The combinatorial chemistry of nature

Gregory L. Verdine

Powerful new technologies such as high-throughput screening and combinatorial chemistry are revolutionizing drug discovery. But natural products still offer unmatched structural variety, especially as new environmental niches are explored, and their usefulness can be further extended by engineering the proteins that produce them and by using them to probe biological pathways.

NATURE is unrivalled in its ability to craft small organic molecules jam-packed with structural complexity and biological potency. In the early 1800s, an emerging fascination with such molecules, later known as 'natural products' or secondary metabolites, gave rise to the field of organic chemistry, so named for its emphasis on the chemistry of living things. Natural products have since provided a focal point for wide-ranging efforts aimed at elucidating the structure, reactivity, covalent bonding and noncovalent interactions of carbon-containing compounds. But it is through their pivotal role in the process of drug discovery that natural products have so profoundly altered the course of human history. In the twentieth century alone, drugs that trace their heritage to secondary metabolites have more than doubled the average lifespan of human beings. Furthermore, they have been able to ameliorate maladies that are not life-threatening but are nonetheless debilitating, such as chronic pain and depression, thus offering new life to those who just decades ago would have been condemned to a life of inexorable grief or madness. Over the past decade, natural products that became drugs have revolutionized medicine by making it possible to replace damaged organs by transplantation.

This is an impressive success record, and one might think that those who wish to discover new drugs need look no further than the nearest microorganism. But the impressive potency of biosynthetically derived compounds must be balanced against their sometimes overwhelming structural complexity, which can seriously hamper efforts to turn the natural product lead molecule into a drug. Furthermore, the pressure on the pharmaceutical industry to cut research and development costs while increasing the speed with which it brings new drugs to market has been steadily rising in recent years[1]. In response, powerful new technologies are washing over the industry in waves, bearing names straight out of a venture-capital prospectus: high-throughput screening[2], structure-based drug design[3], combinatorial chemistry[4]. Together, these new ideas seem certain to revolutionize drug discovery as we know it. What then is the role of the venerable natural product in this brave new world? Here I discuss several ongoing initiatives in the field of natural products science, and their interrelationship with drug discovery as it is likely to be practised in the future.

Exotic chemicals from unexplored niches

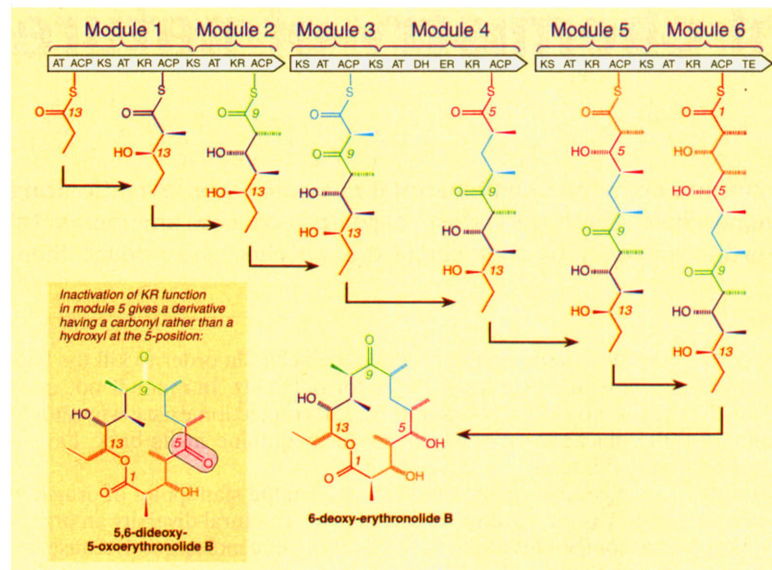
Organisms have evolved the capacity to produce secondary metabolites in response to the specific needs and challenges of their local environments. The process of evolving a natural product entails the random recombination and mutation of existing genetic material to generate new biosynthetic enzymes that catalyse the assembly of new organic compounds. When these compounds serve a useful purpose to the producer organism, the genes programmed to carry out their biosynthesis are retained and perhaps fine-tuned through further genetic changes. Thus the credit for having invented combinatorial chemistry of molecules both large and small clearly rests with nature, although mere mortals deserve an honourable mention for having become such worthy practitioners of the art. Sometimes a secondary metabolite serves the same purpose to the producer organism as that served to man who uses it pharmacologically; for example, the *Penicillium* mould produces penicillin in order to kill the bacteria with which it competes for nutrients. In many if not most cases, no such obvious teleological connection exists: why should the cinchona tree, which produces quinine in its bark, have a vested interest in curing malaria?

Of course, what matters most from the standpoint of drug discovery is the novelty and richness of structural diversity an organism contributes to the universe of organic molecules. Because life forms that are closely related genetically tend to produce similar secondary metabolites, the broadest representation of structures requires the sampling of organisms across the most diverse spectrum of environments. On the basis of this principle, earnest efforts are underway to collect life forms from various niche environments that have been overlooked in the past: coral reefs, deep-sea hydrothermal vents and tropical rainforests, for example. This is a race against time; because of human encroachment and pollution, among other factors, such niche environments and the natural products they contain are rapidly disappearing from the globe[5]. In an extraordinary move that recognizes the potential consequences of such devastation for the future of the pharmaceutical industry, the drug company Merck recently signed a multimillion dollar agreement with the government of Costa Rica to protect its tropical rainforests and screen them broadly for new drug leads.

Beating nature at its own game

The recent application of genetic approaches towards the study of biosynthesis has created a new and exciting avenue for the discovery of new bioactive molecules. Natural products are made by multienzyme systems in which each enzyme carries out one of the many transformations required to make the final product from simple building blocks such as acetate. The genes encoding these enzymes are often clustered into so-called 'biosynthetic operons', self-contained cassettes of genetic material bearing the blueprint for building a natural product. This notion of a genetic blueprint is especially appropriate for the case of macrolide polyketide synthases (PKSs), gargantuan enzymes that assemble acetate and propionate units into a vast array of complex organic molecules. Remarkably, the sequential order of enzymatic domains in PKS genes corresponds precisely to the order in which these domains process the growing polyketide chain, such that one can divine the structure of the final polyketide product based on the DNA sequence of its biosynthetic operon[6] (see figure). This direct relationship suggested that it might be possible to engineer 'designer' natural products through rational manipulation of biosynthetic operons. The concept of biosynthetic engineering was first demonstrated for the erythronolide biosynthetic operon, in which mutational inactivation of a single enzymatic module led to the biosynthesis of an erythromycin analogue having a keto group in place of the normal hydroxyl substituent at the 5-position (ref. 2). Even more strikingly, enzymatic domains can be added to[7] or repositioned within[8] the erythronolide biosynthetic operon, thereby reprogramming polyketide biosynthesis through the addition of further processing steps.

These results are important on several levels. Many natural products are themselves unsuitable for use as drugs, instead providing a lead structure for a drug discovery effort. Converting a



Modular organization of the biosynthetic operon for 6-deoxy-erythronolide B in *Streptomyces erythraea* (adapted from ref. 6). Each of the three tandemly arranged genes in the operon contains two modules, for a total of six. A module encodes a series of linked enzymatic domains that together catalyse the addition and subsequent processing of a single propionate unit: ACP, acyl-carrier protein, the domain in each module that carries the growing polyketide chain bound to a pantotheinyl prosthetic group as a thioester; KS, the ketosynthase subunit that catalyses the elongation of the polyketide chain by a single propionate unit; AT, acyltransferase, which loads malonate onto ACP, to prepare it for participation in chain elongation; KR, ketoreductase, a domain that catalyses the NADPH-dependent reduction of the newly generated keto unit to an alcohol; DH, dehydratase, catalyses the elimination of water from the β-hydroxy thioester to form an α,β-unsaturated thioester; ER, enoyl reductase, which is responsible for catalysing the NADPH-dependent reduction of the α,β-unsaturated thioester to a saturated thioester; and TE, thiol esterase, a domain that promotes the final cyclization of the linear thioester chain to a cyclic ester macrolide. The modules are ordered sequentially in the operon such that the structure of the natural product (neglecting stereochemistry) can be divined from the types of enzyme domains present in each module. Functional disruption of the KR domain in module 5 prevents co-synthetic conversion of the C5 carbonyl to a hydroxyl group, thereby leading to the production of a 'designed' natural product analogue, 5,6-dideoxy-5-oxoerythronolide B.

natural product lead structure into a drug requires chemical modification of the lead, which is often difficult to perform by either total chemical synthesis or synthetic manipulation of the natural product ('semisynthesis'). An obvious and time-honoured alternative is to use the biosynthetic machinery of the producer organism to generate new analogues for testing and semisynthesis. For decades, microbiologists have used methods such as random mutagenesis in an attempt to coerce producer strains into generating natural product analogues. This random approach has turned out to be both labour-intensive and capricious; perhaps more importantly, it generates only loss-of-function mutations and does not allow the structure of the analogue to be predetermined. Biosynthetic engineering can in principle overcome these limitations. Of course, the advantages of biosynthetic engineering must be weighed against the investment in cloning and DNA sequencing that it requires. This investment may be substantial: the biosynthetic genes for the immunosuppressive agent rapamycin occupy a roughly 100,000 base-pair region of the *Streptomyces hygroscopicus* chromosome, on which are encoded an estimated 27 polypeptides[9]. Furthermore, biosynthetic engineering can only be used in those cases for which an expression system is available or the producer organism is efficient at homologous recombination.

The successful transplantation of new functional units into PKS gene modules has raised the exciting possibility that entirely new classes of 'non-natural' polyketides might be generated through what has been termed a 'combinatorial biosynthesis' strategy[10] of mixing, matching and shuffling PKS gene segments. Indeed, it has already been shown that fragmentation[11] of PKS genes, or their recombination[12] or transformation into a heterologous host[13], can direct the biosynthesis of new polyketides. On the basis of these examples and the brisk pace of research in this burgeoning new field, it seems likely that the future will bring a rich and structurally diverse bounty of previously unknown polyketides. Because many successful drugs have come from this class of molecules — tetracycline, erythromycin and Adriamycin, to name a few — it seems likely that others will emerge from the ranks of the new polyketides. What remains to be seen is whether biosynthetic pathways for structural families other than PKSs, for example isoprenoids, will be as amenable to manipulation.

A parallel and complementary effort to *in vivo* biosynthetic engineering is *in vitro* multistep enzymatic synthesis. In brief, this strategy involves overexpressing individual biosynthetic genes in *Escherichia coli*, purifying the encoded enzymes, then mixing them to create a cocktail of enzymatic catalysts. These cocktails are 'fed' simple biosynthetic precursors, which they process according to their substrate specificity. The ease with which such *in vitro* biosynthetic reactions assemble complex organic molecules can be truly remarkable: in one case, a cocktail of 12 enzymes converted 5-aminolevulinic acid to hydrogenobyrininc acid, an advanced intermediate along the vitamin-B_{12} biosynthetic pathway, giving a 20% overall yield[14]. In a matter of hours, this bioreactor system was able to catalyse a 17-step conversion with an average stepwise yield of 90%. Such is the stuff of dreams and future promise for those who make natural products through total chemical synthesis[15,16]. As recently as ten years ago, the notion of overproducing a dozen enzymes would have been greeted with derision, but recent advances in polymerase chain reaction-assisted overproduction systems[17,18] and affinity tagging technology[19–21] have greatly reduced the investment of time and labour necessary to obtain pure proteins in quantity. Thus, the day seems close at hand when combinatorial libraries of diverse organic compounds will be produced through *in vitro* biosynthesis using factorial enzyme arrays with a variety of natural and non-natural building blocks.

Natural products as pathfinders

Just as the technology for presenting chemical diversity to a macromolecule is rapidly expanding, so is the universe of potential targets. By the end of the decade, large-scale DNA sequencing efforts will have identified a large proportion of the unique complementary DNAs encoded by the human genome, and powerful technology for analysis of expression patterns will fingerprint cDNAs according to phenotypic criteria such as the cell type and developmental stage in which they are expressed. With this embarrassment of riches comes the problem of deciding which of the vast number of potential macromolecular targets are worth pursuing.

The early hope that targeted gene disruption ('knockout') technology in mice might provide a universal and definitive means of validating potential drug targets seems to be waning; developmental compensation all too often obscures the expected effect of removing a target molecule from differentiated cells. An alternative and widely practised strategy is to use high-throughput *in vitro* screening to discover a molecule that will potently inhibit a specific target of interest, followed by evaluation of the molecule in animal models. This approach suffers from the problem that it requires a substantial investment of time and resources (and animals) into a target whose importance is, by definition, poorly understood. If only we could narrow down the focus to a small number of 'hardened' targets — those known to

be crucial in the disease process — a greater proportion of precious resources could then be put to productive use.

An altogether different way to look at the problem is to focus on finding and elucidating important pathways, then selecting one or more targets from the macromolecules that lie along it. This very strategy has been tremendously successful in the development of drugs that lower cholesterol by inhibiting its *de novo* biosynthesis. How, then, to discover new pathways and the potential targets that make them up? This could be accomplished by high-throughput screening of whole cells against libraries using 'smart' assays tailored to produce a very specific readout, such as downregulation of the Ras signalling pathway for potential cancer therapeutics. Only time will tell whether the majority of the useful hits are derived from combinatorial or natural product libraries, so the prudent course of action would be to screen both.

After having identified a potently active and selective molecule, the next step is to identify its target. A highly efficient strategy has been developed for this purpose. The first stage consists of modifying the structure of the molecule at various sites, then carrying out cell-based assays on the derivatives to determine those sites at which modification does not lead to a loss of biological activity. This perturbational information is used to guide the construction of a ligand-specific affinity chromatography resin, which selectively extracts the target from cellular extracts. In what is now a classic example, this approach was used to identify calcineurin as the target of the immunosuppressant drugs cyclosporin A and FK506 (ref. 22). Although calcineurin itself has fallen into disfavour as a drug target, other proteins that lie along the same signal transduction pathway, such as ZAP-70 (the protein kinase specific to T cells) are of great current interest. An analogous approach has resulted in the discovery of *inter alia* human FRAP as the target for rapamycin[23], the proteasome as the target for the neuritogenic agent lactacystin[24], and histone deacetylase as the target for the antimitotic agent trapoxin[25].

Future strategies and prospects

Now, as in the foreseeable future, small-molecule drug discovery efforts will rely heavily on high-throughput screening of libraries. Combinatorial libraries will always be designed with ease and efficiency of synthesis in mind, so they are likely to contain relatively simple structures. But organisms do not seem to pay much attention to synthetic expediency, hence they contribute relatively exotic structures to the universe of organic compounds. For example, natural products are more likely to be rich in stereochemistry, concatenated rings and reactive functional groups than the structures in combinatorial libraries. Notwithstanding the many virtues of simplicity, complex problems can demand complex solutions, and a 'baroque' natural product lead having nanomolar affinity will always arouse more excitement than a micromolar combinatorial lead.

Several additional points are worth keeping in mind. The pharmaceutical industry, emboldened by its fantastic success at discovering drugs, has become ever more ambitious in the types of targets it pursues. Whereas two decades ago the majority of targets were enzymes, the ongoing revolution in molecular and cell biology has filled development portfolios with fascinating exotica such as cell-surface receptors and their ligands, intracellular signalling molecules, and even nuclear targets such as transcription factors and their recognition sites. For the most part, these are uncharted waters, and navigating them will require the use of every means available. Basic science may provide some important clues and encouragement: just when everyone was beginning to think that protein–protein interfaces may be too large and diffuse to yield to antagonism by a small molecule, it was discovered that only a few clustered residues of the human growth hormone and its receptor contribute most of the binding energy[26].

Second, even though a natural product may itself be unsuitable for use as a drug, it may provide invaluable structural clues that can form the basis for a development effort. The very recent advances in image-plate detection, high-flux synchrotron X-ray beam sources and cryocooling technology have reduced the timescale of crystallographic analysis such that structure can now be used prospectively rather than retrospectively, as in the past. How better to shave down a complex natural product to its bare essentials than to begin with a structure of the small molecule bound to its macromolecular target? Furthermore, a structural motif of a natural product can sometimes be used as a 'biasing element' around which to build a combinatorial library[27]. The natural product statine has recently been used as a key structural motif in a family of HIV protease inhibitors[28]; statine could be used as a biasing element for a library of protease inhibitors. A library of prospective protein kinase inhibitors might be built using the adenine moiety of ATP. Surely the impetus to develop libraries dedicated through structural biasing at certain protein families will only increase as we move to the post-genomics era, when the crucial issue will be the function rather than the structure of genes.

Finally, even though natural products will always be harder than combinatorial products to optimize, the strategies for systematically altering natural product structures discussed here should greatly help the process of converting a natural product lead into a drug. The advances are much needed; chemical synthesis of natural products has made 'heroic' advances over the past half-century, but the problems it faces are sufficiently challenging that many complex natural products still cannot be made into economically feasible drugs.

In summary, dramatic advances in cell culture and extraction techniques, high-throughput screening, genetics and protein biochemistry, structural biology and synthetic chemistry have converged to create a bright future at the crossroads of natural products science and drug discovery. □

Gregory L. Verdine is at the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

1. Vagelos, P. R. *Science* **252**, 1080–1084 (1991).
2. Broach, J. R. & Thorner, J. *Nature* **384** Suppl., 14–16 (1996).
3. Blundell, T. L. *Nature* **384** Suppl., 23–26 (1996).
4. Hogan, J. *Nature* **384** Suppl., 17–19 (1996).
5. Lee, J. C., Yang, X., Schwartz, M., Strobel, G. & Clardy, J. *Chem. Biol.* **2**, 721–727 (1995).
6. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. *Science* **252**, 675–679 (1991).
7. Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M. & Katz, L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7119–7123 (1993).
8. Cortes, J. *et al. Science* **268**, 1487–1489 (1995).
9. Schwecke, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 7839–7843 (1995).
10. Tsoi, J. & Khosla, C. *Chem. Biol.* **2**, 355–362 (1995).
11. Kao, C. M., Luo, G., Katz, L., Cane, D. E. & Khosla, C. *J. Am. chem. Soc.* **116**, 11612–11613 (1994).
12. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *Nature* **375**, 549–554 (1995).
13. Hopwood, D. A. *et al. Nature* **314**, 642–644 (1985).
14. Roessner, C. A. *et al. Chem. Biol.* **1**, 119–124 (1994).
15. Corey, E. J. & Cheng, X.-M. *The Logic of Chemical Synthesis* (Wiley, New York, 1989).
16. Nicolaou, K. C. & Sorensen, E. J. *Classics in Total Synthesis* (VCH, New York, 1995).
17. MacFerrin, K. D., Terranova, M. P., Schreiber, S. L. & Verdine, G. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1937–1941 (1990).
18. MacFerrin, K. D., Chen, L., Terranova, M. P., Schreiber, S. L. & Verdine, G. L. *Meth. Enzym.* **217**, 79–102 (1993).
19. Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. & Stüber, D. *Bio/technology* **6**, 1321–1325 (1988).
20. Smith, M. C., Furman, T. C., Ingolia, T. D. & Pidgeon, C. *J. biol. Chem.* **263**, 7211–7215 (1988).
21. Arnold, F. H. *Bio/technology* **9**, 151–156 (1991).
22. Liu, J. *et al. Cell* **66**, 1–9 (1991).
23. Brown, E. J. *et al. Nature* **369**, 756–758 (1994).
24. Fenteany, G. *et al. Science* **268**, 726–729 (1995).
25. Taunton, J., Hassig, C. A. & Schreiber, S. L. *Science* **272**, 408–411 (1996).
26. Clackson, T. & Wells, J. A. *Science* **267**, 383–386 (1995).
27. Chen, J. K., Lane, W. S., Brauer, A., Tanaka, A. & Schreiber, S. L. *J. Am. chem. Soc.* **115**, 12591–12592 (1993).
28. Wlodawer, A. & Erickson, J. W. *A. Rev. Biochem.* **62**, 543–585 (1993).

ACKNOWLEDGEMENTS. The keen insights of R. Ward greatly improved this article. I thank the NIH and NSF for ongoing research support in the area of natural products science. The generous support of these organizations, the Hoffmann-La Roche Institute of Chemistry in Medicine and Ariad Pharmaceuticals for research in the author's laboratory is also gratefully acknowledged.