Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Chemical Biology

Combinatorial biosynthesis of polyketides — a perspective

Fong T Wong[1] and Chaitan Khosla[1,2]

Since their discovery, polyketide synthases have been attractive targets of biosynthetic engineering to make 'unnatural' natural products. Although combinatorial biosynthesis has made encouraging advances over the past two decades, the field remains in its infancy. In this enzyme-centric perspective, we discuss the scientific and technological challenges that could accelerate the adoption of combinatorial biosynthesis as a method of choice for the preparation of encoded libraries of bioactive small molecules. Borrowing a page from the protein structure prediction community, we propose a periodic challenge program to vet the most promising methods in the field, and to foster the collective development of useful tools and algorithms.

Addresses
[1] Department of Chemical Engineering, Stanford University, Stanford, CA 94305, United States
[2] Department of Chemistry, Stanford University, Stanford, CA 94305, United States

Corresponding author: Khosla, Chaitan (khosla@stanford.edu)

Current Opinion in Chemical Biology 2012, **16**:117–123

This review comes from a themed issue on
Biocatalysis and Biotransformation
Edited by Jon S Thorson and Ben Shen

Available online 16th February 2012

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2012.01.018

Introduction

Polyketides are a structurally diverse but biosynthetically related family of natural products that includes a number of medicinally important substances such as lovastatin (a cholesterol-lowering agent), erythromycin (an antibiotic), and FK506 (an immunosuppressant) [1]. Their structural and stereochemical complexity makes systematic chemical manipulation a formidable undertaking. Consequently, there has been considerable interest in the potential of harnessing combinatorial biosynthesis to introduce novel functionality into these bioactive compounds and to produce altogether new chemotypes.

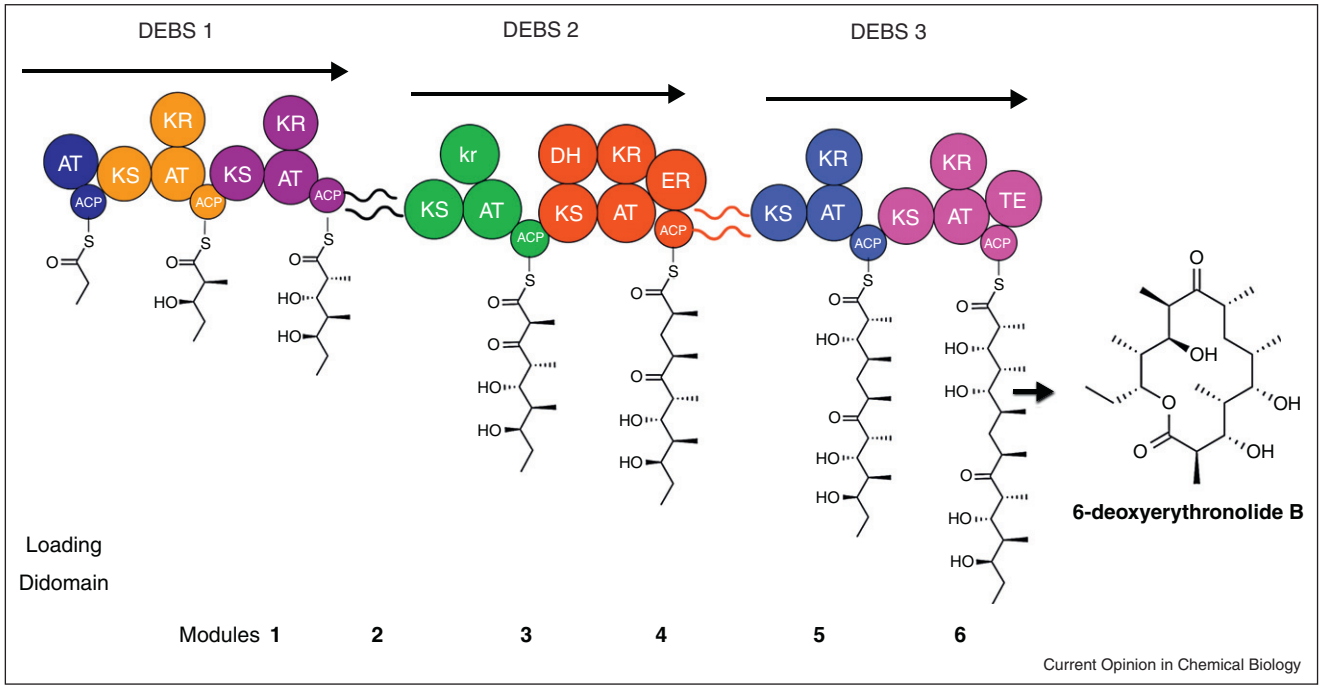
In this review, combinatorial biosynthesis is defined as the genetic manipulation of two or more enzymes involved in polyketide biosynthesis. According to this enzyme-centric definition, combinatorial biosynthesis could even yield the natural product itself, as long as the corresponding polyketide synthase (PKS) harbors two or more genetically modified enzymes. These enzymatic modifications can be accomplished by either genetic manipulation of the original enzyme or by replacing it with a homolog (although the latter approach is more common at the present time). By contrast, a product-centric definition of combinatorial biosynthesis would encompass natural product analogs with two or more functional group transformations, regardless of how these modifications are achieved. For example, products of combinatorial biosynthesis could be derived via precursor directed biosynthesis or through other metabolic engineering strategies [2]. We have chosen an enzyme-centric definition because, in our opinion, it highlights the fundamental and technological challenges to exploiting the functional modularity of PKSs [3]. Specifically, combinatorial biosynthesis can be achieved by manipulating enzymes responsible for primer unit incorporation, chain elongation, and chain termination.

The state of the art

The feasibility of combinatorial biosynthesis has been demonstrated in the context of different types of multifunctional PKSs [4–8]. Whereas the architectures of these PKS subfamilies are variable, all multifunctional PKSs harbor one or more ketosynthases (KS), acyltransferases (AT) and acyl carrier proteins (ACP). In addition, most PKSs also include auxiliary enzymes such as reductases, dehydratases, transferases, cyclases, and thioesterases. To highlight the scope of combinatorial biosynthesis, here we primarily focus on multimodular PKSs with assembly line architectures. Like an automobile assembly line, these PKSs have multiple way stations (called 'modules'), each of which harbors distinct protein domains. Except in a few rare cases, each module is deployed only once in the PKS catalytic cycle; this one-to-one correspondence facilitates convenient mapping of each enzyme domain in the PKS to a unique reaction in the polyketide biosynthetic pathway. A prototypical example of this PKS subfamily is the 6-deoxyerythronolide B synthase (DEBS) (Figure 1) [9]. A particularly impressive showcase for the enzymatic complexity of assembly line PKSs is the FR901464 biosynthetic synthase (Figure 2) [10••]. FR901464 is synthesized by an assembly line encompassing a PKS with several atypical architectural and enzymatic features, including a nonribosomal peptide synthetase and an HMG-CoA reductase.

Phylogenetic and structural analysis of assembly line PKSs suggests that nature has harnessed gene duplication, mutation, and recombination to pursue combinatorial biosynthesis over evolutionary time [11]. In a presumably analogous laboratory investigation, ca. 50 analogs of

Figure 1



The 6-deoxyerythronolide B synthase (three genes, 32 kbp) is a canonical multimodular PKS [9]. The growing chain is shown as it moves down the assembly line. N and C terminal linkers are also shown. KS: ketosynthase, AT: acyltransferase, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, ACP: acyl carrier protein. The ketoreductase domain in module 3 is inactive (shown in lower caps).

6-deoxyerythronolide B were produced by engineering two or more domains of DEBS [5]. The latter study was enabled by the establishment of tools for the reconstitution of complete PKS pathways into genetically amenable hosts such as *Streptomyces coelicolor* [12,13].

Notwithstanding encouraging progress over the past two decades [7,14–16,17••], the promise of rationally guided combinatorial biosynthesis remains unrealized. In the sections that follow, we discuss key ecological, enzymological, and technological challenges that must be addressed in order to efficiently synthesize libraries of 'unnatural' natural products.

Ecological challenges

Until recently, a major obstacle to combinatorial biosynthesis was the availability of DNA sequences of an adequately large number of cloned PKS genes. Less than 20 multifunctional PKS gene clusters had been fully sequenced by the turn of the millennium. As high-throughput sequencing techniques gained momentum, this number increased exponentially. Whereas the growth in PKSs corresponding to structurally characterized natural products has remained modest, the emergence of whole genome sequencing methods has resulted in the discovery of cryptic gene clusters at an explosive pace (Figure 3). Not only has there been an immense growth in the repertoire of enzyme domains and modules, but new assembly line architectures have also been discovered (e.g. 'AT-less' PKSs [18,19]). Today, an aspiring biosynthetic engineer has access to a virtually infinite palette of genetic raw material, although much of it remains to be functionally decoded.

Notwithstanding breathtaking advances in mining nature's PKS gene clusters [20], the ability to identify complete PKSs from unculturable microorganisms remains seriously constrained. The development of resource-efficient strategies for cloning and sequencing large (20–100 kb) contigs from metagenomic sources will enable at least two related types of opportunities in combinatorial biosynthesis. First, the DNA encoding unprecedented chemotypes could become accessible. For example, close structural analogs of marine natural products such as discodermolide [21] or spongistatin [22] have not yet been isolated from cultured microorganisms. Combinatorial biosynthesis of discodermolide or spongistatin analogs is therefore predicated upon cloning their complete gene clusters. Second, thus far, the vast majority of cloned PKS genes have been isolated from terrestrial bacteria, primarily the actinomycetes, bacilli and myxobacteria. As the genetic content of the earth's oceans is mined for PKSs, new biocatalytic strategies will surely emerge, which in turn could be exploited through