JOC
The Journal of Organic Chemistry

VOLUME 65, NUMBER 24 DECEMBER 1, 2000

Perspective

Natural Product Biosynthesis: A New Interface between Enzymology and Medicine

Chaitan Khosla

Departments of Chemistry, Chemical Engineering and Biochemistry, Stanford University, Stanford, California 94305-5025

ck@chemeng.stanford.edu

Received June 2, 2000

Introduction

Given the importance of enzymes in human health and disease, the study of enzyme mechanisms has historically provided a rich interface between organic chemistry and drug discovery. Many of the fundamental questions in enzymology focus around the structural and mechanistic basis for two particular properties of these macromolecules: their exquisite selectivity toward cognate substrates and the extraordinary rate enhancements achieved in their catalytic chemistry. These problems are a natural extension of the principles of physical organic chemistry into biology; this is reflected by the fact that many specific questions in enzymology could be posed, and many enzyme mechanisms could be predicted with remarkable accuracy even before structural biology and protein engineering made their mark on the field within the past two decades.[1,2] Indeed, perhaps the finest testament to the enormous contributions of chemistry to the process of drug discovery is reflected by the observation that, in this post-genomic era, the molecular targets most frequently chosen for therapeutic intervention continue to be members of well-studied enzyme families such as proteases, kinases, phosphatases, and oxidoreductases, where mechanistic detail can rapidly be translated into potent and selective small-molecule inhibitors.

In addition to being targets for medical intervention, the above-mentioned properties of enzymes have also sparked the interest of organic chemists in a different context: their potential use as catalysts for regio- and stereocontrolled synthesis. Here, the goal is to exploit the remarkable ability of enzymes to achieve huge rate accelerations at room temperature and in aqueous environments for the transformation of cheap substrates into value-added products. These transformations can involve the action of a single enzyme (usually as an isolated protein) or multiple enzymes (typically a fermentation process involving intact cells). Notable advances have been made in this area with selected enzymes (e.g., esterases and oxidoreductases) and for certain categories of products (e.g., bioactive microbial products such as vitamins and antibiotics).[3] However, there remain major limitations to the more extensive use of enzymes in synthetic chemistry. Perhaps the most notable barrier is their inherent selectivity against unnatural substrates, which adversely affects the volumetric productivity, and even the yield (especially in cases where cellular catalysts are used) of enzyme-catalyzed processes. As a result, synthetic applications of biocatalysis have primarily been

* To whom correspondence should be addressed. Phone/Fax: 650-723-6538.

(1) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover: Mineola, NY, 1969.

(2) Walsh, C. T. *Enzymatic Reaction Mechanisms*; W. H. Freeman: New York, 1979.

(3) Wong, C. H.; Whitesides, G. M. *Enzymes in Synthetic Organic Chemistry*; Pergamon: Elmsford, NY, 1994.

10.1021/jo000849y CCC: $19.00
Published on Web 09/20/2000

successful in the context of structurally complex, high-value materials such as medicinally relevant natural products.

This paper provides a personal perspective of the interplay between natural product biosynthesis and the development of natural product-based drugs. I will start with an overview of the process of natural product drug discovery, and the problems it has encountered in recent times. I will then briefly review chemical and biological aspects of polyketide natural product biosynthesis and recap the events that drew me to this most rewarding field of research. The remainder of this paper will be focused on the interplay between chemistry and biology to understand polyketide synthase structure and mechanism and the application of this basic knowledge to solve the major problems associated with natural product drug development. Toward the end, I will conclude with some thoughts about the future of the interface between natural product biosynthesis and drug development.

Natural Product Drug Discovery and Development

Drug discovery in the 21st century is commonly associated with dazzling new concepts and technologies including functional genomics, massively parallel combinatorial chemistry, ultrahigh throughput screening, and structure-based ligand design. In stark contrast, the pathway for developing new natural product drugs has remained relatively unchanged over the past several decades. New natural products have been discovered by screening microbial (or occasionally plant) extracts against appropriate molecular or cellular targets. Identification of an active extract is followed by a deconvolution process that typically requires extensive purification, re-testing, and spectroscopic analysis, and can take several months. Once an active substance has been isolated and characterized, on rare occasions it has adequate potency, selectivity, pharmacokinetics, and proprietary status so as to enable it to enter clinical studies without further modifications. More common are situations where the natural product serves as a lead for further development. Microbiologists are recruited to ensure a reliable supply of the natural product, whereas medicinal chemists are responsible for converting the lead into a clinical candidate. As a necessary prerequisite for establishing quantitative structure−activity relationships, the first task of the chemist is to develop practical chemistries for selectively modifying the polyfunctional organic molecule at different positions. Through a cumulative effort that typically involves tens of years, the project generates a few hundred analogues of the parent natural product, at which point either a clinical candidate emerges, or the project is abandoned altogether.

The impact of natural products on drug discovery has been enormous: in a recent report by the National Cancer Instiute, Cragg and co-workers estimated that approximately 40% of the new drugs approved in the period 1983−1994 owed their origins to natural product leads.[4] This number would have been considerably higher if one were to focus the analysis on certain therapeutic areas such as infectious diseases and oncology, or if one went back in time to include "the Golden Era of Antibiotics" in the survey. Notwithstanding this enviable track record, however, natural product programs in industry appear to be in search of a new identity. While several explanations can be offered for this turn-around, they all appear to boil down to a simple economic reality. Given the large investment required for natural products drug discovery, the marginal returns have lately been disappointing: out of a few thousand new antibiotics that have been discovered since 1975, fewer than 10 have been commercialized.

What can be the reasons for this gradual decline of natural products in drug discovery? Certainly it is not because the flow of new natural products into the pipeline has decreased. On the contrary, the rate of discovery of new bioactive microbial products has consistently increased since World War II. Two major hurdles are encountered in the development of a newly discovered natural product lead. First, its structural complexity is often in stark contrast to synthetic lead molecules. In an era of rapidly shrinking time scales for lead optimization, this often results in up-front elimination of the natural product lead from further consideration. The ability to rapidly derivatize natural products represents a major challenge at the interface of chemistry and biology. The second hurdle in natural product drug development is the extraordinarily high cost of producing the molecule. This presents a challenge both for the medicinal chemist, who is now presented with the added difficulty of performing challenging chemistry on a small amount of material, and for the pharmacologist or toxicologist, who seeks to study the properties of the natural product (and semisynthetic derivatives) in assays that require exponentially greater quantities of material. In the initial stages following the isolation of a novel natural product, the cost of production typically exceeds $1000 per gram of purified material. As both the producing strain and the production process are further developed and scaled up, material costs decrease but often remain greater than $10000/kg at the time when the product reaches the market. Several decades of further strain and process improvement can bring the cost of goods down further to under $100/kg, but even this is still significantly more expensive than synthetic substances. The identification of rapid and reliable ways to overproduce natural products has major implications for the future of natural products drug development.

Work in this laboratory has been motivated by the goal of developing fundamental approaches for solving the two problems presented above (especially the former) in the context of polyketide natural products. Therefore, before proceeding further, it would be appropriate to introduce this family of microbial natural products to the reader not familiar with them.

Polyketide Natural Products and Their Biosynthesis

Polyketide natural products are known to possess a wealth of pharmacologically important activities, including antibacterial, antifungal, antiparasitic, antitumor, and immunosuppressive properties. Notwithstanding their remarkable structural variety (Figure 1), the biosynthesis of these natural products bears close mechanistic relationships. Polyketide synthases (PKSs) are large (M_r 100−10000 kDa) multienzyme systems that are responsible for the stepwise biosynthesis of the carbon

(4) Cragg, G. M.; Newman, D. J.; Snader, K. M. *J. Nat. Prod.* **1997**, *60*, 52−60.

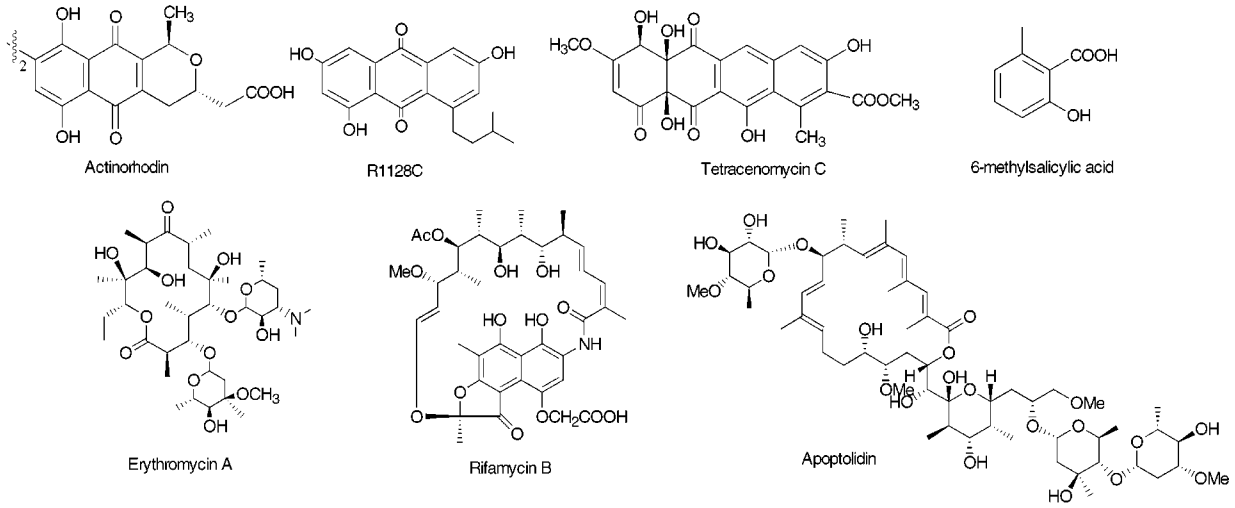

Figure 1. Examples of polyketides that have been studied in this laboratory.

chain backbones of polyketides from simple two-, three-, and four-carbon building blocks such as acetyl-CoA, propionyl-CoA, and butyryl-CoA and their activated derivatives malonyl-, methylmalonyl-, and ethylmalonyl-CoA. The key chain-building step of polyketide biosynthesis is a decarboxylative condensation analogous to the chain elongation step of classical fatty acid biosynthesis, and indeed polyketide synthases and fatty acid synthases show remarkable genetic, protein structural, and mechanistic similarities. Unlike fatty acid biosynthesis, however, in which each successive chain elongation step is followed by a fixed sequence of ketoreduction, dehydration, and enoyl reduction, the individual chain elongation intermediates of polyketide biosynthesis undergo all, some, or none of these functional group modifications, resulting in a striking level of chemical complexity in the products. Additional degrees of complexity arise from the use of different starter and chain elongation units, the generation of new stereocenters, and variations in cyclization of the carbon chain backbone. The subject has been extensively reviewed in the recent literature, and will not be discussed in detail here.[5−11] It is however worth noting that, although chemical approaches toward understanding polyketide biosynthesis have evolved over more than half a century, biology has only made its impact on this field of research within the past 10−15 years. As is becoming increasingly common these days, the most exciting and innovative work in biosynthesis is being performed at the interface of these two disciplines.

I was drawn to the study of polyketide biosynthesis by three factors. The first was the recognition that, although the questions associated with polyketide biosynthesis were quintessentially chemical in nature, biological tools would be crucial for obtaining satisfactory answers. Therefore, this research area offered me the opportunity to stay close to my intellectual roots, while at the same time reaching out into the fascinating world of biology. The second factor was the firm (but not very well-thought out) belief that polyketide synthases would not only represent a fertile ground for uncovering new principles in enzymology, but also for translating them into practical opportunities for enzyme engineering. The timing of my early biochemical education closely paralleled the dawn of protein engineering, and it was stimulating to see the power of this technology being applied to the decoding of protein chemical mechanisms, yet frustrating to witness the difficulties associated with the practical engineering of proteins. The third factor was a chance meeting with two individuals, who have subsequently had a huge impact on my research. The first was David Hopwood, Professor of Genetics at the John Innes Centre in the U.K., whose contributions to the field of actinomyces genetics are unparalleled. (The actinomyces are a family of bacteria that produce more than half of all known bioactive microbial products.) The second individual was David Cane, Professor of Chemistry at Brown University, who has made seminal contributions to our understanding of mechanistic aspects of polyketide and terpene biosynthesis, and who arrived on sabbatical in the Hopwood laboratory at the same time as I arrived there as a postdoctoral fellow. It did not take long to recognize that both of them were exceptionally talented and accomplished individuals with razor-sharp clarity of thought in their own fields of expertise and (more intriguingly) an unusually deep appreciation for the challenges and opportunities in each other's fields of research. Having only a limited backgound in natural product biosynthesis, I was indeed fortunate to be able to tap into such an immense wealth of knowledge of both genetic and chemical aspects of the field at such an early stage in my career. My scientific association with them has continued over the years and has taught me much about the power of collaborative science in a highly interdisciplinary area of research.

The Impact of Genetics on the Study of Polyketide Biosynthesis

The impact of genetics on the study of polyketide biosynthesis was initially felt through the cloning and nucleotide sequence analysis of biosynthetic genes. In

(5) Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet.* **1990**, *24*, 37−66.
(6) O'Hagan, D. *The Polyketide Metabolites*; Ellis Horwood: Chichester, U.K., 1991.
(7) Katz, L.; Donadio, S. *Annu. Rev. Microbiol.* **1993**, *47*, 875−912.
(8) Hutchinson, C. R.; Fujii, I. *Annu. Rev. Microbiol* **1995**, *49*, 201−238.
(9) Staunton, J.; Wilkinson, B. *Chem. Rev.* **1997**, *97*, 2611−2630.
(10) Cane, D. E.; Walsh, C. T.; Khosla, C. *Science* **1998**, *282*, 63−68.
(11) Khosla, C.; Gokhale, R.; Jacobsen, J. R.; Cane, D. E. *Annu. Rev. Biochem.* **1999**, *68*, 219−253.

1984, Malpartida and Hopwood reported the cloning of the entire gene set required for biosynthesis of the benzoisochromanequinone antibiotic, actinorhodin.[12] Remarkably all the genes (ca. 25) involved in actinorhodin biosynthesis, regulation, and self-resistance were clustered within the genome of the producing bacterium *Streptomyces coelicolor*. Since then, the genes encoding the biosynthesis of numerous bacterial and fungal natural products have been cloned and sequenced. In all cases the biosynthetic, regulatory, and resistance genes are tightly linked.[13] This monumental discovery has contributed significantly to the emergence of a new genetics-led paradigm for understanding and manipulating natural product biosynthesis in microorganisms. A variety of genetic strategies has been developed over the years to take advantage of nature's benevolence in order to clone complete biosynthetic gene clusters of interest. Frequently used strategies include complementation of blocked mutants,[14] transfer of partial or complete pathways in a surrogate host,[12] homology-based gene isolation,[15] identification of resistance gene(s) through selection in a heterologous host,[16] and reverse genetics based on limited amino acid sequence of a purified pathway enzyme.[17]

In addition to facilitating the cloning of new PKS genes, genetics has also contributed to our understanding of polyketide biosynthesis by facilitating the construction and analysis of recombinant PKS gene clusters. Originally, the strategy of choice for this purpose involved utilization of homologous recombination to replace or delete individual base pairs or even entire genetic segments in the chromosome of the native polyketide producer. All other native genes required for natural product biosynthesis, regulation, and precursor formation remain intact in the recombinant host. As a consequence, the reporter polyketide product is produced in the natural intracellular environment and frequently undergoes some or all of the normal post-PKS transformations that are typically associated with polyketide natural product biosynthesis. For example, the early studies of Hopwood and co-workers on actinorhodin biosynthesis[18] and of Katz and co-workers on erythromycin biosynthesis[19] involved such methodologies. Homologous recombination is a particularly attractive strategy for manipulating extremely complex PKSs encoded by large gene clusters; however, it is often technically difficult and relatively slow and, therefore, places serious constraints on the systems that can be studied and the types of experiments that can be performed on a PKS of interest.

An alternative strategy involves heterologous expression of selected PKS genes in suitable hosts. Upon arrival at Stanford, the development of generally applicable methods for this purpose was the principal focus of our efforts. Again, the history of molecular biology was an important motivating force. By facilitating the production of reagent quantities of virtually any polypeptide of interest in simple and genetics-friendly hosts such as *E. coli*, biology provided a huge impetus to the study of protein chemistry in the past quarter century. My limited exposure to the genetics of a variety of natural product producing organisms during my years in the Hopwood laboratory convinced me that, for nonspecialists such as myself to harness genetic tools for dissecting and manipulating polyketide biosynthesis, the development of versatile heterologous expression systems would be important. Moreover, thanks to David Hopwood's influence, the choice of a heterologous host was straightforward: *Streptomyces coelicolor*.

Streptomyces coeliccolor A3(2) is a model actinomyces with well-developed genetic tools and, more recently, an 8 Mbp genome whose sequence is nearing completion. It is known to synthesize at least five structurally distinct natural products including actinorhodin, a polyketide spore pigment of unknown structure, and CDA (a nonribosomal peptide antibiotic of unknown structure). Thus, it has a solid track record for expressing PKSs and their close cousins, the nonribosomal peptide synthetases (NRPSs).[20] Actinorhodin is perhaps one of the most well-studied polyketide biosynthetic pathways, especially at the genetic level. Back in 1992, more was known about the genetics of actinorhodin biosynthesis and its regulation than about any other polyketide natural product. We therefore argued that development of a heterologous system in this host might be accomplished by a two-step procedure involving (a) surgical deletion of the entire actinorhodin gene cluster from its genome and (b) development of a plasmid-borne expression system that incorporates the requisite genetic tools (promoters, activator) from the deleted gene cluster. The resulting bifunctional actinomyces-*E. coli* vector, pRM5,[21] has successfully been used to functionally express the actinorhodin,[21] granaticin,[21] tetracenomycin,[21] frenolicin,[22] oxytetracycline,[23] erythromycin,[24] picromycin,[25] epothilone,[26] and 6-methylsalicylic acid [27] PKSs in CH999, the deletion host of *S. coelicolor*. In each case, the expected polyketide product was synthesized by the recombinant strain, demonstrating that the heterologous multienzyme systems were metabolically active. Moreover, PKS proteins are produced at 1−5% total cellular protein levels, which in turn has facilitated the development of convenient cell-free systems for polyketide synthesis.[28−30] More recently, the use of *E. coli* as a heterologous host for expressing

(12) Malpartida, F.; Hopwood, D. A. *Nature* **1984**, *309*, 462−464.
(13) Hopwood, D. A. *Microbiology* **1999**, *145*, 2183−2202.
(14) Sherman, D. H.; Malpartida, F.; Bibb, M. J.; Kieser, H. M.; Bibb, M. J.; Hopwood, D. A. *EMBO J.* **1989**, *8*, 2717−2725.
(15) Malpartida, F.; Hallam, S. E.; Kieser, H. M.; Motamedi, H.; Hutchinson, C. R.; Butler, M. J.; Sugden, D. A.; Warren, M.; McKillop, C.; Bailey, C. R.; Humphreys, G. O.; Hopwood, D. A. *Nature* **1987**, *325*, 818−821.
(16) Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. *Nature* **1990**, *348*, 176−178.
(17) Revill, W. P.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1995**, *177*, 3946−52.
(18) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. *Mol. Microbiol.* **1992**, *6*, 3237−3249.
(19) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675−679.
(20) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651−2674.
(21) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546−1550.
(22) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1993**, *115*, 11671−11675.
(23) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 6443−6444.
(24) Kao, C. M.; Katz, L.; Khosla, C. *Science* **1994**, *265*, 509−512.
(25) Tang, L.; Fu, H.; Betlach, M.; McDaniel, R. *Chem. Biol.* **1999**, *6*, 553−558.
(26) Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. *Science* **2000**, *287*, 640−642.
(27) Bedford, D. J.; Schweizer, E.; Hopwood, D. A.; Khosla, C. *J. Bacteriol.* **1995**, *177*, 4544−4548.
(28) Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C. *Nature* **1995**, *378*, 263−266.
(29) Wiesmann, K. E. H.; Cortes, J.; Brown, M. J. B.; Cutter, A. L.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1995**, *2*, 583−589.
(30) Carreras, C.; Pieper, R.; Khosla, C. *J. Am. Chem. Soc.* **1996**, *118*, 5158−5159.

PKS genes has become feasible[31,32] and is expected to gain further relevance as this consummately friendly host is optimized for efficient polyketide production. When successfully implemented, these heterologous expression systems are the tools of choice for in vivo and in vitro studies on account of their greatly enhanced speed and convenience for manipulating PKS genes, and their ability to offer relatively easy access to PKS proteins through overexpression compared to native producing hosts.

The Logic of Polyketide Synthases, As Revealed through Genetic Analysis

Several spectacular features of this enzyme superfamily were revealed as the first few PKS gene clusters were cloned and sequenced in the late 1980s and early 1990s. Contemporaneous characterization of the actinorhodin,[33] granaticin,[14] tetracenomycin,[34] frenolicin,[35] oxytetracycline,[36] and doxorubicin [37] gene clusters by the Hopwood and Hutchinson laboratories revealed close architectural and mechanistic relationships between this subfamily of PKSs (often referred to as "Type II PKSs" in the literature) and bacterial fatty acid synthases. Each of these PKSs possessed a distinct acyl carrier protein (ACP) and a ketosynthase (KS), to which the nucleophilic and electrophilic species that undergo C−C bond formation are attached. However, genetic analysis also exposed some mysteries, which could only be addressed through subsequent biochemical and chemical studies. For example, since no acyltransferase could be found in most gene clusters, the mechanism for charging the ACP with a malonyl group remained unknown. (It turns out that these PKSs "borrow" a malonyl transferase from the endogenous fatty acid synthase for this purpose.[38]) Likewise, although most ketosynthases exist as homodimers, this KS appeared to form a heterodimer with a homologous protein that lacked the signature active site of the KS. (This homologous protein was subsequently shown to play a role in controlling chain length specificity.[21])

While Type II PKSs were being genetically characterized in the U.K. and in Wisconsin, Schweizer and coworkers in Germany cloned and sequenced the 6-methylsalicylic acid synthase.[39] Their studies revealed a "Type I PKS", architecturally related to vertebrate fatty acid synthases, in which all the active sites responsible for polyketide biosynthesis resided in a single large polypeptide.

Perhaps the most architecturally remarkable subclass are the "modular PKSs", first characterized by Katz and co-workers at Abbott Laboratories[19] and by Leadlay and co-workers in Cambridge.[16] These giant megasynthases are comprised of several "modules" of active sites, where each module is responsible for one round of chain elongation and associated modification. In contrast to Type I and Type II PKSs, the logic by which these PKSs control product structure appears to be more overtly inscribed in their genetic code. Around the same time, similar modular megasynthases were also uncovered in the context of nonribosomal peptide biosynthesis by Marahiel and co-workers in Germany working on gramicidin biosynthesis[40,41] and by Turner and co-workers in the U.K. working on β-lactam biosynthesis.[42] Modular PKSs and NRPSs are closely related, and nature often intersperses these two types of modules to make interesting "hybrid" molecules such as rapamycin,[43] FK506,[44] and epothilone.[26,45]

Combinatorial Biosynthesis: A Tool for Dissecting and Exploiting Polyketide Biosynthesis

The development of a robust heterologous expression system for PKSs allowed us to team up with David Hopwood, whose laboratory had isolated numerous Type II PKS gene clusters, to rapidly map the molecular recognition features associated with different biosynthetic pathways using an approach that became known as "combinatorial biosynthesis". Pathways such as actinorhodin, frenolicin, and tetracenomycin had related PKS gene sets that were somehow "programmed" to synthesize products possessing subtle structural variations. To track down the source of these variations, the genes were systematically recombined, expressed, and resulting products were structurally characterized. Numerous experiments of this sort from our own laboratory, as well as from the laboratories of Hutchinson and Floss, led to the emergence of a draft set of "design rules" for predictably altering product structure through genetic manipulation (reviewed in ref 11).

In contrast to the Type II PKS systems, the molecular logic of modular PKSs was more overtly programmed into their gene structure. Here again, numerous studies involving recombination of domains, modules, and subunits were contemporaneously performed in several laboratories including those of Katz, Leadlay, McDaniel, and our own (in collaboration with that of David Cane). These studies, which mostly used the erythromycin PKS as a scaffold, led to a working definition of domain and module junctions, identification of domains responsible for the primary control of key aspects of product structure such as building block specificity, regiospecificity of cyclization, and stereoselectivity of alcohol reduction (reviewed in ref 11). Together, they also provided a vivid

(31) Kealey, J.; Liu, L.; Santi, D. V.; Betlach, M.; Barr, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 505−509.
(32) Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Science* **1999**, *284*, 482−485.
(33) Fernández-Moreno, M. A.; Martínez, E.; Boto, L.; Hopwood, D. A.; Malpartida, F. *J. Biol. Chem.* **1992**, *267*, 19278−19290.
(34) Bibb, M. J.; Biró, S.; Motamedi, H.; Collins, J. F.; Hutchinson, C. R. *EMBO J.* **1989**, *8*, 2727−2736.
(35) Bibb, M. J.; Sherman, D. H.; Omura, S.; Hopwood, D. A. *Gene* **1994**, *142*, 31−39.
(36) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. *Gene* **1994**, *141*, 141−142.
(37) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. *Gene* **1994**, *151*, 1−10.
(38) Carreras, C. W.; Khosla, C. *Biochemistry* **1998**, *37*, 2084−2088.
(39) Beck, J.; Ripka, S.; Siegner, A.; Schiltz, E.; Schweizer, E. *Eur. J. Biochem.* **1990**, *192*, 487−498.
(40) Kratzschmar, J.; Krause, M.; Marahiel, M. A. *J. Bacteriol.* **1989**, *171*, 5422−5429.
(41) Turgay, K.; Krause, M.; Marahiel, M. A. *Mol. Microbiol.* **1992**, *6*, 2743−2744.
(42) Smith, D. J.; Earl, A. J.; Turner, G. *EMBO J.* **1990**, *9*, 2743−2750.
(43) Schweke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J. B.; Bohm, G.; Staunton, J.; Leadlay, P. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839−7843.
(44) Motamedi, H.; Shafiee, A. *Eur. J. Biochem.* **1998**, *256*, 528−534.
(45) Molnar, I.; Schupp, T.; Ono, M.; Zirkle, R.; Milnamow, M.; Nowak-Thompson, B.; Engel, N.; Toupet, C.; Stratmann, A.; Cyr, D. D.; Gorlach, J.; Mayo, J. M.; Hu, A.; Goff, S.; Schmid, J.; Ligon, J. M. *Chem. Biol.* **2000**, *7*, 97−109.

testament for the functional modularity of these megasynthases.

In both of the above cases the availability of heterologous expression systems played a vital role in the speed with which these blueprints for further studies emerged. At the same time, these studies also highlighted the limitations of an in vivo approach toward the dissection of PKS function, in which a PKS gene cluster was genetically manipulated and the structure of a reporter product was spectroscopically analyzed. Many questions could only be answered qualitatively via such an approach; others could not be addressed at all. Direct structural and mechanistic dissection of natural and engineered PKSs would be essential for a better understanding of the properties of these multi-enzyme systems. A key prerequisite for this would be reconstitution of PKS activity in vitro. Again, heterologous expression systems played a key role in achieving this objective.

Biochemical Analysis of Polyketide Synthases

Although methylsalicylic acid synthase was reconstituted in vitro 30 years ago,[46] the ability to extend this principle to more complex PKSs had to wait until the application of recombinant DNA techniques to the field.[28−30,47] The power of cell-free systems stems from the fact that the properties of PKS proteins or their variants can be qualitatively or quantitatively probed in a manner that is unencumbered by any metabolic constraints associated with qualitative or quantitative aspects of precursor availability. However, since small (μg) quantities of polyketide products are ordinarily generated in such experiments, highly sensitive assays (typically radioactive assays) for product formation are required.

Detection and quantitation of cell-free PKS activity using a variety of model systems was followed by purification and kinetic characterization of these proteins.[48,32,38,49,50] The initial (actinomyces-based) heterologous expression systems yielded milligram quantities of purified proteins, a quantity that was adequate for steady-state kinetic analysis but insufficient for more elaborate pre-steady-state kinetic analysis, protein chemical analysis, or structural investigations. More recently, the ability to use *E. coli*-based expression systems for producing soluble, active PKS proteins has greatly loosened this constraint.

As is always the case, the availability of reagent quantities of an interesting protein (or multi-protein system) spawns interesting chemistry. In the case of PKSs, a number of longstanding questions about polyketide biosynthesis are finally within grasp. What are the salient architectural features of PKSs, and how do these vary between different subclasses of this enzyme superfamily? How are their molecular recognition features parsed among different subunits, domains, and modules? How strong is the discrimination of individual enzymes toward unnatural substrates? Is there a good correlation between the intrinsic substrate specificity of individual enzymes and the structures of their natural substrates? And finally, what is the balance of power between protein−protein interactions and protein−substrate interactions as reactive biosynthetic intermediates are channelled from one active site to the next?

Answers to these questions are only just beginning to emerge. For example, it is becoming increasingly evident that, although most PKS enzymes display some substrate specificity, in many cases this selectivity is not strong enough to have a serious impact on the overall turnover of the megasynthase in the presence of an unnatural substrate.[51] Another remarkable feature of PKSs that is becoming apparent is the importance of protein−protein interactions on substrate channelling from one active site to the next.[32,52,53] Although the role of protein−protein interactions is keenly appreciated in biological signal transduction processes (such as the regulation of gene expression), it has been poorly explored in metabolic biochemistry, where individual enzyme are expected to have exquiste specificity for the product of the previous reaction. As the structural and mechanistic basis for these and other properties becomes clearer, this information will undeniably loop back to the design drawing board through the emergence of a new generation of strategies for combinatorial biosynthesis.

The Role of Organic Chemistry: Past, Present, and Future

There are two principal levels at which the impact of organic chemistry has been felt on the above studies. First, analytical and spectroscopic tools have been essential to the process of elucidating the structures of new metabolites, or "unnatural" natural products, generated through combinatorial biosynthesis. Some of these structures have closely resembled known natural products; in these cases the assignment of structures has been relatively straightforward. In many cases however, genetic manipulation has yielded completely new (and unexpected) classes of compounds, and the challenge of structure elucidation has been akin to that encountered in classical natural products chemistry, but has nevertheless been somewhat simplified by the prior knowledge that the compound in question is a polyketide, and the ability to selectively incorporate isotopically labeled precursors such as acetate and propionate.

In addition to the above, synthetic organic chemistry has also proven to be an invaluable companion to genetics and biochemistry in many recent studies on polyketide biosynthesis. Historically, biosynthetic mechanisms have often been probed through the use of isotopically labeled forms of putative pathway intermediates that are fed to cultures of producing microorganisms. If the biosynthetic model is correct, then the expectation is that the labeled compound will be incorporated into the final natural product. Negative results are inconclusive and, even when incorporation is observed, it is often inefficient. (Incorporation levels of 0.1−1% are not uncommon.) Moreover, since a single experiment can often require large quantities (several hundreds of milligrams) of synthetic material, proposals for studies that involve

(46) Dimroth, P.; Walter, H.; Lynen, F. *Eur. J. Biochem.* **1970**, *13*, 98−108.
(47) Shen, B.; Hutchinson, C. R. *Science* **1993**, *262*, 1535−1540.
(48) Pieper, R.; Gokhale, R. S.; Luo, G.; Cane, D. E.; Khosla, C. *Biochemistry* **1997**, *36*, 1846−1851.
(49) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. *Biochemistry* **1998**, *37*, 8132−8138.
(50) Bycroft, M.; Weissman, K. J.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **2000**, *267*, 520−526.
(51) Wu, N.; Kudo, F.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2000**, *122*, 4847−4852.
(52) Dreier, J.; Shah, A. N.; Khosla, C. *J. Biol. Chem.* **1999**, *274*, 25108−25112.
(53) Dreier, J.; Khosla, C. *Biochemistry* **2000**, *39*, 2088−2095.

anything beyond relatively straightforward synthetic schemes are cautiously evaluated, and successful results are looked upon as small acts of heroism. The impact of modern biology has changed the outlook for such studies dramatically in two significant ways. First, by using genetically tractable and physiologically defined heterologous hosts such as *Streptomyces coelicolor* and *E. coli* for natural product biosynthesis, one can dramatically enhance the efficiency of incorporation of isotopically labeled molecules into the natural product,[54] thereby reducing the need for large quantities of synthetic material. Indeed, by reconstituting the PKSs in vitro (again derived from these same heterologous systems), one can reduce material requirements to sub-milligram quantities, thereby opening the door to the usage of more complex synthetic reagents and even reagents derived from parallel combinatorial chemistry. Second, by using precisely engineered, genetically blocked mutants, one can altogether eliminate competition from the endogenous biosynthetic pathway, thereby opening the door for incorporation of unnatural synthetic precursors that may not be able to efficiently compete with the natural pathway.[55,56] A variety of new "unnatural" natural products with nonbiological functional groups have been generated in preparative quantities using these approaches. Soon one can expect to see significant further developments in this area of "chemo-biosynthesis", as the tactics for introducing unnatural amino acids into polypeptides, using a combination of synthetic building blocks and transferases with orthogonal molecular recognition features,[57] are adapted to polyketide biosynthesis.

Implications for Natural Product Drug Discovery and Development

So how are these fundamental investigations into PKS structure and mechanism going to impact the discovery and development of new natural product drugs in the future? I started this paper by alluding to the two central problems that hamper the emergence of new natural product drugs—the difficulties associated with regio- and stereoselective modification of these complex molecules and the lack of availability of adequate materials for further biological and chemical investigations. The application of new protein engineering tools and methods to PKSs has clear implications for ameliorating the former limitation. Especially valuable will be new ways to selectively introduce orthogonal functional groups into different segments of naturally occurring polyketide scaffolds, so as to simplify the medicinal chemist's task of developing structure–activity relationships. As illustrated by the history of protein chemistry, the easier it is to manipulate polyketide structure, the wider will be the applicability of these technologies. Therefore, future efforts aimed at decoding the structures and mechanisms of PKSs, in conjunction with the development of more sophisticated tools for gene cloning, expression and manipulation, are likely to remain the focus of such research in the next decade.

But what about the latter problem? The enhancement of both the volumetric productivity as well as the atom economy of a polyketide fermentation process represents a major untapped frontier at the interface of chemistry, biology, and engineering. Heterologous expression of complex biosynthetic pathways offers a convenient springboard by shifting the relevant chemistry from highly diverse and often esoteric hosts into biologically well-studied environments, but it is no more than a starting point. To overproduce polyketides routinely, systematic improvements will have to be engineered in selected hosts to enhance their quantitative biosynthetic capabilities. Again, this is completely analogous to the role that molecular biology has played in present-day protein chemistry. At first glance the problem may appear substantially more complex in the case of natural product chemistry; however, the impact of genomics and functional genomics is sure to be felt in this area. With the *E. coli* and *B. subtilis* genomes completely sequenced, and the *S. coelicolor* genome over 70% sequenced at the time of this writing, the stage is set to explore how the metabolism of these organisms might be adapted and controlled for the production of these high-value natural products.

Until relatively recently, chemicals such as nucleic acids and proteins were not regarded as fertile grounds for drug discovery and development. Molecular biology provided the fundamental scientific underpinnings and associated technology developments that changed this dogma. In contrast, the drug-like features of polyketide natural products are universally accepted, and this family of biomolecules continues to yield unique leads in a broad spectrum of pharmacological assays. However, the difficulties associated with translating this promise into practical utility are believed to be enormous. The coming decade will determine whether our rapidly burgeoning knowledge of the chemistry of polyketide biosynthesis will be harnessed to rise to this challenge.

Acknowledgment. Research in this laboratory has been supported by grants from the National Science Foundation, the National Institutes of Health, the Packard Foundation, and generous gifts from Sankyo Co. and AstraZeneca. I also acknowledge the contributions of co-workers and collaborators over the past decade.

JO000849Y

(54) Cane, D. E.; Luo, G.; Khosla, C.; Kao, C. M.; Katz, L. *J. Antibiot.* **1995**, *48*, 647–651.
(55) Jacobsen, J. R.; Khosla, C. *Curr. Opin. Chem. Biol.* **1998**, *2*, 133–137.
(56) Marsden, A. F. A.; Wilkinson, B.; Cortes, J.; Dunster, N. J.; Staunton, J.; Leadlay, P. F. *Science* **1998**, *279*, 199–203.
(57) Liu, D. R.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4780–4785.