COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 3, Issue: 4, October 2012, e201210021, http://dx.doi.org/10.5936/csbj.201210021

CSBJ

COMBINATORIAL APPROACHES FOR INVERSE METABOLIC ENGINEERING APPLICATIONS

Georgios Skretas [a,*], Fragiskos N. Kolisis [b,*]

Abstract: Traditional metabolic engineering analyzes biosynthetic and physiological pathways, identifies bottlenecks, and makes targeted genetic modifications with the ultimate goal of increasing the production of high-value products in living cells. Such efforts have led to the development of a variety of organisms with industrially relevant properties. However, there are a number of cellular phenotypes important for research and the industry for which the rational selection of cellular targets for modification is not easy or possible. In these cases, strain engineering can be alternatively carried out using "inverse metabolic engineering", an approach that first generates genetic diversity by subjecting a population of cells to a particular mutagenic process, and then utilizes genetic screens or selections to identify the clones exhibiting the desired phenotype. Given the availability of an appropriate screen for a particular property, the success of inverse metabolic engineering efforts usually depends on the level and quality of genetic diversity which can be generated. Here, we review classic and recently developed combinatorial approaches for creating such genetic diversity and discuss the use of these methodologies in inverse metabolic engineering applications.

REVIEW ARTICLE

1. Introduction

Metabolic engineering has been a well established scientific discipline for over two decades now [1]. During this period, advances in the understanding of the functional role of thousands of genes in various organisms, and the development of theoretical and experimental tools for determining the flow of metabolites through different biochemical pathways [2,3], have provided hints for numerous potential cellular targets whose modification can lead to optimal metabolism and improved properties.

More recently, advances in sequencing technologies have resulted in an explosion in the number of sequenced microbial genomes, revealing a plethora of novel enzymes, biochemical reactions and pathways, while the development of efficient methodologies for performing directed protein evolution has enabled the engineering of enzymes with tailored activities [4,5]. Furthermore, the emergence of more global approaches for analysing cell function, such as systems biology, have contributed to our understanding of how biochemical pathways operate, not in the form of sequential "isolated" reactions, but instead as complex, interdependent and dynamic networks of such reactions [6]. Finally, synthetic biology has extended the capabilities of classical metabolic engineering as it has provided the rationale and tools to combine biological components and generate designed gene networks, artificial metabolic pathways, and organisms with man-made genomes [7,8].

[a]*Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, Athens, Greece*

[b]*Biotechnology Laboratory, School of Chemical Engineering, National Technical University of Athens - Zografou Campus, Athens, Greece*

* Corresponding author.

E-mail address: gskretas@eie.gr (Georgios Skretas)

kolisis@chemeng.ntua.gr(Fragiskos N. Kolisis)

These technologies, in combination with the significantly advanced analytical tools available today for analyzing levels of DNA, RNA, proteins and metabolites in the cell, have enabled the development of engineered prokaryotic and eukaryotic organisms with the ability to produce a wide array of chemicals useful for research and the industry: amino acids [9], antibiotics [10], alkaloids [11], isoprenoids [12], vitamines, fragrances, peptides and polyketides [13], organic solvents and biofuels [14,15], nutraceuticals, polymers [16] and others.

A prerequisite for the success of classical (forward) metabolic engineering is a detailed knowledge of the biochemical pathways involved in the biosynthesis of a particular metabolite or in the appearance of a desired phenotype. This knowledge is necessary in order to be able to make targeted and rationally selected genetic modifications. However, a variety of cellular properties important for the industry and research, such as resistance to organic solvents and optimal production of certain metabolites, are still poorly characterized and frequently arise from modifications in pathways that involve genes of unknown function or ones that would be impossible to predict by rational engineering, possibly because they act by indirect or compensatory mechanisms. In these cases, where educated guesses about possible interventions are hard or impossible to make, a different strategy can be applied, which is termed "inverse metabolic engineering" [17]. In this type of metabolic engineering, the desired property is first linked to a readily detectable phenotype, e.g. changes in cell growth, color, fluorescence etc. Then, random genetic modifications, such as chromosomal point mutations, gene deletions, gene over-expressions etc., are introduced into the host so that a library of cells with genetic diversity is generated. This library is then screened and the clones that exhibit the desired phenotype are identified. Genetic analysis of the isolated clones subsequently can reveal the factors responsible for the improved properties and lead to enhanced understanding of the (previously unidentified) biochemical processes involved.

Given the availability of an effective genetic screening or selection system for the desired property, the potential success of an inverse

metabolic effort is highly dependent on the level and quality of genetic diversity which can be generated. Here, we review combinatorial approaches for generating genetic diversity within a host cell population, and discuss the applications of these methodologies in inverse metabolic engineering. Emphasis is placed on approaches which have been developed in recent years.

2. Classical approaches in inverse metabolic engineering

2.1. Spontaneous mutagenesis

Spontaneously acquired mutations leading to increased fitness under specific growth conditions is the classical adaptive evolution paradigm. This has been used by mankind for millennia, and by the industry and academia for decades in order to generate useful organisms for a variety of applications. Numerous successes in inverse metabolic engineering mediated through spontaneous mutagenesis have been reported and include increased tolerance to isobutanol and ethanol [18,19], growth on citrate [20]; production of D-lactate [21] and hard-to-express recombinant proteins in *Escherichia coli* [22]; production of 1,3-propanediol in *Klebsiella pneumoniae* [23]; xylose and galactose utilization in *Saccharomyces cerevisiae* [24,25]. Analysis of the genetic lesions that lead to improved phenotypes in spontaneously evolved strains used to be a complicated or impossible process for variants carrying multiple and distantly located mutations, but the recent development of -omics technologies coupled with the increased throughput and decreased cost of sequencing technologies has made this a very tractable task [19-21,26].

2.2. Random mutagenesis with chemical mutagens

Libraries of cells containing lesions randomly distributed over the entire chromosome can be readily generated by exposing a population to sub-lethal doses of mutagenic chemicals, such as N-methyl-N′-nitro-N-nitrosoguanidine and ethyl methanesulfonate or other mutagenic agents, such as UV irradiation. The use of such agents has resulted in the development of microbial strains with the ability to produce enhanced amounts of isobutanol [27], full-length IgG antibodies [28] and membrane proteins [29].

2.3. Transposon mutagenesis

Genes whose products have a negative impact on a desired property can be readily identified by transposon mutagenesis. This type of mutagenesis results in random insertions of transposable elements throughout the genome with concomitant functional disruption of the gene sequence that received the insertion. Transposon mutagenesis has led to the development of improved strains and the identification of inhibitory roles for genes involved in the production of biomass [30], lycopene [31], and membrane proteins [32] in *E. coli*; riboflavin production in *Bacillus subtilis* [33]; and poly-3-hydroxybutyrate in *Synechocystis* sp. PCC6803 [34], to name a few examples. Very useful tools for studying the effect of gene knockouts is the Keio collection, a publicly available library of all single knockouts of all the non-essential *E. coli* K-12 genes [35] and the yeast deletion collection [36]. The utility of these libraries in inverse metabolic engineering has already been demonstrated [37] and it is expected that such libraries will be increasingly used in the coming years.

2.4. Gene overexpression libraries

Genes, gene fragments or fragments of entire operons that favorably affect a desired property can be isolated from vector libraries co-expressing genomic fragments. Genomic libraries have been screened in order to identify genes that enhance alcohol tolerance/production and galactose fermentation in *S. cerevisiae* [38-40]; acetate and butanol tolerance [41,42], lycopene [43] and membrane protein production [44] in *E. coli*; butyrate tolerance in *Clostridium acetobutylicum* [45], and in other cases. In addition, individual enhancer genes can be identified using the ASKA library, a library of all the *E. coli* open reading frames (ORFs) transcribed from the strong and inducible T5*lac* promoter [46] or the FLEXgene collection, an analogous library encoding yeast ORFs from *S. cerevisiae* [47], both of which are publicly available. Again, such collections have already been used successfully for inverse metabolic engineering applications [48,49]. To explore the functional genomic content of unculturable organisms, similar screens can also be carried out by constructing and screening metagenomic libraries [50].

Finally, additive positive effects from genes and/or operons distantly located within a genome can be identified by a recently developed tool termed coexisting/coexpressing genomic libraries (CoGeLs) [51]. CoGeLs allow the simultaneous screening of two genomic libraries encoded in different vectors with compatible origins of replication which can coexist in the same host. These vectors can be regular bacterial plasmids that contain small- or medium-size inserts (up to 10 kbases) or fosmids, which are low-copy number vectors where large DNA fragments (~40 kbases) can be inserted. Using a combination of two plasmid-encoded *E. coli* genomic libraries, Nicolaou et al. demonstrated that CoGeLs can be used to identify combinations of distantly located factors that impart increased acid resistance in *E. coli* [51].

3. Recently developed approaches in inverse metabolic engineering

In recent years, a number of new approaches have been developed that attempt to create more "global" changes on cellular pathways and physiology as means of generating complex phenotypes more effectively. The majority of those strategies aim at modifying the transcriptional landscape of an organism, e.g. by generating libraries of randomized transcription factors or by mutating components of the RNA polymerase. Since certain components of the transcriptional machinery in prokaryotic as well as eukaryotic organisms regulate the expression of a wide repertoire of genes, subtle changes in these components can have a pronounced effect on the transcriptome of the cell, thus offering the potential for the emergence of diverse and complex phenotypes [52]. Some examples of these methods are described below and are summarized in Table 1.

3.1 Zinc finger-based artificial transcription factors

Zinc fingers are highly specific DNA-binding protein domains which recognize three-base pair sequences and are found in many proteins that regulate transcription in a variety of organisms. One transcription factor can include several of these motifs. The ones that contain more fingers recognize larger stretches of DNA and are, generally, more specific about the genomic loci that they bind. Currently, at least one natural or engineered zinc finger exists for every possible triplet of DNA bases (4x4x4=64 possible triplets) [53-55]. Park et al. took advantage of the wide repertoire of DNA-binding specificities and the highly modular way with which these proteins can be assembled [56] to generate libraries of artificial transcriptional activators and repressors that can activate or silence practically any gene within a eukaryotic genome [57], with the ultimate goal of evolving complex phenotypes. First, they selected 40 and 25 zinc fingers with diverse DNA-binding specificities and used