, and drug-likeness: an analysis of some recently launched drugs**. *Bioorg Med Chem Lett* 2002, **12**:1647-1650.

27. • Erlanson DA: **Fragment-based lead discovery: a chemical update**. *Curr Opin Biotechnol* 2006, **17**:643-652.
A review summarizing recent developments in fragment-based drug discovery.

28. Roe JM, Webster RAB, Ganesan A: **Total synthesis of (+)-okaramine J featuring an exceptionally facile *N*-reverse-prenyl to *C*-prenyl aza-Claisen rearrangement**. *Org Lett* 2003, **5**:2825-2827.

29. Yurek-George A, Habens F, Brimmell M, Packham G, Ganesan A: **Total synthesis of spiruchostatin A, a potent histone deacetylase inhibitor**. *J Am Chem Soc* 2004, **126**: 1030-1031.

30. •• Wen S, Carey KL, Nakao Y, Fusetani N, Packham G, Ganesan A: **Total synthesis of azumamides A and azumamide E, evaluation as histone deacetylase inhibitors, and design of a more potent analogue**. *Org Lett* 2007, **9**:1105-1108.
An example of total synthesis that is a means to an end, and a prelude to biological characterization and the rational design of unnatural analogs.

31. •• Yurek-George A, Cecil A, Mo AHK, Wen S, Rogers H, Habens F, Maeda S, Yoshida M, Packham G, Ganesan A: **The first biologically active synthetic analogues of FK228, the depsipeptide histone deacetylase inhibitor**. *J Med Chem* 2007, **50**:5720-5726.
A nice example of the power of total synthesis. The SAR and conclusions reached in this paper would be impossible to reach by studying the natural product alone.

32. Butler MS: **Natural products to drugs: natural product derived** •• **compounds in clinical trials**. *Nat Prod Rep* 2005, **22**:162-195.
A comprehensive survey of recently approved natural product drugs and those in clinical development.

33. Pucheault M: **Natural products: chemical instruments to** •• **apprehend biological symphony**. *Org Biomol Chem* 2008, **6**:424-432.
Examples of natural products used in target identification and validation, ranging from compounds that are biological tools to approved drugs.

34. Kennedy J: **Mutasynthesis, chemobiosynthesis, and back to** • **semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products**. *Nat Prod Rep* 2008, **25**:25-34.
A review with recent examples of biosynthetic pathway alteration to produce 'unnatural' natural products.

35. Hughes B: **2007 FDA drug approvals: a year of flux**. *Nat Rev Drug Discov* 2008, **7**:107-109.

Databases and internet resources

36. Scifinder Scholar. Comprehensive coverage of all chemical matter, and helpful in tracing the history of particular natural product leads.

37. PubChem, http://pubchem.ncbi.nlm.nih.gov. The NIH's archive of biologically active small molecules, and used for all the metrics in Tables 3 and 4.

38. Wikipedia, http://en.wikipedia.org/wiki/Main_Page. This freely accessible encyclopedia is ever growing, and contains many detailed entries related to drugs and their discovery.

39. DrugBank, http://redpoll.pharmacy.ualberta.ca/drugbank. Although not comprehensive, a useful source of information on specific drugs.

40. FDA, http://www.fda.gov/. A useful repository of information on all drugs approved for use in the USA.