Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Advances in analytical tools for high throughput strain engineering

Esteban Marcellin[1,3] and Lars Keld Nielsen[1,2,3]

The emergence of inexpensive, base-perfect genome editing is revolutionising biology. Modern industrial biotechnology exploits the advances in genome editing in combination with automation, analytics and data integration to build high-throughput automated strain engineering pipelines also known as biofoundries. Biofoundries replace the slow and inconsistent artisanal processes used to build microbial cell factories with an automated design–build–test cycle, considerably reducing the time needed to deliver commercially viable strains. Testing and hence learning remains relatively shallow, but recent advances in analytical chemistry promise to increase the depth of characterization possible. Analytics combined with models of cellular physiology in automated systems biology pipelines should enable deeper learning and hence a steeper pitch of the learning cycle. This review explores the progress, advances and remaining bottlenecks of analytical tools for high throughput strain engineering.

Addresses
[1] Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, St. Lucia 4072, Australia
[2] Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark
[3] Queensland Node of Metabolomics Australia, The University of Queensland, St. Lucia 4072, Australia

Corresponding author: Nielsen, Lars Keld (lars.nielsen@uq.edu.au)

Current Opinion in Biotechnology 2018, **54**:33–40

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

Edited by **Hiroshi Shimizu** and **Fumio Matsuda**

https://doi.org/10.1016/j.copbio.2018.01.027

Introduction

The past decade has seen several metabolic engineering projects progress to commercial bioprocesses for the sustainable, green production of chemicals [1,2]. In parallel, the process of engineering microbial cell factories has transformed from artisanal to industrial, with the establishment of biofundries employing the design–build–test cycle (Figure 1). Powered by inexpensive, base perfect genome editing technology and automation [3], the field is rapidly moving towards standardisation, big data and robust operation.

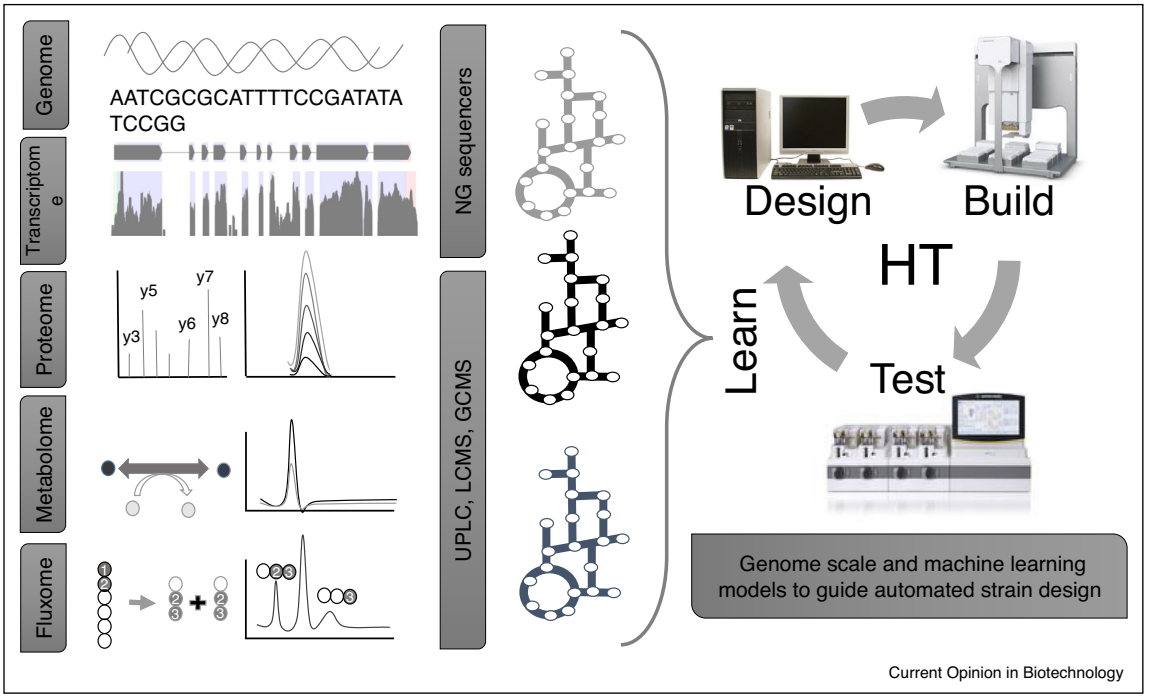
Biofoundries transform the speed and cost of industrial strain design [2,4•]. Until recently, it took at least 5 years, 100 person years and US$50 million to develop a commercial strain. Now, leading companies can move from product ideas to commercial levels in record time with minimal human interference, a concept unimaginable only a few years ago. Amyris, for example, has completed phase one of their DARPA collaboration, building 400 strains for 400 different molecules in record time. At the same time, Ginkgo Bioworks and Zymergen are growing rapidly and partnering with the top chemical companies. Academic initiatives such as iBioFAB [5••], the MIT-Broad Foundry, SynbiCITE at Imperial College, the Edinburgh genome foundry and NUS Synthetic Biology Foundry in Singapore, have all recently established, thus bringing strain engineering capabilities to the academic world.

The initial focus has been on designing strain faster, cheaper and more accurately. The strains produced are typically validated by genomics and characterized by growth pattern and product accumulation. While machine learning (ML) is used to develop statistical models from these data and artificial intelligence (AI) is used to guide the design process, it is inherently difficult to extract meaningful learnings about complex biological systems from design and superficial characterization data only. A comprehensive *omics* characterization of strains would provide a far richer dataset to guide subsequent design. Recent advances in high-throughput analytics promise to deliver quantitative *omics* data at a lower costs. With efficient integration of these data, it should be possible to design commercial strains with fewer iterations and hence at an overall lower cost. This review explores the frontiers of analytics that can complement high-throughput strain design to create scientific opportunities for the near future and contribute to the emerging bioeconomy [6].

Automation in cell factory design

Modern strain design starts on a computer, and subsequent automation occurs in the design, assembly and characterization of the strain (Figure 2). Even searching the literature to guide the design is fully automated. For example, Semantic Scholar is a search engine that in addition to words, extract graphs and influential citations. Iris.AI is a browsing tool that explores papers by concept

Figure 1



Biofoundries are automated platforms for strain design integrating biology with software and robots at high-throughput (HT). Computer-aided design is used to develop strains layouts that are executed on robots through computer-aided manufacturing. The strains are tested using automated fermentation and analytics and the data returned to the designer. Through iterations of design–build–test cycles, machine learning algorithms interpret large datasets to achieve deep learning of the biosystems. The learning achieved through DNA sequencing, mRNA sequencing, protein quantification and fluxomics is far deeper through the computer than humanly possible. Technological advances in biology and incorporation with AI algorithms are opening up the possibility of automating, complex, non-routine cognitive tasks. The learning, deeply enhanced through very large datasets from the *omics* datasets, can accelerate the learning and design to advance understanding of biological systems.

and can be used to predict missing citations in papers. At the end of the pipeline, Nutonian's software Eureqa interprets genetic and biochemical results by formulating mathematical theories that explain patterns in very large datasets.

A cellular design to produce a non-native molecule requires designing heterologous pathways and is often guided through metabolic network reconstructions [7] and strain engineering software and algorithm [8–10]. Pathways not existing in nature are also designed *in silico* using, for example, Biochemical Network Integrated Computational Explorer (BNICE), which predicts enzymatic steps required to convert a given substrate to a desired molecule [11] ranking pathways by thermodynamics, length and yield [12]. Similarly, M-Path builds a synthetic pathway using an iterative approach to search enzymatic reactions from databases [13] and eXTended Metabolic Space (XTMS) predicts novel pathways that can be introduced into *E. coli* [14].

The experimental design can be automated using for example Double Dutch, which uses combinatorial libraries of pathway variants to reduce human effort and increase the efficiency of part design [15]. Automation in the building cycle includes tools such as J5 or Raven, which guide simultaneous assembly of multiple DNA segments [16,17] and are compatible with public libraries of promoters, terminators and RBS [18••]. Once the pathway has been designed, liquid handling robots can build, transfer (using CRISPR/CAS9 or other methods) and select clones automatically (Figure 2).

Industrial biofoundries are capable of constructing hundreds of production strains per week each strain containing an assembly of up to 20 new genes. The strains are characterized in an automated culture system and one or more samples collected per strain for molecular characterization. Integrated, high-throughput analytics is critical to close the loop. The depth (and cost) of characterization must be weighed against the cost of strain construction and the expected gain from analytics, which changes across the engineering cycle. Early in the cycle, when identifying good parts is critical, a simple concentration measure may be adequate. Later, when the highly expressed heterologous pathway interacts with cellular