Part One
Natural Products as Sources of Potential Drugs and Systematic Compound Collections
1
Natural Products as Drugs and Leads to Drugs: An Introduction and Perspective as of the End of 2012

David J. Newman and Gordon M. Cragg

1.1
Introduction

Two very frequent comments (together or separately) that have been made, in writing and verbally, over the last 15–20 years can be summarized as follows:

- The use (or pursuit) of natural products as either drugs or as leads to new chemistry that will lead to drugs is now passé, and that what is needed is the use of very high-throughput screens, coupled to large numbers of novel molecules produced by combinatorial chemistry.
- The clever use of computational methods to fit compounds into the active sites of the enzyme (or receptor) of interest will permit the derivation of large numbers of drugs to be discovered and then commercialized rapidly as a result.

We think that perhaps the best answer to comments such as these can be seen in two simple graphical models shown in Figures 1.1 and 1.2. In Figure 1.1, we have plotted the number of small ie meaning up to roughly 45 amino acid residues, with Byetta™ being the upper limit, against the number of “N” and “S” classifications as defined in Ref. [1] from January 1, 1981 through December 31, 2012. In Figure 1.2, we have taken the total number of “N-related” approved drugs over the same time frame as a percentage of the approved drugs for that year. The mean percentage per year of “N-derived drugs” ± the standard deviation over this time frame is 33.4 ± 8.9%, and in 2010, 50% of the 18 approved small-molecule drugs were in this category.

What must be borne in mind is that these are the most conservative figures as we only count a drug once, in the United States if it was first approved by the FDA (Food and Drug Administration) or the approving country’s equivalent of the FDA. Thus, compounds that are subsequently approved for another disease either in the same or in a different country, or whose pharmaceutical properties are extended by slow release or by combination with other agents, are not counted again. There are a few exceptions to this general rule such as the use of nanoparticle-associated

1) The opinions expressed in this chapter are those of authors, and not of the US Government.
albumins in the case of some versions of Taxol\textsuperscript{1} and combinations of different modified insulins, but these, however, account for less than 0.3% of about 1500 compounds (small and large) approved in the last 32 years.

In Figure 1.3, we have shown the breakdown by category, again using the classifications used previously [1] of all drugs and small drugs approved over the last 32 years from January 1, 1981 through December 31, 2012, which should be studied by the interested reader. Again, if one looks at these diagrams, the role of natural product structures as leads (N- and S'-linked materials) is still very significant and even in 2011–2012, 41 of the 62 small-molecule drugs fell into these categories (data not shown but available from the authors on request).

In addition, in Figure 1.4, as befits authors from the US National Cancer Institute (NCI), we have shown the breakdown for all antitumor drugs from the beginning of chemotherapy treatments in the mid-1930s, using variations on the mustard gas used in warfare in World War I, through to the large number of tyrosine protein kinase inhibitors approved in the last few years, with almost all being isosteres of ATP and binding at the ATP site. As already mentioned, in the 2011–2012 N to S' breakdown, 16 of the 18 small-molecule antitumor drugs fell into these classifications. The isostere link was reconfirmed by an excellent presentation given by Fabbro [2] of Novartis at the recent NAD 2012 Meeting in Olomouc, the Czech Republic in July

---

**Figure 1.1** Numbers of natural product-related (N plus S') small molecules per year (1981–2012).

**Figure 1.2** Percentage of natural product related (N) small molecules per year (1981–2012).
2012. Finally in this section, the influence of natural product structures on antitumor agents is such that if one sums the “N-related” then the answer is 89 or 47%, with the “S*-related” equaling 38 or 20% overall. Thus, one can see that natural product-related compounds in this disease category equal 67% of all approved small-molecule drug entities in this time frame. Although not shown, comparable figures are also seen for anti-infective agents over the 32-year time frame covered by Figure 1.3 (we have not gone back to the late 1930s for these data, but may well do so in time).

1.2
The Sponge-Derived Nucleoside Link to Drugs

Until about the 1960s, it was axiomatic that if you wished to make a biologically active nucleoside-derived molecule, you could modify the base including substitutions that
differed entirely from a pyrimidine or purine so that the base could comprise a multiplicity of heterocycles and even carbocycles. However, you had to use either ribose or deoxyribose as the sugar moiety, thus generating large numbers of molecules in pharmaceutical and academic laboratories that met these criteria; none, however, came to fruition as agents, aside from perhaps 5-fluorocytosine, first reported as being synthesized in 1957 [3] and launched as an antifungal agent in 1972.

However, these “conditions” changed as a result of the reports of Bergmann and coworkers [4–6] on the discovery and subsequent identification of spongothymidine (1) and spongouridine (2) in the early 1950s from the Caribbean sponge Tethya crypta as biologically active agents with arabinose instead of ribose or deoxyribose derivatives. These reports led to a complete reversal of the then current dogma
whereby chemists took their substituted bases and initially coupled them to arabinose and then expanded (and contracted!) the sugar moieties to include halogens and other chemical groups (cyano, azido, etc.) and in other examples, reducing the sugar component to 3 carbon units that led to active agents such as acyclovir (3). The subsequent explosion of compounds was described with the relevant citations by Suckling [7] and Newman et al. [8]. These discoveries led to the identification of a close analog, cytosine arabinoside, as a potent antileukemic agent; this compound was subsequently commercialized by Upjohn (then Pharmacia, now Pfizer) as Ara-C (4).

1; Spongthymidine  
2; Spongouridine  
3; Acyclovir

Other closely related compounds such as adenine arabinoside (Ara-A) (5), an antiviral compound synthesized and commercialized by Burroughs Wellcome (now GlaxoSmithKline (GSK)), and later found in the Mediterranean gorgonian, *Eunicella cavolini*, together with spongouridine (AraU) as reported by Cimino et al. [9] in 1984, and even azidothymidine (AZT) (6) can be traced back to this initial discovery of the “other than ribose-substituted bioactive nucleosides.”

4; Ara-C  
5; Ara-A  
6; AZT

In the mid-2000s, two excellent reviews on natural product-sourced nucleosides covering purines [10] and pyrimidines [11] were published in the then new journal *Chemistry & Biodiversity*. In the intervening years, these two interesting papers have been cited over 50 times by multiple investigators. In the purine case, two more recent reviews are worth reading, one on purine modulation of key biological targets by Legraverend and Grierson [12] in 2006 and a later one on privileged structures by Welsch et al. [13] in 2010. In the case of pyrimidines, an interesting example of what is now being produced for biological evaluation is shown in the 2010 paper by Elmarrouni et al. [14] on the synthesis of pyrimidyl α-amino acids. These publications demonstrate that Bergmann and coworkers were definitely prescient in their discoveries.
1.3 Initial Recognition of Microbial Secondary Metabolites as Antibacterial Drugs

If one were asked to name the single microbial-sourced natural product that has saved the most lives, directly or indirectly since its original discovery, there is no doubt that penicillin G (7) would be the molecule of choice. At this time, there are few people in developed countries who can remember the pre-antibiotic age with any clarity. Some, over the age of 75, may have hazy memories of relatives dying at young ages due to bacterial infections, but that is not the norm.

The initial usage of microbial natural products as true antibacterials rather than as surface sterilants was in the later stage of World War II (WWII), roughly June 1944, with the use of secondary metabolites such as penicillin and streptomycin being the examples known in the West. This occurred as a result of the recognition by Fleming in the late 1920s of the activity of penicillin (although there were anecdotal reports of scientists such as Tyndall, Roberts, and Pasteur in the 1870s recognizing antagonism between various bacteria), leading ultimately to the well-known and documented use of penicillins G and V [15] and streptomycin (discovered by Waksman and coworkers) [16] in the early 1940s. However, it also appears that in the same time frame in the former Soviet Union, the antibiotic Gramicidin S (8; Soviet Gramicidin) [17–19] was being used as a treatment for war wounded.

Although the aminoglycosides such as streptomycin, neomycin, and the gentamicins have a long and storied history as treatments for antibacterial infections, particularly in the early days when streptomycin was a treatment for both infected wounds and tuberculosis, there were only a few modifications of the basic molecule(s) that went into clinical use predominately due to the complexity of chemical modification at that time of saccharide-based structures, although in the late 1990s the then Schering-Plough company was working on modifications of the everninomicin complex known as evernimicin or Sch-27899 (9) for the treatment
of resistant *Staphylococcus aureus* strains [20]. This molecule was discontinued for business reasons in late 2000.

![Diagram of Evernimicin (Sch-27899)](image)

Although we could discuss, *ad nauseam*, the countless modifications made to antibiotic classes such as the rifamycins, we will instead show how β-lactams, tetracyclines, macrolides, glycopeptides, lipopeptide, and pleuromutilins, all “ancient antibiotic structures,” are even today being used as base structures upon which to build molecules.

### 1.4 β-Lactams of All Classes

The number of penicillin- and cephalosporin-based molecules produced by semisynthesis and total synthesis to date is well in excess of 30,000. Most started with modification of the fermentation product 6-amino-penicillanic acid (10) or the corresponding cephalosporin 7-aminocephalosporanic acid (11). This number is only approximate as a significant number of structures from industry were never formally published, or were only mentioned in the patent literature. To gain an idea of the multiplicity of these natural product structures that have been reported through 1979, the reader should consult the excellent review from investigators at Fujisawa [21].

In 1948, the ring-expanded version of penicillin, cephalosporin C, was reported from *Cephalosporium* sp. by Brotzu [22–24] with the structure being reported in 1961 by the Oxford group [25,26]. As with the penicillin nucleus, this ring-expanded molecule served as the building block (as 7-aminocephalosporanic acid) for many thousands of cephalosporin structures, with the first orally active molecule cephalexin (12) being introduced in 1970.

![Diagram of 6-Aminopenicillanic Acid, 7-Aminocephalosporanic Acid, and Cephalexin](image)
To extend the “medicinal life” of β-lactams that were substrates for both constitutive and inducible β-lactamases, in the late 1960s and early 1970s, efforts by Beecham (now part of GlaxoSmithKline) and Pfizer found molecules with similar pharmacokinetics to the β-lactams and were inhibitors of the “regular” β-lactamases that were part of the pathogenic microbe’s defense systems. Beecham [27–29] reported the microbial clavulanate family with clavulanic acid (13) being incorporated into the combination known as Augmentin™, a 1:1 mixture of amoxicillin and clavulanic acid launched in 1981. The Pfizer [30] entrant (CP-45,899 or sulbactam (14)) was basically penicillanic acid with a sulfoxide in place of the sulfur. In tazobactam (15), one of the gem methyl groups was replaced by a 1,2,3-triazol-1-yl-methyl substituent by Lederle, now Pfizer [31]. Even today, ~20 years after the last introduction, no other inhibitors have made it to commercialization, although a non-β-lactam β-lactamase covalent slow-released inhibitor known by a variety of names (including avibactam (16), NXL-104, and AVE1330A) [32] as it moved from one company to another, is now in phase III trials with ceftazidime against Gram-negative infections and in phase II with ceftaroline for predominately methicillin-resistant S. aureus infections, both under the aegis of AstraZeneca [33].

Conventionally, clavulanate is normally linked with amoxicillin or ticarcillin, sulbactam with ampicillin, and tazobactam with piperacillin. All of these inhibit only class A serine-based β-lactamases, leaving a significant number of other β-lactamase enzymes where inhibitors are required, including the pharmacologically important zinc-containing β-lactamases [34].

Concomitantly, efforts were underway to obtain the simplest β-lactam, a monobactam (17). Following many years of unsuccessful research at major pharmaceutical houses came the reports from Imada et al. [35] in 1981 and a Squibb group led by Sykes [36] demonstrating the same basic monobactam nucleus. What is important to realize is that no molecules synthesized before the discoveries of these natural products had a sulfonyl group attached to the lactam nitrogen, which is an excellent method for stabilizing the single four-membered ring. Since that time, a significant number of variations have been placed into clinical trials, and one Aztreonam (18) has been introduced to the market. In 2009, the lysinate salt of Aztreonam was launched in the European Union for the
inhalation treatment of *Pseudomonas aeruginosa* in cystic fibrosis under the trade name Cayston\textsuperscript{®}, and in 2010 FDA approval was given for the same indication.

Even in the twenty-first century, these “ancient molecular structures in drug terms” and others discovered after the early 1940s [21] are still valid as scaffolds upon which to base drugs. Perhaps the best way to demonstrate this is to show the data on drugs approved since 2000 that have a \(\beta\)-lactam in their structure. Since 2000, four synthetic penems known as biapenem (19; 2002), ertapenem (20; 2002), doripenem (21; 2005), and tebipenem (22; 2009), which were based upon the structure of the natural product thienamycin (23), reported in 1978 [37], have been approved.
Three cephalosporins – one, cefovecin (24; 2006), which was a veterinary drug, and two human-use drugs, ceftobiprole medocaril (25; 2008, but withdrawn in 2010, although still in advanced trials for other indications) and ceftaroline fosamil acetate (26; 2011), which was launched in the United States for treatment of MRSA – have also been approved by the relevant authorities for use as drugs.

1.5 Tetracycline Derivatives

The structures, basic chemistry, structure–activity relationships, clinical microbiology, and resistant phenotypes of the first (Achromycin®, Aureomycin®, and Terramycin®) and second generation (Minocin®) are given with extensive commentary in the excellent 2001 review by Chopra and Roberts [38], which should be read by the interested reader. As already mentioned, the result of clinical reports of the recognition of the evolution of tetracycline resistance in *Shigella dysenteriae* in 1953 and of a multiresistant *Shigella* in 1955 [39], by classical and the later use of molecular genetics approaches, led to the recognition of the multiple tetracycline efflux pumps and of protective ribosomal mechanisms, discussed in detail in Ref. [38]; the suggestive evidence of the monophyletic origin of these genes plus the potential for cross-contamination from animal sources was covered in 2002 by Aminov et al. [40].

Following on the major resistance problems with the first- and second-generation tetracyclines, a series of synthetic and semisynthetic modifications of the base pharmacophore were made with special emphasis on position 9 of the base molecule. Although prior attempts to modify at this position led to molecules with
poor antibacterial activities, scientists at the then Lederle Laboratories (then Wyeth, now Pfizer) discovered that 9-acylamido derivatives of minocycline (Minocin) had activities comparable to first- and second-generation molecules, but did not have activity against resistant organisms [41]. Following these initial discoveries came a publication in 1999 on the synthesis of GAR-936 [42], a glycol derivative of a modified doxycycline molecule, now known as tigecycline (27), which had broad-spectrum activity including both Gram-positive and Gram-negative bacteria and MRSA, and was approved in 2005 by the FDA. Thus, by utilizing what are effectively relatively simple chemical modifications to an old molecule, these base structures can have a new lease on life and provide activity against clinically important infections [43,44].

Even today, 64 years after the original reports by Duggar [45], this class of antibiotics is generating significant interest both chemically and biologically. Knowledge from genetic analyses of tetracycline biosynthesis in bacteria, coupled with the advances in the biosynthetic processes as reported by Pickens and Tang [46] in 2009, bode well for the future of this old compound class.

![Tigecycline](image)

27; Tigecycline

### 1.6 Glycopeptide Antibacterials

Vancomycin (28), the initial member of the glycopeptide class of antibiotics, was first approved in 1955, and is still the prototype for variations around the same mechanism of action, namely, the binding to the terminal L-Lys–D-Ala–D-Ala tripeptide when the Gram-positive cell wall is undergoing extension during growth. The compounds discussed in this section are semisynthetic modifications of the same basic structural class, thus following in the “chemical footsteps” of the β-lactams and the tetracyclines discussed previously and the macrolides discussed in a later section.

By December 2012, there were a number of such molecules either approved or in clinical trials. Televancin (29) was approved in 2009 in the United States and then approved for a different indication in the European Union in 2011. In addition, two more semisynthetic glycopeptides, oritavancin (30) and dalbavancin (31), are in phase III trials. In all cases, as with vancomycin, their antibacterial mechanism is via inhibition of cell wall production, although the exact mechanisms can vary with the individual agent. In the case of oritavancin, it would appear that the agent is
comparable to vancomycin in its inhibition of transglycosylation, but more effective as a transpeptidation inhibitor [47]. As noted earlier, all are semisynthetic derivatives of natural products, with oritavancin [48] being a modified chloroeremomycin (a vancomycin analog), dalbavancin [49] being based on the teicoplanin relative, B0-A40926, and telavancin (TD-6424) being directly based on a chemical modification of vancomycin [50].
Theravance (also the originator of telavancin) has successfully combined a cephalosporin with vancomycin to produce TD-1792 (32), which is currently in phase II trials against complicated skin and soft tissue infections in human patients [51]. Thus, combining two old antibiotic classes can produce novel agents,
again underscoring the possibilities of reworking older structures if one understands their history.

1.7 Lipopeptide Antibacterials

Although vancomycin has activity against lipid II, with the onset of the VanR phenotype in pathogenic bacteria, microbial metabolites that had languished for a number of years were reinvestigated. The first example was ramoplanin (33), a lipopeptide antibiotic complex isolated from Actinoplanes sp. ATCC33076, consisting of factors A1, A2 (the major component), and A3 [52,53]. This mixture exerted its antibacterial activity by binding to the peptidoglycan intermediate lipid II (C₃₅−MurNac−peptide−GlcNac) and thus disrupting bacterial cell wall synthesis [54–56]. At the time of writing, this mixture was in phase III clinical trials.

Another older cyclic lipopeptide that had moved around from one company to another, starting at Lilly, moving to the then Lederle (then Wyeth, now Pfizer), and finally developed by Cubist as a new antibiotic against MRSA is daptomycin (34; launched 2003), a member of a large class of complex cyclic peptides with variations in the peptidic components and the acylating fatty acids. These included the mixtures identified as the daptomycin/A21978 complex, the A54145 complex,
the CDA complex, the friulimicins/amphomycins, and the laspartomycin/glycinocins, whose base structures, biosyntheses, and potential for genetic manipulation were discussed in detail in 2005 by Baltz et al. [57] from Cubist. Further examples on the potential for modifications were published from 2006 to late 2008 demonstrating the potential for such “combinatorial biochemistry” to produce complex structures with modified activities [58–60]. The potential was realized by the entry of a modified daptomycin known by the name surotomycin (35) into clinical trials by Cubist. The molecule has the same cyclic peptide moiety as daptomycin but a changed lipid tail and it is currently in phase III trials with an emphasis upon the treatment of Clostridium difficile-associated diarrhea [61–63].

33; Ramoplanin A2

34; Daptomycin
If one follows novel modifications of old structures that bind to ribosomes and therefore inhibit protein synthesis [64], from 2000 there have been three molecules formally known as “ketolides” that are based on the erythromycin chemotype that were either approved or entered advanced clinical trials in this time frame. Telithromycin (36) was approved in 2001 and was later found to be both a substrate and an inhibitor of cytochrome P450 3A (CYP3A4) [65]. Two others either entered or are in phase III trials. Thus, cethromycin (ABT-773; 37) entered clinical trials and was in phase III under the Chicago-based company Advanced Life Sciences (ALS), with good activity against respiratory infections [66] and also showed in vivo activity against plague (Yersinia pestis) in rats [67]. However, since ALS ceased operations in mid-2011, the current status of this compound is unknown. In contrast, the product of glyco-optimization, now known as solithromycin (CEM-101; 38), quoted as the “most potent macrolide-based antibiotic known” [68] with excellent activity against plasmodium [69] and Neisseria gonorrhoeae isolates [70], has just moved into phase III clinical trials against community-acquired pneumonia (CAP).

However, not all modifications of older structures succeed, as was demonstrated by the discontinuation in 2010 of the interesting modification of the base erythromycin structure, the “bicyclolide” known as modithromycin (39) (also known as EDP-420, EP-013420, and S-013420). This compound, a novel, bridged bicyclic derivative originally designed by Enanta Pharmaceuticals [71,72], was in phase II trials for treatment of CAP by both Enanta and Shionogi before discontinuation.
Demonstrating again that older structures with antibiotic activity have significant validity for today’s diseases (even though it was for an old disease common in the early 1940s in the early days of antibiotics), in 2007 GSK received approval for a
modified pleuromutilin, retapamulin (40), for the treatment of impetigo in pediatric patients [73]. The base structure, pleuromutilin (41), dates from a report in 1951 of its isolation from the basidiomycete *Pleurotus mutilus* (FR.) Sacc. and *Pleurotus passeckerianus* Pilat [74]. In the mid-1970s, a significant amount of work was reported on the use of derivatives of pleuromutilin as veterinary use antibiotics [75], including approval of valnemulin (42) in 1999 under the trade name of Econor® by Sandoz. The use of the base molecule as a source of human-use antibiotics is reminiscent of the work that led to the approval of Synercid® in the late 1990s, as the synergistic molecules that led to that mixture were extensively used in veterinary applications, predominately in the alteration of metabolism in ruminants.

A number of human-use antibiotics based on this elderly structure in addition to retapamulin are currently in clinical trials. Thus, Nabriva (an Austrian company) signed a codevelopment agreement with Forest Laboratories in the United States in 2012 for the phase II development of BC-3781 (43) as both an oral and IV therapy against MRSA and other resistant Gram-positive organisms [76]. Based on the same structure and also from Nabriva, two other agents BC-3205 (44) [77] and BC-7013 (45) are currently in phase I clinical trials, with the latter being developed as a topical agent.
Privileged Structures

One can claim that secondary metabolites – that is, those compounds produced by an organism, usually in response to a stimulus of some type, that are not required for the basic life of the organism – are “privileged structures.” This term was first defined by Evans et al. [78,79] when the Merck group in the United States was discussing the biological activities of synthetic benzodiazepines based on known anxiolytic structures as potential cholecystokinin antagonists. The influence of this term/concept can be seen by almost 700 citations to Evans’s original paper listed in a search in late 2012.

Of the most recent papers that cited Evans, three are of interest. Two use the privileged structure concept as defined by Evans [80,81] but with some variations. The third, from Ganesan’s group [82], which will be discussed later in the chapter, has an interesting “twist” on the privileged structure concept that can perhaps be best described by quoting part of their introduction: “A scaffold that leads to biologically active compounds will attract interest by medicinal chemists who will then produce more examples of the same and discover new active compounds that further confirm the hypothesis. There should then exist examples of “underprivileged scaffolds” that are intrinsically suitable for drug discovery applications but in practice are underrepresented or absent.”

The Origin of the Benzodiazepines

A seven-membered di-aza ring was reported in the late 1800s as a potential dyestuff under the name “benzheptodiazine”; 4,5-benzo-[hept-1,2,6-oxdiazine] in the German literature. Sternbach remembered this work from his years working in Poland before WWII while at Roche, and revised the structure to be a quinazoline 3-oxide (not the previously reported seven-membered ring system). However, the irony in the story is that Sternbach et al. [83–85] used the initial but incorrect structure as the basis for his syntheses of the extremely well-known psychoactive drugs, Librium® (46) and Valium® (47), by building upon the simple 1934 syntheses of the benzodiazepines from o-phenylene diamine and benzaldehyde [86].

However, it was not for almost 40 years after the original synthesis of the benzodiazepines in 1934 that in 1971 the same basic pharmacophore was identified in natural product molecules – the tomatymycins (base structure; 48). This identification was followed with a publication the following year [87]. A much earlier example of what might be termed as a “temporal disconnect” between synthesis and identification as a natural product was the case of histamine whose synthesis was reported in the late 1880s, well prior to its discovery in the bloodstream.
Since Evans’ introduction of this concept, major chemistry groups interested in natural product syntheses and structural modification have shown the potential of these natural product-derived scaffolds, and we will cover three basic examples plus one novel extension of the biosynthetic process. These in publication order are from Nicolaou’s group covering benzopyran skeletons derived nominally from plants, from Waldmann’s group with molecules derived from the marine-sourced metabolite dysidiolide (with a segue into Quinn’s biosynthetic analyses; the novel extension), and finally, the recent norbenzomorphan work by Sahn and Martin.

The brief section on the work from Waldmann’s group (Section 1.13) should be read in conjunction with Chapter 2, which covers the potential of these methods in much greater detail.

1.12
Benzopyrans: A Source of Unusual Antibacterial and Other Agents

In the case of benzopyrans, the natural product literature yielded nearly 4000 analogs, and then if one included a slight structural modification, another 8000 structures were identified. In the late 1990s, using these skeletons, Nicolaou’s group [88–90] successfully used the combinatorial concept of “structures from structures” to produce iterative derivatives, ending up with relatively simple molecules based on benzopyran (49) and dihydrobenzopyran (50) skeletons, which led to the identification and subsequent optimization of benzopyrans with a cyanostilbene substitution (51) that were effective against vancomycin-resistant bacteria, a structural class that did not have any antibiotic activities reported prior to these publications.

Quite recently, a more extensive report covering roughly 12 years since the original series was published by Lee and Gong [91]. This demonstrated the power of modifying these basic skeletons and the addition of other substituents including fused carbocycles and heterocycles. These papers
demonstrate how simple modifications of such privileged structures can lead to novel potential agents.

\[
\begin{align*}
49; \text{Benzopyran} & & 50; \text{Dihydropyran} & & 51; \text{Cyanostilbenes}
\end{align*}
\]

1.13 Multiple Enzymatic Inhibitors from Relatively Simple Natural Product Secondary Metabolites

Waldmann and his group [92,93] explored an interesting concept for the design of combinatorial libraries based on natural products, initially reported in a series of papers commencing in 2002. The guiding principles were derived from recognition of the fundamental and complementary properties of natural products and their protein targets. The overall idea can be described in the following manner.

Nature, as a result of the evolution of natural products, has explored only a small fraction of the available “small-molecule chemical space,” and the same holds true for the biological targets of natural products, which are mainly proteins. Due to the fact that topologically similar shapes (i.e., the outer surfaces) can result from different underlying amino acid sequences, the number and topology of three-dimensional protein folds have been shown to be even more conserved during evolution than the underlying sequences. Although estimates of the number of proteins in humans range between 100,000 and 450,000, the number of topologically different protein folds is actually much lower, with estimates ranging between 600 and 8000 [94]. Since natural product space and protein structure space explored by Nature are limited in size and highly conserved, these structure spaces have to be highly complementary.

As a result of this conservation and complementarity, if a natural product is described as a competitive inhibitor of a specific protein/fold (i.e., it binds at the active site of the enzyme), then it represents what may be considered as a biologically validated starting point for the subsequent chemical development of closely related structures. These “derived” structures may inhibit proteins with similar folds, and perhaps allow for the discovery of specificity. This idea was formalized by the Waldmann group [95–97] under the acronym PSSC or “protein structure similarity clustering.” In this paradigm, proteins are clustered by their three-dimensional shape (surface topology) around the ligand binding sites, regardless of sequence similarity. This concept is fundamentally similar to the privileged structure concept [78,79], but PSSC
has the extra dimension of using protein folding patterns (surface topology) as the basis for subsequent screens.

The second concept was that the base scaffold of natural products can be mapped in a hierarchical manner thus creating a scaffold tree and was given the acronym SCONP or “structural classification of natural products” [98,99]. This concept permitted the derivation of logical pathways for the structural simplification of scaffolds. Merging of both of these concepts then led to the BIOS (biology-oriented synthesis) approach [100]. Thus, the ligand of any member of a PSSC could be expected to exhibit some degree of complementarity toward other members of the PSSC and hence serve as a starting point for the development of modulators of the other members.

The initial success of what ultimately came to be known as the “BIOS approach” was demonstrated by a combinatorial library inspired by the marine natural product dysidiolide (52). By postulating that the γ-hydroxy-butenolide group of dysidiolide was the major determinant of phosphatase activity, testing of a 147-member library built around this molecule yielded a compound (53) that was 10-fold more potent (IC_{50} = 350 nM) than the parent compound against Cdc25A [101]. What was very significant in this work was that other members of the library were identified with low micromolar activities against the enzymes acetylcholinesterase and 11β-hydroxysteroid dehydrogenase type 1, which fall within the same PSSC as Cdc25A [102], whereas from classical enzymology, none of these other enzymes would have been considered to be inhibited by a Cdc25A inhibitor.

Very significant efforts have been made in the past 20 plus years with respect to the discovery and development of novel kinase inhibitors, using the term kinase in its enzymatic sense – a phosphorylator of hydroxyl groups. This was frequently through the “design of structures that resemble purines and/or ATP itself and will bind at ATP-binding sites,” an approach that has been quite successful at producing structures for clinical trials [103–105]. In an alternative approach, which did not formally concentrate on the specifics of the ATP-binding site, the Waldman group successfully used BIOS to search for kinase inhibitors.

The marine sponge-derived metabolite nakijiquinone C (54), first reported by Kobayashi et al. [106] in 1995, was shown to be an inhibitor of epidermal growth factor receptor (EGFR), c-ErbB2, and protein kinase C (PKC), in addition to having cytotoxic activity against L1210 and KB cell lines. Using this compound as the starting structure, a library of 74 compounds was constructed around the basic nakijiquinone C structure by the Waldmann group [107] and tested against a battery of kinases with similar protein domain folds. These compounds yielded seven new inhibitors with low micromolar activity in vitro, including one VEGFR-2 inhibitor (55) and four inhibitors of Tie-2 kinase (56–59), a protein intimately involved in angiogenesis and for which, at the beginning of the study, no inhibitors were known. However, during the study, the first natural product inhibitor of Tie-2 kinase (60) was reported [108] from the plant Acacia aulacocarpa, with a set of four papers from another research group demonstrating the activity of synthetic pyrrolo [2,3-d]pyrimidines (61) as inhibitors of the same class of kinases [109–112].
Quite recently, details of the evolution and utility of this approach as an integrated program were given in two reviews by the Waldmann group [113,114], and very recently, an extension demonstrating the use of “fragment-based ligand
discovery” from natural product-derived fragments was published by the same group [115]. All of these should be consulted for the specific details of the processes involved, in particular the latest one in *Nature Chemistry* [115].

1.14
**A Variation on BIOS: The “Inside—Out” Approach**

In the mid-2000s, Quinn’s group in Australia was considering secondary metabolite biosynthetic processes, specifically the production of flavonoids in plant systems that also had potential as kinase inhibitors. Quinn *et al.* [116] considered that the active site of the last enzyme in the biosynthetic cascade (if the structure was known or could be modeled) would share a common protein fold topology (PFT) with the target (active site) of the compound produced. The concept was extended further in a later paper from the same group [117] covering a different set of biosynthetic metabolites.

Effectively, Waldmann’s BIOS approach comes from the “outside” (protein folds but from the surface) to the active site, whereas Quinn’s approach considers that the “active site of the target” is effectively the mirror image of the active site of the last biosynthetic enzyme. Thus, these concepts are complementary, not competitors.

1.15
**Other Privileged Structures**

If one studies the naturally occurring azanaphthalene scaffolds (i.e., the quinolones and isoquinolines), then their influence as pharmacophores would be very significant when one looks at the number of bioactive compounds, both drugs and candidates as shown in the recent review by Polanski *et al.* [80]. An interesting aspect of their paper is that they did not consider topological mimics of natural product structures such as ATP. Although they discuss bis-azanaphthalene structures and show some of the compounds currently under clinical trials as potential kinase inhibitors, the concept of an NP-mimic is not addressed.

Of the current approved kinase inhibitors, the majority act as direct competitive inhibitors of ATP, but this type of interaction does not show up in a regular computerized analysis of structural motifs. Similarly, peptide isosteres such as the angiotensin receptor 1 antagonists, the sartans, or the HIV protease inhibitors (the vast majority of which are isosteres of the natural hexapeptide substrate) do not show up in such analyses.

The recent paper by Sahn and Martin [81], however, demonstrates what can be done if like Nicolaou *et al.* [88–90], one starts with a known series of bioactive agents, in this particular case, the morphine (62) alkaloids, which are now known to be peptide isosteres of the endorphins, the endogenous substrates for the opioid receptors in man. By taking the base tricyclic structure of the benzomorphan (63)
and removing one carbon in the central ring system, a [6.5.6] tricyclic motif with one nitrogen atom was generated (the norbenzomorphans). Further modification led to compounds such as 64, which exhibited activity as an acetylcholinesterase inhibitor (AChE inhibitor) as active but less toxic than (−)-physostigmine (65). By utilizing substituted benzaldehydes as the starting materials, a 124-member library was constructed that is currently being tested in a variety of biological screens with current activities ranging from an inhibitor of the topoisomerase I of Y. pestis to an antagonist of the human M1 muscarinic receptor. What other biological activities will be found are yet to be revealed.

1.16 Privileged Structures as Inhibitors of Protein–Protein Interactions

A further extension of both the BIOS and PFT concepts is implied in a recent review on the use of secondary structure information in drug design by Koch [118]. In this review, Koch demonstrates that the concepts can be extended to interactions at protein–protein contact positions that are termed “hot spots” [119,120]. These contact interfaces are approximately 1200–2000 Å² in area, and as in the examples described earlier with BIOS and PFT, not all of the interface residues are of equal importance. Hot spots appear on average to comprise ~10% of interfacial residues and overlap with conserved regions on the surface of the proteins, with complementarity in “pockets” on either side of dimeric interfaces [119,120].
Wells and McClendon [120] described a number of “synthetic small molecules” that successfully interacted with IL2, BCL-XL, HDM2, HPV E2, ZipA, and TNF with affinities comparable to or greater than the natural partners. In one case, the molecule was based upon a familiar scaffold, that of the benzodiazepines, where the structure is known to be a mimic of a β-turn [121], with a derivative called benzodiazepinedione (66). A second example is shown in the recent report on the activity of thio-benzodiazepines (67, 68) as nanomolar-level inhibitors of the p53-MDM2 protein–protein interaction [122]. Other relatively simple structures such as the terphenyl (69) moiety mimic an α-helix [123]. There are computerized tools that can help in the prediction of “turn structures” from sequence data in the absence of a crystal structure, thus perhaps permitting analyses of a significant number of proteins from this aspect [124].

![Molecule 66: Benzodiazepinedione](image)

**66; Benzodiazepinedione**

![Molecule 67: Ki = 91 nM p53-MDM2](image)

**67; Ki = 91 nM p53-MDM2**

![Molecule 68: Ki = 89 nM p53-MDM2](image)

**68; Ki = 89 nM p53-MDM2**

![Molecule 69: o-Terphenyl](image)

**69. o-Terphenyl**

However, for an excellent example of where one base molecule from microbial sources has become a “poster child” for protein–protein interactions, one does not have to look any further than the story of the molecules related to rapamycin. There are currently six molecules including rapamycin (70; sirolimus) that are in clinical use. The other five are everolimus (71), zotarolimus (72), temsirolimus (73), biolimus A9 (74), and novolimus (75). The last two are components of stents and are not used as isolated drug entities. There is also one other in the series that is currently in phase III clinical trials as an antitumor agent, deforolimus (76), though it has had a fairly checkered career to date as to clinical trials and lack of approval. The
genesis of most of these agents has been given in many articles and need not be repeated here, although the data up through 2007 were given in a 2008 perspective by Newman [125]. As a very current example of how these agents are being used in cancer treatment, the meta-analysis by Dittmer et al. [126] covering 150 clinical trials registered with the NCI covering only viral cancers should be consulted. In particular, their Figure 2 is illustrative of the range of these agents.

It should also be noted that all of the agents except one differ only at one position on the large macrolide ring (the C43 position) from the original rapamycin molecule, thus demonstrating that very small changes in the overall “shape” of the molecules cause quite different effects [126]. Readers might also wish to consult another review that demonstrates the value of these agents against all cancers rather than the subset used by Dittmer et al. To this end, the 2012 review by Pópulo et al. [127] covers a broader range of cancers and should be read in conjunction with the more restrictive one, in order to gain a slightly different perspective.

What was also of interest was the approval of the novolimus-containing biodegradable stent in the European Union in late 2012. This is the only molecule in the post-rapamycin (rapalog) series that does not have a modification at the C43 locus. It is in fact a metabolite of rapamycin (sirolimus) where the methoxy group at C16 has been demethylated to the alcohol.

Rapamycin and all of the other rapalogs bind at the interface of the proteins mTOR (mammalian target of rapamycin) and FKBP12 (FK binding protein 12). mTOR is a serine-threonine kinase and is homologous to phosphatidylinositide 3-kinase (PI3K) with a formal sequence similarity of >30%; however, one needs to take into account the caveats under BIOS and PFT with respect to similarities, so the actual resemblance may be a lot higher.

On binding the “rapalogs” to FKBP12, the resulting complex then binds to and inhibits the protein kinase activity of mTOR. Thus, rapamycin and its analogs are formal protein kinase inhibitors but in an “indirect fashion.” With this information, Tanneeru and Guruprasad [128], using the crystal structure of PI3K and molecular dynamic (MD) modifications, were able to derive a model of the human mTOR kinase domain, and then model in 27 ATP-competitive inhibitors (structures in references 18–20 in their review) to derive fundamental data for the design of other mTOR inhibitors. Further discussion on the utility of MD calculations in this type of work was recently presented by Caballero and Alzate-Morales [129], whose review should be consulted for further information.

The potential of mTOR inhibitors, and by extension inhibitors of the pathways that this kinase leads into, has recently been discussed in reasonable detail by Gentzler et al. [130], and their review should be consulted for further information. Another example of the influence that this series of molecules has had on the scientific literature can be seen from almost 2400 references found as of November 2012 when searching the Scopus database using just the phrase “rapamycin binding to mTOR.”
Ganesan’s group [82] at the University of East Anglia explored the potential of the well-known class of natural product-based molecules, the diketopiperazines (77; DKPs) that are very easily synthesized from dipeptides. The biological activities of
what might be considered to be “regular DKPs” are well publicized, covering a wide variety of drug targets [131,132], although as might be expected, synthetic and medicinal chemists have synthesized large numbers of nitrogen-based heterocyclic compounds such as the DKPs, even though analyses of natural product sources 13 years apart showed that in 1999, oxygen-related heterocycles predominated [133] and these findings were still as valid in 2012 [134].

If one now considers “underrepresented scaffolds,” in 2009, chemists at UCB-Celltech [135] in the United Kingdom identified approximately 25,000 small aromatic ring systems (mono and bicyclic rings with five or six atoms in the ring(s)). They limited the atoms to C, H, N, O, and S, and all putative structures had to obey Hückel’s aromaticity rules. As of that date, less than 1800 had been reported in the literature, following searches of research papers and patents. Thus, there are very significant numbers of “not yet represented” scaffolds open for synthesis and/or discovery.

The Ganesan group [82] therefore elected to investigate a simple modification of the “normal” DKP structure where a nitrogen atom would replace a ring carbon atom in the basic diaza-dione system (77), thus generating a triazadione (78) analog of the basic DKP structure. Following some excellent chemistry using solid-phase combinatorial methodologies, they reported synthesizing 32 examples, using as the starting materials variations on regular amino acids, variations on aldehydes, and in particular, a propargyl derivative that hopefully may well be amenable to “click chemistry” linkages with potential targets. To date, no biological activities related to these compounds have yet been published, but with the previous record of DKPs we consider that it is only a matter of time before biologically active compounds from this or a similar series will be identified.

1.18 So Where Should One Look in the Twenty-First Century for Novel Structures from Natural Sources?

Our suggestion may seem unusual to scientists who have spent their professional lives performing medicinal and natural product chemistry around structures isolated from plants, marine organisms, and terrestrial microbes, but we consider that it is at the interface of microbial interactions with their hosts and commensals where novel agents can be found. As can be seen from the previous sections, each
one of the earlier sources have proven to be excellent reservoirs of novel structures that have produced a multitude of drug candidates against a large number of disease entities.

However, what has become quite apparent from genomic work on the total sequences of free-living microbes of all Kingdoms is that we have barely scratched the surface of potential biosynthetic mechanisms in single-celled organisms. From analyses of the then relatively few published genomes of actinomycetes, it was becoming obvious in the early 2000s (and from the work of companies such as Ecopia in Canada) that each of the bacteria that was studied contained multiple potential biosynthetic clusters (the so-called cryptic clusters) with the implied potential to produce previously unknown molecules if they could be activated. Over the next few years, as the cost of genome sequencing decreased dramatically (the <$2500 US sequence is effectively here at the time of writing), massive amounts of data have been placed into open databases by groups such as the Department of Energy’s Genome Sequencing laboratory in the United States, or from a significant number of academic groups in the Americas and Europe; the stage is set for identification and hopefully expression of these biosynthetic clusters.

One probable reason why these clusters have not been recognized previously was that researchers concentrated on the use of “pure” single-celled organisms for their fermentation experiments, whereas in Nature these organisms exist in consortia and do “talk among themselves using chemical cues.” One should not be surprised by such findings as within each “biological niche,” chemical cues had been recognized in the past. One can think of quorum sensing agents in bacteria, or mating factors in sexual forms of fungi, or “elicitors” in plants, or pheromones in insects and even humans. What was missing was the recognition that these organisms “talk to dissimilars” as well as to their compatriots within a single group of organisms. Thus, there is evidence that the human microbiome (and bear in mind that humans are roughly 90% microbe and 10% mammalian on a cell number basis) can mediate the health of their human host, if one can say that 10% is the “host.” The 2012 review by Cho and Blaser [136] makes extremely interesting reading for people without a background in this field.

Such host–microbe interactions can include, but not be limited to,

- other microbes, as in the case of the rhizoxins [137],
- interplay between insects and microbes [138,139], and
- mining the massive numbers of what are known as “cryptic clusters” in bacteria and fungi [140–142].

As an example of the possibilities in the last suggestion above, if one looks at the number of putative secondary metabolite clusters in the published DNA sequences of just nine Aspergillus species, there are between 33 and 79 putative clusters covering most of the potential biosynthetic routes but not including terpenes, identified to date, as given in Table 1 in Ref. [142]. There are techniques already in the literature for the identification and expression of such clusters, so the
methodology is already available, but it needs to be used on a significant scale in order to obtain the maximum benefit [143–145].

In a similar fashion, the number of “cryptic clusters” in marine-sourced actinomycetes has been determined in some specific cases and the compounds encoded have been expressed and their structure determined. This is shown in great detail in the work around the genetics of the Salinispora species that produce salinosporamide A and a plethora of unrelated compounds. The recent reviews from the Moore, Jensen, and Fenical groups [143,146,147] at the Scripps Institution of Oceanography (University of California, San Diego) should be read to know the wealth of opportunities that are present but not yet realized by most investigators.

1.19 Conclusions

We hope that in this relatively short chapter we have been able to demonstrate that even in the second decade of the twenty-first century, natural product-based structures are still alive and acting both as drugs in their own right and as leads from which to generate novel agents of utility against the manifold diseases of man.

What is also of perhaps even more import is that the linkage of genomics, computerized genome mining, and perhaps variations on combinatorial chemistry to be used in the optimization of novel structures from Nature will enable scientists (biologists of all “types” and chemists) to unlock the enormous potential of materials produced from the interaction of entirely different domains of life, let alone across Kingdoms. The biosynthetic processes for novel structural classes are there, we just have to learn how to switch them on in a productive manner.

References


48 Cooper, R.D., Snyder, N.J., Zweifel, M.J., Staszak, M.A., Wilkie, S.C., Nicas, T.L.,


References


