Metabolic Engineering ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Genome scale engineering techniques for metabolic engineering

Rongming Liu [a,1], Marcelo C. Bassalo [b,1], Ramsey I. Zeitouna [a,b], Ryan T. Gill [a,*]

[a] Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309, United States
[b] Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, CO 80309, United States

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form
15 August 2015
Accepted 2 September 2015

Keywords:
Metabolic engineering
Synthetic biology
Genome scale engineering
Design-Build-Test-Learn

ABSTRACT

Metabolic engineering has expanded from a focus on designs requiring a small number of genetic modifications to increasingly complex designs driven by advances in genome-scale engineering technologies. Metabolic engineering has been generally defined by the use of iterative cycles of rational genome modifications, strain analysis and characterization, and a synthesis step that fuels additional hypothesis generation. This cycle mirrors the Design-Build-Test-Learn cycle followed throughout various engineering fields that has recently become a defining aspect of synthetic biology. This review will attempt to summarize recent genome-scale design, build, test, and learn technologies and relate their use to a range of metabolic engineering applications.

1. Introduction

Metabolic engineering is concerned with the engineering of biological systems for the purpose of manipulating flux towards desired products. A central goal of the field is to develop *forward* engineering approaches that are driven by predictive models and associated theory. Since such approaches require both sufficient understanding to develop models and genetic engineering tools to construct and test model predictions, the history of the field has focused primarily on the modification of a small number of genes with clear links to a targeted pathway. Typical modifications include overexpression of rate-limiting steps in the pathway, introduction of heterologous genes, and/or removal of competing pathways. Efforts along these lines have proven successful in increasing production titers from a broad range of platform strains, with applications ranging from bulk chemicals (Song et al., 2013; Yang et al., 2014), biofuels (Choi et al., 2012; Jang et al., 2012), to pharmaceuticals (Martin et al., 2003; Paddon and Keasling, 2014) and food derivatives (Kaur et al., 2014), among others.

In the last decade, metabolic engineering has shifted from designs targeting a handful of genes with close metabolic network relationships to increasingly complex designs requiring the modification of dozens of genes spanning a broad range of metabolic functions (transporters, pathway enzymes, tolerance genes, etc.). To support this increased engineering complexity, metabolic engineering can now be generally defined by the use of iterative cycles of rational genome modification, systems level characterization, and sophisticated analysis. This approach mirrors the Design-Build-Test-Learn (DBTL) cycle from the computational and engineering sciences (Fig. 1). Here, we review applications and successes of genome scale engineering techniques for metabolic engineering based on the DBTL concept that link i) pathway design algorithms with active machine learning, ii) next-generation DNA synthesis and assembly with genome-engineering, and iii) laboratory automation with ultra-high throughput and sensitive genomics methods.

2. Pathway design algorithms with active machine learning

Conventional "design" typically involves a combination of literature searching, metabolic modeling, and heuristics. This design approach has limited throughput, where typically only a handful of designs are considered in depth. Recently, our understanding of microbial metabolism has greatly increased with accumulating bio-information on gene functions (Kan et al., 2012), genome structures (Lam et al., 2012), biological pathways (Peralta-Yahya et al., 2012), metabolic and regulatory networks (Gerosa and Sauer, 2011), and evolution of genomes (Blount et al., 2012). This knowledge makes it possible for the DBTL design to provide the complete set of build instructions for any target molecule, enabling rapid discovery of pathway configurations for reliable target molecule production.

Computational algorithms such as constraint-based flux balance analysis (FBA) are essential tools to predict phenotypic properties in genome scale modeling, which was widely used in

* Corresponding author.
E-mail addresses: rongming.liu@colorado.edu (R. Liu), marcelo.bassalo@colorado.edu (M.C. Bassalo), ramsey.zeitoun@gmail.com (R.I. Zeitouna), rtg@colorado.edu (R.T. Gill).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.ymben.2015.09.013
1096-7176/

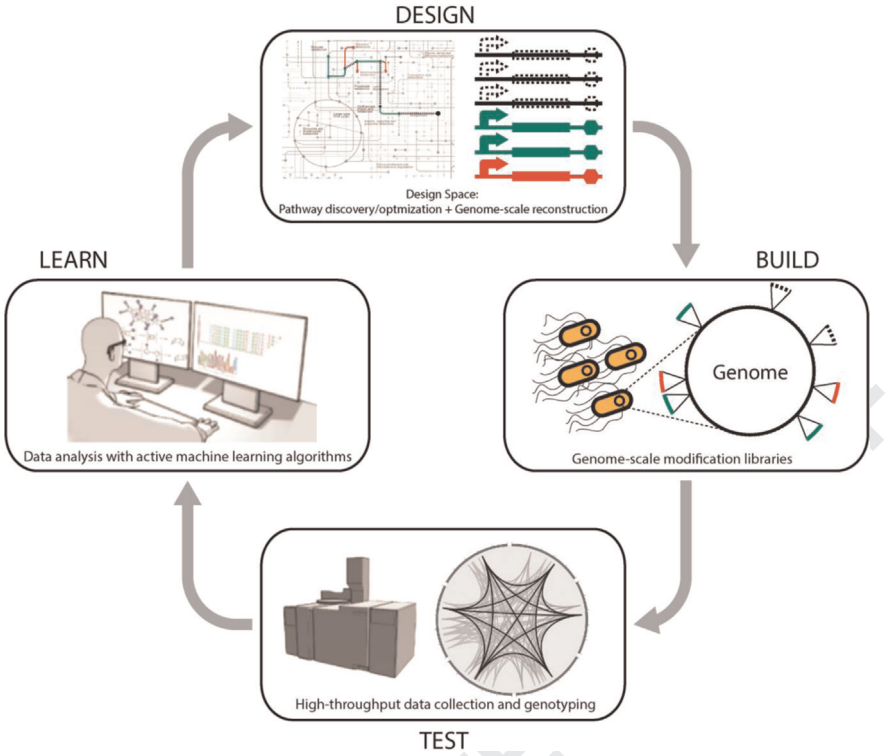

Fig. 1. The DBTL cycle applied to synthetic biology.

different model strains. As an example, *E. coli*'s genome-scale metabolic network models have been updated over 20 years (McCloskey et al., 2013; Orth et al., 2011). These databases are critical to improve the accuracy of the prediction of cellular phenotypes. More than 100 genome-scale metabolic network models were constructed for a wide range of different microorganisms, including *Saccharomyces cerevisiae* (Förster et al., 2003), *Corynebacterium glutamicum* (Shinfuku et al., 2009), *Mannheimia succiniciproducens* (Kim et al., 2007), *Bacillus subtilis* (Henry et al., 2009), *Clostridium acetobutylicum* (Lee et al., 2008), *Clostridium beijerinckii* (Milne et al., 2011), *Lactococcus lactis* (Flahaut et al., 2013), *Pichia pastoris* (Sohn et al., 2010), *Pseudomonas putida* (Puchałka et al., 2008), and so on. Recently, the ensemble modeling (EM) approach has shown promise in capturing kinetic and regulatory effects in the modeling of metabolic networks in comparison to FBA (Tran et al., 2008). It can simultaneously consider alternative model structures and parameter sets, such as identifying genetic/enzyme perturbations to minimize the number of models retained in the ensemble after each round of model screening (Zomorrodi et al., 2013). Ensemble Modeling for Robustness Analysis (EMRA), which combines a continuation method with the Ensemble Modeling approach, can be used for investigating the robustness of non-native pathways. By comparing possible designs of two nonnative pathways (non-oxidative glycolysis and reverse glyoxylate cycle), EMRA resulted in the selection of targets for flux improvement by considering both performance and robustness (Lee et al., 2014).

A number of algorithms based on the above genome-scale models have been developed to identify network manipulation strategies while predicting their system-wide effects (Table 1). Q4 OptKnock (Choon et al., 2014) is one popular computational algorithm, capable of suggesting gene deletion strategies that lead to the overproduction of a target metabolite. A nested optimization framework identifies gene deletions targets considering both the production of the desired compounds and biomass formation. OptKnock was applied to develop strategies for the metabolic engineering of *E. coli* for the production of 1,4-butanediol (BDO), leading to a strain capable of producing 18 g/L BDO from renewable carbohydrate feedstocks. Beyond gene knockouts, the design of strains involving overexpression and down-regulation have also been shown to enhance biochemical production by computational algorithms. OptForce contrasts the metabolic flux patterns observed in a parent strain and a strain overproducing the chemical at the targeted yield (Ranganathan et al., 2010). By applying the OptForce Q5 algorithm, the effect of redirecting malonyl-CoA flux towards resveratrol production was evaluated, and shake flask experiments yielded 1.6 g/L of resveratrol without the need of using expensive inhibitors of fatty acid metabolism (Bhan et al., 2013). Q6 Q7 Q8

Every predicted mutation should be associated with a specific design and measured effect on metabolism. However, to fully learn microbial metabolism and its responses to environmental factors, it is necessary to functionally characterize and accurately quantify all levels of gene products, mRNAs, proteins and metabolites, as well as their interaction. These requirements led to the generation of omics platform techniques, such as transcriptomics (Sorek and Cossart, 2010), proteomics (Otto et al., 2012), metabolomics (Hou et al., 2012) and interactomics (Janga et al., 2011). However, these techniques also generate a substantial amount of data that is hard to process and analyze for functional patterns. Several tools for Omics data analysis have been developed, such as GIMME(Becker and Palsson, 2008), E-Flux(Colijn et al., 2009), TIGER(Jensen et al., 2011), GIMMEp(Bordbar et al., 2012) (Table 1). GIMME produces a guaranteed functional metabolic model specific to transcriptomics data and quantifies the agreement between gene expression data and one or more metabolic objectives, which can be used for adaptive evolution of bacteria and rational design of metabolic engineering strains. Furthermore, by integrating proteomics and metabolomics data, GIMMEp and GIM^{3}E methods were developed based on the GIMME.

Machine-learning methods, instead, seek to use intrinsic data structure, as well as the expert annotations of biologists to infer models that can be used to solve versatile data analysis tasks.