We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Introductory Chapter: From BioBricks to Synthetic Genomes

Luis Humberto Reyes and Miguel Fernández-Niño

1. Introduction

One of the goals of Synthetic Biology is to design novel biological systems by the rational assembly of biological parts (BioBricks) into artificial metabolic networks [1–4]. These engineered networks can be further coupled to create highly complex biological systems [3]. Such systems are arranged in living organisms that work as biological chassis to hold and express the engineer networks and produce a desired phenotype [4–7]. Several technologies have been developed during the last 20 years to improve our ability to engineer novel biological systems. This has been observed at different levels of complexity, including the development of technologies for the identification, design, and synthesis of BioBricks and the expansion of innovative protocols for pathway assembly/modeling and genome-editing technologies in the fine-tuning of biological systems (**Figure 1**).

Currently, BioBricks can be identified (mined) from the large amount of information contained in public databases (e.g., BRENDA [8], GenBank [9], PANTHER [10], UniProt [11], etc.) or by selecting pre-designed BioBricks from specialized databases such as the iGEM Parts Registry [12] and the BioMaster DataBase [13]. These engineered BioBricks can be obtained by traditional methods, including PCR [14], or by using cutting-edge technologies such as the *de novo* synthesis of large fragments of nucleic acids (synthetic genes [15]), whose prices have been remarkably decreasing during the last years. The selected BioBricks can be further engineered to optimize their heterologous expression and ease subsequent assembly, expression, and purification [2, 16]. Different protein tags have been designed for this purpose, including solubility and affinity tags and tags aimed to simplify subsequent cloning in expression vectors [17]. Once the desired BioBricks have been obtained, they are usually coupled to other BioBricks to create artificial networks of higher complexity [3, 18]. Several technologies have been developed to assemble multiple BioBricks in artificial networks, including PCR-dependent cloning [19], seamless cloning [20], recombinational approaches [21], among others.

The genetic expression of these engineered artificial networks can be further optimized by modifying the network topology (e.g., changing from operon to monocistronic topologies [20]) or by adding regulatory elements such as feedback loops, oscillators, riboswitches, and protein scaffolds [3]. The behavior of these artificial networks and their regulatory elements can be studied through *in silico* modeling to predict the most appropriate topology for the artificial system to be designed. Several tools and software for *in silico* modeling of metabolic pathways have been designed to determine the effect of the expression of a particular artificial network on the global metabolic response of the host organism [22–26]. For example, it is currently possible to identify genes in the host organism through *in silico* modeling, whose deletion may result in a higher expression of the desired phenotype or even predict the most efficient set of reactions required to produce valuable compounds

Figure 1. Frequently used steps to engineer artificial biological systems and synthetic genomes using BioBricks as building blocks.

using reconstructed metabolic models [24, 25]. The predictive ability of these models relies on our understanding of the genome, transcriptome, and the global metabolism of the selected host. Thus, it is not surprising that most of the host organisms used in synthetic biology are widespread model organisms such as well-characterized strains of *Escherichia coli* and *Saccharomyces cerevisiae*, for instance. These organisms have been studied for generations, and different toolboxes for their metabolic engineering/genome editing have been previously designed [27].

Interestingly, with the reduction of sequencing prices and the development of novel methodologies for long-read sequencing (e.g., Oxford Nanopore [28] and PacBio technologies [29]), the genome of a large number of new (so far not characterized) organisms have been recently elucidated. This has been concurrent with the development of new technologies for the identification, characterization, and quantification of metabolites, proteins, and lipids using last-generation liquid and gas chromatographic columns coupled to mass spectrometry (LC-MS and GC-MS) and the development of new devices for nuclear magnetic resonance (NMR) analytics [30–32]. In addition, novel technologies for improving genome editing such as CRISPR-Cas9 have opened the possibility to expand our ability to engineer novel biological systems in living cells, as never before [33, 34]. Thus, the opportunity to design and construct an entire genome is now a reality with the current technological advances. This has opened a new field of research (known as Synthetic Genomics [35]) to engineer and assemble entire artificial genomes or larger parts of genomes in living organisms by using the principles of synthetic biology previously summarized. A genome is considered synthetic if the building blocks used for its assembly were originated by chemical synthesis [36].

2. Current advances in Synthetic Genomics

“One needs to write synthetic DNA sequences in order to better understand the grammar of life”-with this sentence, Schindler and co-workers, in their 2018 review, accurately summarized the primary motivation behind the fast progress of synthetic genomics [36]. Accordingly, one of the main goals of synthetic genomics is to improve our understanding of genome fundamentals [35, 36]. Thus, the construction of entire synthetic genomes has allowed the study of their constitutive building blocks, considering the effect of the whole biological system on biological function.

As previously described in this chapter, synthetic genomes are engineered following a hierarchical and modular assembly starting from synthetic genes, gene clusters, artificial metabolic pathways, and chromosomes. Currently, two approaches can be utilized to assemble a synthetic genome into an organism: (1) using a heterologous host or (2) using a native host as a chassis for chromosome replacement [35, 36]. Heterologous hosts are well-known model organisms with an extensive toolbox for genetic engineering that simplify the subsequent assembly (e.g., *E. coli* and *S. cerevisiae*). However, their engineering capacity might be restricted by the size and number of synthetic chromosomes to be assembled. On the other hand, native hosts are advantageous for chromosome replacement. Still, most of them are not well-characterized organisms, or there is a lack of tools for their genetic engineering. Currently, many native microorganisms have been characterized in response to the fast development of sequencing and genome editing techniques as previously described [27–29].

Most of the synthetic genomes that have been successfully assembled are viral and bacterial, with a smaller genome size than eukaryotes. For example, the viral genome of the Poliovirus (7.5 kb size) was entirely synthesized 20 years ago [37] with the technology available at the time. More than 15 years later, a fully synthetic genome was assembled for the Horsepox virus (212 kb size) using the latest technology, which allowed the assembly of a synthetic genome that is more than 20 times bigger as compared to Polio genome size [38]. Remarkably, Thao et al. (2020) have recently engineered and assembled the entire genome of the virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) responsible for the current pandemic disease (COVID-19) using an *S. cerevisiae* platform [39]. This heterologous platform for the synthetic assembly of viral genomes constitutes a valuable tool to reconstruct different RNA viruses (from different families) in a short time using pre-designed synthetic building blocks [39]. Regarding the reconstruction of synthetic bacterial genomes, most of the research has been focused on species from the genus *Mycoplasma* with genome sizes ranging from 583 kb (in *Mycoplasma genitalium*) to 1079 kb (in *Mycoplasma mycoides*) [40, 41]. It is important to mention that up to now, there are no eukaryotic genomes that have been entirely reconstructed. However, there are still several efforts to engineer eukaryotic genomes, such as the Sc2.0 international project (Synthetic Yeast), aimed to assemble the world first eukaryotic synthetic genome from scratch and the Genome Project-write (GP-write) that is an international project aimed to reconstruct the entire genome of a large number of cell lines within the next years.

Currently, the research in synthetic genomics has moved one step forward to design new genomes that are different from the ones found in nature [35, 36]. Thus, the synthetic genome of *M. mycoides* has been reduced in size by 49.3%, which resulted in a new platform to discover new biological functions [42]. Similarly, a synthetic *E. coli* genome is currently being developed [43] to remove/replace codons in the codon sequence that allow the expression of proteins with non-natural amino acids. These examples show the relevance of synthetic genomics as a promising area to explore novel biological functions using completely unnatural biological

systems. The research in this area is expected to increase in response to developing novel technologies and bioinformatics tools to rational design and analyze complex biological systems. Consequently, new platforms must be characterized to increase the number of organisms used as a host for chromosome assembly. New bioinformatics tools (Genomic design software) must still be designed to predict and study the behavior of the engineered synthetic genomes. Finally, it is important to mention that this emerging area has not only the potential to boost science but also can be used for harmful purposes. Consequently, a deeper discussion on synthetic genomics' ethical, social, and ecological consequences must still be conducted with scientists, politicians, and communities.

3. Impact and risks

The intentional or accidental release of genetically modified organisms into the environment could have significant negative impacts on both human and environmental health. This biological revolution, together with advances in biotechnology, could be used to improve the biological properties of viruses simply by altering resistance to antiviral agents, modifying antigenic properties, modifying the tropism, pathogenesis, and transmissibility of tissues, "humanizing" zoonotic viruses, and creating designer super-pathogens. The main paradigm shift may be that the approach is less technically demanding and more design-based, requiring only limited technical expertise because the genome can be synthesized and purchased from commercial vendors, government-sponsored facilities, or from rogue basement operations (e.g., bioterrorist sponsored organizations or private entrepreneur). The main technical support could include a competent research technician and minimal equipment to isolate recombinant pathogens from recombinant DNAs.

These potential impacts require governance methods and research guidelines that promote their ethical and responsible use. Under the precautionary principle, a rigorous risk assessment and inclusion of diverse stakeholder perspectives should be applied in the development and management of innovative synthetic biology applications and products. The precautionary principle states that when human activities can lead to unacceptable harm that is scientifically plausible but uncertain, steps must be taken to avoid or lessen that harm.

IntechOpen

Author details

Luis Humberto Reyes¹ and Miguel Fernández-Niño^{2*}

¹ Grupo de Diseño de Productos y procesos (GDPP), Departamento de Ingeniería Química y de Alimentos, Universidad de los Andes, Bogotá, Colombia

² Department of Bioorganic Chemistry, Leibniz-Institute of Plant Biochemistry, Halle, Germany

*Address all correspondence to: mfernand@ipb-halle.de

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] Leggieri PA, Liu Y, Hayes M, Connors B, Seppälä S, Malley MAO, et al. Integrating systems and synthetic biology to understand and engineer microbiomes. *Annual Review of Biomedical Engineering*. 2021;**23**:169-201
- [2] Hughes RA, Ellington AD. Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harbor Perspectives in Biology*. 2017;**9**:a023812
- [3] Agapakis C, Silver P. Synthetic biology: Exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*. 2009;**5**:704-713. DOI: 10.1039/b901484e
- [4] Singh V. Recent advancements in synthetic biology: Current status and challenges. *Gene*. 2014;**535**:1-11. DOI: 10.1016/j.gene.2013.11.025
- [5] Calero P, Nikel PI. Minireview chasing bacterial chassis for metabolic engineering: A perspective review from classical to non-traditional microorganisms. *Microbial Biotechnology*. 2019;**12**:98-124. DOI: 10.1111/1751-7915.13292
- [6] Yu D, Wang M, Zhu G, Ge B, Liu S, Huang F. Enhanced photocurrent production by bio-dyes of photosynthetic macromolecules on designed TiO₂ film. *Scientific Reports*. 2015;**5**:9375
- [7] Liu J, Wu X, Yao M, Xiao W. Chassis engineering for microbial production of chemicals: From natural microbes to synthetic organisms. *Current Opinion in Biotechnology*. 2020;**66**:105-112. DOI: 10.1016/j.copbio.2020.06.013
- [8] Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, et al. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research*. 2021;**49**:498-508. DOI: 10.1093/nar/gkaa1025
- [9] Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Research*. 2016;**44**:67-72. DOI: 10.1093/nar/gkv1276
- [10] Mi H, Ebert D, Muruganujan A, Mills C, Albou L, Mushayamaha T, et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*. 2021;**49**:394-403. DOI: 10.1093/nar/gkaa1106
- [11] The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021;**49**:480-489. DOI: 10.1093/nar/gkaa1100
- [12] Smolke CD. Building outside of the box: iGEM and the BioBricks foundation. *Nature Biotechnology*. 2009;**27**:1099-1102. DOI: 10.1038/nbt1209-1099
- [13] Wang B, Yang H, Sun J, Dou C, Huang J, Charles TC. BioMaster: An integrated database and analytic platform to provide comprehensive information about BioBrick parts. *Frontiers in Microbiology*. 2021;**12**:1-6. DOI: 10.3389/fmicb.2021.593979
- [14] Shetty RP, Endy D, Knight TF Jr. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*. 2008;**12**:1-12. DOI: 10.1186/1754-1611-2-5
- [15] Kosuri S, Church GM. Large-scale de novo DNA synthesis: Technologies and applications. *Nature Methods*. 2014;**11**:499-507. DOI: 10.1038/nmeth.2918
- [16] Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG. Expression of codon optimized genes in microbial

- systems: Current industrial applications and perspectives. *Frontiers in Microbiology*. 2014;5:1-8. DOI: 10.3389/fmicb.2014.00021
- [17] Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. *Current Protocols in Protein Science*. 2013;73:1-23. DOI: 10.1002/0471140864.ps0909s73
- [18] Young R, Haines M, Storch M, Freemont PS. Combinatorial metabolic pathway assembly approaches and toolkits for modular assembly. *Metabolic Engineering*. 2021;63:81-101. DOI: 10.1016/j.jymben.2020.12.001
- [19] Tillett D, Neilan BA. Enzyme-free cloning: A rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Research*. 1999;27:26-28
- [20] Xu P, Vansiri A, Bhan N. ePathBrick: A synthetic biology platform for engineering metabolic pathways in *E. coli*. *ACS Synthetic Biology*. 2012;7:256-266
- [21] Park J, Throop AL, Labaer J. Site-specific recombinational cloning using gateway and in-fusion cloning schemes. *Current Protocols in Molecular Biology*. 2015;110:1-23. DOI: 10.1002/0471142727.mb0320s110
- [22] Kumar VS, Maranas CD. GrowMatch: An automated method for reconciling in silico/in vivo growth predictions. *PLoS Computational Biology*. 2009;5:18-20. DOI: 10.1371/journal.pcbi.1000308
- [23] Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D. ClustScan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research*. 2008;36:6882-6892. DOI: 10.1093/nar/gkn685
- [24] Pharkya P, Burgard AP, Maranas CD. OptStrain: A computational framework for redesign of microbial production systems. *Genome Research*. 2004;14:2367-2376. DOI: 10.1101/gr.2872004
- [25] Burgard AP, Pharkya P, Maranas CD. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*. 2003;84(6):647-657. DOI: 10.1002/bit.10803
- [26] Fong SS, Palsson BØ, Herrga MJ. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Computational Biology*. 2006;2:e72. DOI: 10.1371/journal.pcbi.0020072
- [27] Redden H, Morse N, Alper HS. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Research*. 2015;15:1-10. DOI: 10.1111/1567-1364.12188
- [28] Jain M, Olsen HE, Paten B, Akeson M. The Oxford nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*. 2016;17:1-11. DOI: 10.1186/s13059-016-1103-0
- [29] Xie H, Yang C, Sun Y, Igarashi Y, Jin T, Luo F. PacBio long reads improve metagenomic assemblies, gene catalogs, and genome binning. *Frontiers in Genetics*. 2020;11:1077. DOI: 10.3389/FGENE.2020.516269/BIBTEX
- [30] Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, et al. Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics*. 2018;14:1-31. DOI: 10.1007/s11306-018-1449-2
- [31] Wang Y, Hui S, Wondisford FE, Su X. Utilizing tandem mass

spectrometry for metabolic flux analysis. *Laboratory Investigation*. 2020;**101**:423-429. DOI: 10.1038/s41374-020-00488-z

[32] Li D, Gaquerel E. Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. *Annual Review of Plant Biology*. 2021;**72**:867-891

[33] Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation*. 2014;**124**:4154-4161. DOI: 10.1172/JCI72992. transcription

[34] Nidhi S, Anand U, Oleksak P, Tripathi P, Lal JA, Thomas G, et al. Novel CRISPR-Cas systems: An updated review of the current achievements, applications, and future research perspectives. *International Journal of Molecular Sciences*. 2021;**22**:1-42

[35] Wang L, Jiang S, Chen C, He W, Wu X, Wang F, et al. Synthetic genomics: From DNA synthesis to genome design. *Angewandte Chemie (International Ed. in English)*. 2018;**57**:1748-1756. DOI: 10.1002/ANIE.201708741

[36] Schindler D, Dai J, Cai Y. Synthetic genomics: A new venture to dissect genome fundamentals and engineer new functions. *Current Opinion in Chemical Biology*. 2018;**46**:56-62. DOI: 10.1016/J.CBPA.2018.04.002

[37] Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science*. 2002;**297**:1016-1018. DOI: 10.1126/SCIENCE.1072266

[38] Koblentz GD. The de novo synthesis of horsepox virus: Implications for biosecurity and recommendations for preventing the reemergence of smallpox. *Health Security*. 2017;**15**: 620-628. DOI: 10.1089/HS.2017.0061

[39] Thao TTN, Labroussaa F, Ebert N, V'kovski P, Stalder H, Portmann J, et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*. 2020;**582**:561-565. DOI: 10.1038/s41586-020-2294-9

[40] Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*. 2008;**319**:1215-1220. DOI: 10.1126/SCIENCE.1151721

[41] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*. 2010;**329**:52-56. DOI: 10.1126/SCIENCE.1190719

[42] Hutchison CA, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, et al. Design and synthesis of a minimal bacterial genome. *Science*. 2016;**351**:aad6253. DOI: 10.1126/SCIENCE.AAD6253

[43] Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, et al. Design, synthesis, and testing toward a 57-codon genome. *Science*. 2016;**353**:819-822. DOI: 10.1126/SCIENCE.AAF3639