Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Biotechnology

Molecular tools for chemical biotechnology

Stephanie Galanie[1,3], Michael S Siddiqui[2] and Christina D Smolke[3]

Biotechnological production of high value chemical products increasingly involves engineering *in vivo* multi-enzyme pathways and host metabolism. Recent approaches to these engineering objectives have made use of molecular tools to advance *de novo* pathway identification, tunable enzyme expression, and rapid pathway construction. Molecular tools also enable optimization of single enzymes and entire genomes through diversity generation and screening, whole cell analytics, and synthetic metabolic control networks. In this review, we focus on advanced molecular tools and their applications to engineered pathways in host organisms, highlighting the degree to which each tool is generalizable.

Addresses
[1] Department of Chemistry, Stanford University, Stanford, CA 94305, United States
[2] Department of Chemical Engineering, Stanford University, Stanford, CA 94305, United States
[3] Department of Bioengineering, Stanford University, 473 Via Ortega, MC 4201, Stanford, CA 94305, United States

Corresponding author: Smolke, Christina D (csmolke@stanford.edu)

Current Opinion in Biotechnology 2013, **24**:1000–1009

This review comes from a themed issue on **Chemical biotechnology**

Edited by **Kristala LJ Prather**

For a complete overview see the Issue and the Editorial

Available online 23rd March 2013

0958-1669/$ – see front matter,

http://dx.doi.org/10.1016/j.copbio.2013.03.001

Introduction

Chemical biotechnology is the use of biocatalysts in engineered systems to produce bulk and fine chemicals [1]. Three waves of biocatalysis have been described: first, realization that biological components could be used for chemical transformations; second, development of genetic engineering techniques needed for industrial production of proteins; and third, development of directed evolution-enabled enzyme engineering [2]. The coming wave relies not only on further improvements in protein engineering and DNA synthesis technologies, but also critically on our ability to engineer controlled, multi-enzyme pathway systems. Although isolated enzymes are widely industrially used today, whole cells are a more feasible system for multi-enzyme pathways. The introduction of heterologous pathways into a host organism and metabolic flux optimization toward the product of interest constitute a synergistic application of concepts from metabolic engineering and synthetic biology [3]. In this review, we describe selected contributions of metabolic engineering, synthetic biology, systems biology, and protein engineering to chemical biotechnology to improve the productivity of multi-enzyme pathways. These fields have provided advanced molecular tools for *de novo* pathway identification, tuned pathway construction, diversity generation and screening, genome-scale identification of optimization targets, and dynamic pathway control. Here, we focus on such molecular tools developed, improved, and applied in new contexts over the past few years.

Enhanced tools for precise biosynthetic pathway construction

Engineered biosynthetic pathways require composition of genetically encoded expression devices that support precise and tunable levels of pathway enzymes. Both the number of characterized control elements — such as ribosome binding sites (RBSs), promoters, and terminators — and the degree to which those control elements can be made to behave in a predictable manner under a range of contexts have expanded. Additionally, improved and new methods have been developed to assemble these control elements with enzymes in order to construct biosynthetic pathways.

One challenge to the rational design of genetically encoded elements is that they often behave in a context-dependent manner, exhibiting properties that depend on the combination of other elements used in the device or exhibiting off-target perturbations of the biological host. This challenge has been addressed by designing for context and then iteratively optimizing to improve behavior. Contextual features to consider range from the specific (e.g. DNA sequence surrounding the element) to the holistic (e.g. environmental growth conditions). For example, the impact of oxygen and glucose conditions on constitutive yeast promoter activities was characterized to permit design for these culture conditions [4]. Known variation in tRNA availabilities among hosts has been used to reduce the host-dependence of protein expression via codon optimization [5,6]. Alternatively, insulated elements have been developed that behave robustly in varying contexts. For example, researchers have developed an insulated constitutive bacterial promoter library with relative protein production rates that span two orders of magnitude and are independent of the coding sequence of the expressed protein [7•]. New insulating elements that use RNA processing to reduce the context dependence of genes in multi-gene operons have also been introduced [8].

Quantitative modeling and characterization of control elements have enabled researchers to create larger

libraries of elements that exhibit predictable behaviors when integrated into gene expression devices. For example, a thermodynamic model of bacterial translation initiation was developed and used to forward design synthetic RBSs with a 47% chance of exhibiting protein expression levels within 2.3-fold of the desired level [9]. In addition, libraries of novel gene control elements have been developed using evolutionary and screening strategies. For example, a set of Rnt1p-cleavable hairpins provides post-transcriptional tuning of protein expression levels ranging from 8 to 84% of a control construct without a hairpin [10]. *In vivo* scaffolds are another set of synthetic control elements and act post-translationally to improve pathway flux by spatially co-localizing enzymes to RNA [11••], DNA [12], protein [13,14], cell surface [15], or a specific organelle [16•].

Along with the diversification and insulation of genetic control elements have come faster, more reliable methods to construct biosynthetic pathways. Notably, an eight-gene biosynthetic pathway was assembled into a shuttle vector or yeast artificial chromosome in a single transformation with over 50% efficiency (Figure 1a) [17]. Pathways can also be integrated iteratively, which may increase the accessible library size for testing variants in a multi-gene pathway [18]. The construction and transplantation of a chemically synthesized bacterial genome showcased the cumulative advances of *in vitro* enzyme-mediated assembly and *in vivo* transformation-associated-recombination in yeast [19]. Similar techniques have been used to replace chromosome arms in yeast with circular or linear synthetic versions [20••]. In *E. coli*, researchers have developed 'recombineering' methods that use phage proteins to facilitate recombination-based genetic engineering. Rec-mediated recombineering was developed for efficient recombination between linear PCR products and linearized plasmids, which complements efficient lambda phage Red-mediated recombination between linear PCR products and circular plasmids [21].

Improved tools for *de novo* pathway identification

Engineered biosynthetic pathways were once painstakingly pieced together from a single organism's cDNA to mimic natural biosynthesis strategies, and optimization of host platforms involved serial knockouts or overexpression of targeted genes. However, modern bioinformatic tools allow rapid mining of huge sequence repositories for functions of interest and even prediction of multiple possible pathways to produce a desired small molecule (Figure 1a). Databases such as BioCyc [22], Kyoto Encyclopedia of Genes and Genomes (KEGG) [23], Rhea [24], and Braunschweig Enzyme Database (BRENDA) [25] facilitate search and display of metabolic pathways. These databases and others are populated by proteins whose discovery is accelerated by functional prediction algorithms, for example, global biochemical reconstruction using sampling (GLOBUS), which takes a systems biology approach to adjusting functional predictions generated by homology and secondary structure [26]. Genome-scale metabolic models of several industrially useful host organisms are available and reviewed elsewhere [27]. These stoichiometric models are used in computational predictions of routes from central metabolites to a product of interest, implementing a variety of rules for reaction qualifications (e.g. known enzymes, metabolites, or reaction chemistries) and scoring schemes (e.g. route length, number of enzymes known, complexity of transformations, flux balance analysis) to rank possible pathways [28]. For example, researchers used in-house pathway prediction and ranking software with a genome-scale *E. coli* model to identify two possible pathways to the non-natural product 1,4-butanediol [29•].

In a creative use of genomic information, an alkane-producing operon was discovered via subtractive genomic analysis of 11 cyanobacteria strains and used in *E. coli* to produce a mixture of alkanes and alkenes [30]. Enzymes variants can be rationally selected when enzymes with similar functions from many organisms are available. In the reverse engineering of the β-oxidation cycle for production of *n*-butanol, enzymes were chosen on the basis of kinetic parameters and co-factor specificity [31•,32]. The combination of genomic information and library-based cloning strategies is a rapid way to obtain rare enzymatic activities. In one example, a 30 kb *Vibrio splendidus* genome fragment, suspected by homology to encode transporters needed for alginate utilization, was cloned into *E. coli* by screening a genomic fosmid library [33••].

Tools for molecular diversity generation and screening

Directed evolution is a powerful method for optimizing biosynthetic pathways and requires molecular tools for generating and screening diversity (Figure 1c). Tools for generating diversity at the single-gene level, such as error-prone PCR, DNA shuffling, saturation mutagenesis, and site-directed mutagenesis, have been used to generate enzyme libraries of varying size for some time. Screening for small molecule production depends on physical characteristics of the substrates, cofactors, products, or strains. Screens employ selection, growth, colorimetric, fluorescent, or UV readouts in a high-throughput format; or direct quantification by a separation method coupled to a detection method, like LC–MS, in lower-throughput formats. Recent work to engineer *E. coli* for increased levopimaradiene production provides examples of both small and large library generation and screening [34]. One enzyme in the pathway, levopmaradiene synthase, was iteratively subjected to site-directed and single-site saturation mutagenesis at amino acid positions selected from a structural homology model, and resulting changes in productivity were analyzed by