Shailza Singh *Editor*

Systems Biology Application in Synthetic Biology

Springer

Systems Biology Application in Synthetic Biology

Shailza Singh
Editor

Systems Biology Application in Synthetic Biology

Springer

Editor
Shailza Singh
Computational and Systems Biology Lab
National Centre for Cell Science
Pune, India

ISBN 978-81-322-2807-3 ISBN 978-81-322-2809-7 (eBook)
DOI 10.1007/978-81-322-2809-7

Library of Congress Control Number: 2016952540

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer (India) Pvt. Ltd.

Preface

Systems and synthetic biology is an investigative and constructive means of understanding the complexities of biology. Discovery of restriction nucleases by Werner Arber, Hamilton Smith, and Daniel Nathans in 1978 revolutionized the way DNA recombinant constructs were made and how individual genes were analyzed for its function and vitality. It also opened the doors to a new era of "synthetic biology" where apart from analysis and description of existing gene, new gene arrangements can be constructed and evaluated. Since then, synthetic biology has emerged from biology as a distinct discipline that quantifies the dynamic physiological processes in the cell in response to a stimulus. Switches, oscillators, digital logic gates, filters, modular – interoperable memory devices, counters, sensors, and protein scaffolds are some of the classic design principles based on which many more novel synthetic gene circuits can be created with possible application in biosensors, biofuels, disease diagnostics, and therapies. Most of these gene networks combine one or more classes of controller components, such as conditional DNA-binding proteins, induced-protein dimerization, RNA controllers, and rewired cell-surface receptors, to modulate transcription and translation that alters protein function and stability.

An iterative design cycle involving molecular and computational biology tools can be capitalized to assemble designer devices from standardized biological components with predictable functions. Research efforts are priming a variety of synthetic biology inspired biomedical applications that have the potential to revolutionize drug discovery and delivery technologies as well as treatment strategies for infectious diseases and metabolic disorders. The building of complex systems from the interconnection of parts or devices can be significantly facilitated by using a forward-engineering where various designs are first optimized, tested *in silico* and their properties are assessed using mathematical analysis and model-based computer simulations. Mathematical models using Ordinary Differential Equations (ODEs), Partial Differential Equations (PDEs), Stochastic Differential Equations (SDEs), or Markov Jump Processes (MJPs) are typically used to model simple synthetic biology circuits. Thus use of computation in synthetic biology can lead us to ways that help integrate systems models to support experimental design and engineering. Synthetic biology has significantly advanced our understanding of complex control dynamics that program living systems. The field is now starting to tackle relevant therapeutic challenges and provide novel diagnostic tools as well as unmatched therapeutic strategies for treating significant

human pathologies. Although synthetic biology-inspired treatment concepts are still far from being applied to any licensed drug or therapy, they are rapidly developing toward clinical trials. Nevertheless, it has provided insights into disorders that are related to deficiencies of the immune system known for its complex control circuits and interaction networks.

Novel-biological mechanism may also be coupled with image-modeling approach to be verified in *in vitro* conditions. Computational techniques can be used in tandem with image analysis to optimally characterize mammalian cells, leading to results that may allow scientists to uncover mechanisms on a wide range of spatio-temporal scales. These elucidated methods and principles used in *in silico* hypotheses generation and testing have the potential to catalyze discovery at the bench. Despite considerable progress in computational cell phenotyping, significant obstacles remain with the magnitude of complexity with experimental validation at the bench. The true power of computational cell phenotyping lies in their strengths to generate insights toward *in vivo* constructs, which is a prerequisite for continued advancements. None of the obstacles is insurmountable. However, advances in imaging and image processing may transcend current limitations which may unlock a wellspring of biological understanding, paving the way to novel hypotheses, targeted therapies, and new drugs. Additionally, phenotyping permits the effects of compounds on cells to be visualized immediately without prior knowledge of target specificity. By harnessing the wealth of quantitative information embedded in images of *in vitro* cellular assays, HCA/HCS provides an automated and unbiased method for high-throughput investigation of physiologically relevant cellular responses that is clearly an improvement over HTS methods, allowing significant time and cost savings for biopharmaceutical companies. The emergence of non-reductionist systems biology aids in drug discovery program with an aim to restore the pathological networks. Unbalance reductionism of the analytical approaches and drug resistance are some of the core conceptual flaws hampering drug discovery. Another area developing and envisaged in this book is system toxicology, which involves the input of data into computer modeling techniques and use differential equations, network models, or cellular automata theory. The input data may be biological information from organisms exposed to pollutants. These inputs are data mostly from the "omics," or traditional biochemical or physiological effects data. The input data must also include environmental chemistry data sets and quantitative information on ecosystems so that geochemistry, toxicology, and ecology are modeled together. The outputs could include complex descriptions of how organisms and ecosystems respond to chemicals or other pollutants and their inter-relationships with the many other environmental variables involved.

The model outputs could be at the cellular, organ, organism, or ecosystem level. Systems toxicology is potentially a very powerful tool, but a number of practical issues remain to be resolved such as the creation and quality assurance of databases for environmental pollutants and their effects, as well as user-friendly software that uses ecological or ecotoxicological parameters and terminology. Cheminformatics and computational tools are discussed in lengths which help identify potential risks including approaches for building

quantitative structure activity relationships using information about molecular descriptors. The assimilation of chapter from various disciplines includes the trade-offs and considerations involved in selecting and using plant and other genetically engineered crops. Systems biology also aid in understanding of plant metabolism, expression, and regulatory networks. Synthetic biology approaches could benefit utilizing plant and bacterial "omics" as a source for the design and development of biological modules for the improvement of plant stress tolerance and crop production. Key engineering principles, genetic parts, and computational tools that can be utilized in plant synthetic biology are emphasized.

The collection of chapters represents the first systematic efforts to demonstrate all the different facets of systems biology application in synthetic biology field.

I would like to thank Mamta Kapila, Raman Shukla, Magesh Karthick Sundaramoorthy, and Springer Publishing group for their assistance and commitment in getting this book ready for publication. I would also like to thank my wonderful graduate students Vineetha, Milsee, Pruthvi, Ritika, Bhavnita, and Dipali for being a rigorous support in the entire endeavor. Finally, I would especially like to thank my family, Isha and Akshaya, my parents for being patient with me during the process. Without their love and support, this book would not have been possible.

Pune, India Shailza Singh

Contents

About the Editor

Shailza Singh is working as Scientist D at National Centre for Cell Science, Pune. She works in the field of Computational and Systems Biology wherein she is trying to integrate the action of regulatory circuits, cross-talk between pathways and the non-linear kinetics of biochemical processes through mathematical models. The current thrust in her laboratory is to explore the possibility of network-based drug design and how rationalized therapies may benefit from Systems Biology. She is the recipient of RGYI (DBT), DST-Young Scientist and INSA (Bilateral Exchange Programme). She is the reviewer of various international and national grants funded from government organizations.

Microbial Chassis Assisting Retrosynthesis

1

Milsee Mol, Vineetha Mandlik, and Shailza Singh

1.1 Introduction

It's a well-known and a documented fact that life has arisen from simple molecules. Therefore the main stay of research in biology is to strip down the inherent complexity associated due to the interaction between these simple molecular assemblies. During the course of evolution, there has been a reduction in the complexity that constitutes the essential features of a living cell. The comprehension (if it is possible to comprehend fully) of the underlying complexities will not only allow us to understand the key regulatory mechanism in numerous diseases, production of important metabolites, etc. but also help us to build a reliable mathematical model for formulating future scientific enquiry. A better understanding of cellular systems can be done via two competing routes the "bottom up" as well as "top-down" synthetic biology approach. Synthetic biology has two goals: to re-engineer existing systems for better quantitative understanding; and, based on this understanding engineer new systems that do not exist in nature [1]. The fundamental principle of synthetic biology is similar to constructing non-biological system e.g. a computer, by putting together composite, well-characterized modular parts. It is an interdisciplinary science drawing expertise from biology, chemistry, physics, computer science, mathematics and engineering [2].

Synthetic biology has re-revolutionized the way biology is done today in laboratories across the globe, also mainly because of the way DNA the blue print of a cells functionality is being synthetized by simply providing the desired sequence to the automated synthesizer. Synthetic biologists are now on the verge of developing 'artificial life' that has enormous applications in biotechnology apart from the fact that it is being used to now understand the origin of life. The 'top-down' approaches in synthetic biology are being used to synthesize the minimal cells by systematically reducing the genome of a cell such that it shows a desired function under environmentally favourable conditions [3, 4]. Successful chemical synthesis of genome and its transfer to the bacterial cytoplasm [5] reveals the power of synthetic biology framework to create a minimal cell for greater application in biotechnology [2]. Such a minimal cell having the minimum required genome could serve as a "chassis" that can be further expanded with the addition of genes for specific functions desired from a tailor-made organism. Further a streamlined chassis based on a minimal genome can simplify the interaction

M. Mol • V. Mandlik
National Centre for Cell Science, NCCS Complex, Ganeshkhind, Pune 411007, India

S. Singh (✉)
Computational and Systems Biology Lab, National Centre for Cell Science, NCCS Complex, Ganeshkhind, Pune 411007, India
e-mail: shailza_iitd@yahoo.com; singhs@nccs.res.in

S. Singh (ed.), *Systems Biology Application in Synthetic Biology*,
DOI 10.1007/978-81-322-2809-7_1

between the host and the system that may have relevance in minimizing the effect of the metabolic burden of the exogenous pathway placed in the cell [6]. Such extensively streamlining is possible for many of the medically and industrially important microorganism as their genomes have been already sequenced and assembled.

Comparative genomics is a useful methodology that delineates genes based on the conservedness of the genes to distant related species. It is based on the hypothesis that the conserved genes are certainly essential for cellular function and may be well approximated to the required minimal gene set [7]. But as more and more genomes are being sequenced there is divergence in the evolutionary tree showing that some of the essential functions can be performed by non-orthologous genes [8]. Therefore, gene persistence rather than gene essentiality should be taken into consideration for constructive way to identify the minimal universal functions supporting robust cellular life [9].

Another approach that of experimental gene inactivation identifies genes, those are important for the viability of the cell. Genome-scale identifications of such genes have been done using the prokaryotic as well as eukaryotic systems using strategies of massive transposon mutagenesis [10, 11], the use of antisense RNA [12] to inhibit gene expression and the systematic inactivation of each individual gene present in a genome [13, 14]. These genome scale identifications have been done under predefined experimental growth conditions. This kind of experimental identification helps us get a complete understanding of the relationship between genotype and phenotype which would facilitate the design of minimal cell [8].

The data generated in such genome scale experimental models is large which needs computer-assisted mathematical treatment to get some meaningful statistically valid approximations. Therefore mathematical models that relate the gene content (genotype) of a cell to its physiological state (phenotype) enables the simulation of minimal gene sets under various environmental growth conditions (constraint-based approach) [15–18]. Thus, *in silico*, with in the complex gene network reaction(s), each gene can be individually "deleted" (flux 'zero') and relate it to the biomass as the fitness function for the system [19]. This flux-based models yield key evolutionary insights on the minimal genome [20].

Integrating all the information from comparative genomics, experimentation and *in silico* predictions, a new approach of retrosynthesis is rising for building *de novo* pathways in host chassis [21–23]. Retrosynthesis is a technique routinely used in synthetic organic chemistry [24, 25], where it starts by conceptually defining the structure and properties of the desired molecule to be produced and working backward through known chemical transformations to identify a suitable precursor or sets of precursors. This approach when applied to biological metabolic transformation can identify the reactions involved and their corresponding enzymes. Thus, by enumerating the biochemical pathways, it can be linked to the final product in the host's metabolism [23].

With the available tool kits for designing biological systems, the future predictability is relatively difficult and may lead to bottleneck situation in the production pipeline. Metabolic pathway models are being made more predictable by incorporating the freedom to tweak the gene expression to achieve a particular flux of each metabolite in the reaction or pathway [26]. Tools that help in debugging bottleneck in the metabolic pathway would reduce development times for optimizing engineered cells. Functional genomic tools can serve this purpose [27], which helps in chalking out the over or under production of a protein/enzyme in the pathway that can lead to a stress response [28, 29]. The information from these tools can be rendered to diagnose the problem and modify expression of genes in the metabolic pathway to improve productivity. Taking advantage of the cell's native stress response pathways, too many desirable chemicals particularly at the high titres needed for industrial-scale production can be an effective way to overcome product toxicity [30].

1.2 Tools for Designing and Optimizing Synthetic Pathway

It is an uphill task to find an optimal solution for a selected pathway, enzymes or chassis organism from an abundance of possibilities. Engineering a synthetic pathway and uploading it into the chassis organism followed by optimizing the production of the desired product involves lot of experimental work which is accompanied by lots of permutation and combinations of conditions. To make life easy for a synthetic biologist powerful computational tools are a necessity. There are many computational tools that can lead for a better informed, rapid design and implementation of novel pathways in a selected host organism with the desired parts and flux of the desired product is listed in Table 1.1. These tools are based on criteria like pathway selection and thereafter ranking them. These prediction help to explore the pathways that are chemically versatile and also help compare their efficiencies as compared to the natural pathways. Organism selection for uploading the novel pathway depends on two approaches: First, choose an organism that already has most of the reactions involved in the pathway, thereby reducing the stochasticity that can be introduced due to the new enzymes in the metabolic network [23]. The second approach is to build genome scale models using constraint-based flux balance analysis. In this approach, steady-state flux distribution of the metabolic network is predicted based on the stoichiometry of each reaction, mass–balance constraints and an objective function specifying the fluxes of components that are to be optimized [31]. Once the prioritized pathway and optimum host is selected, the next step is to construct the pathway by using parts such as the RBS, promoters, terminators, etc. with the regulatory elements incorporated. A range of standardized and characterized parts are available at the parts registry [32]. Efforts are underway to increase the catalogue available at the registry, as they are suitable for finding regulatory elements rather than the coding sequences. Since the coding sequences for the enzymes are part of a specific synthetic pathway, they are not catalogued and for this purpose genome-mining is a crucial step. The last part of the process design is to synthesize the DNA parts that are codon optimized for the host chassis. Many variants of the basic DNA sequence can also be synthesized from which an efficient sequence can be picked up. After all the above steps are succesfully completed a functional design can be arrived to, which can then be inserted into the chromosome of the host genome [33] or as a multigene expression plasmid [34]. The workflow designing a synthetic pathway into a microbial chassis system can be depicted pictorial in Fig. 1.1.

1.3 Choosing a Host and Vector for Synthetic Pathway Construction

Choosing a correct heterologous host for the production of a desired product is an important and uphill task in metabolic engineering of microbes. A host must be chosen based on the fact whether the desired metabolic pathway already exists or can it be reconstituted in that host. If so, then the host can survive under the desired process conditions of pH, temperature, ionic strength, etc. for the optimum titre of the desired product. The host should be genetically robust and should not be susceptible to phage attacks and at the same time should be amenable to available genetic tools. Although *E. coli* can be treated with different genetic tools available, it has disadvantage of being susceptible to phage attack. The host should be able to grow on simple, inexpensive carbon sources without or with minimal additions to the process media, thereby reducing the production cost of the product [63, 64]. Another aspect that should be considered is the level of expression of the heterologous enzymes in the host strain. The enzymes should be expressed in amounts that are catalytically important for the conversion of the starting material to the desired product. Toxicity of the intermediate metabolites for the hosts should also be dealt with, because any intermediate that is toxic will have a profound effect on the final titres of the desired product.

Table 1.1 Computational tools currently being employed for synthetic pathway construction

Tool		Description
Pathway prediction	BNICE (Biochemical Network Integrated Computational Explorer) [35]	Identification of possible pathways for the degradation or production of a desired compound within a thermodynamic purview
	DESHARKY [36]	Best match pathway identification specific to a host; provides phylogenetically related enzymes
	RetroPath [37]	Retrosynthetic pathway design, pathway prioritization, host compatibility prediction, toxicity prediction and metabolic modelling
	FMM (From Metabolite to Metabolite) [38]	Finds an alternate biosynthetic routes between two metabolites within the KEGG database
	OptStrain [39]	Optimization of the host's metabolic network by suggesting addition or deletion of a reaction
Parts identification	Standard Biological Parts knowledgebase [40]	Knowledgebase with parts for easy computation; includes all the parts from Registry of Standard Biological Parts
	IMG (Integrated Microbial Genomes) [41]	Comparative and evolutionary analysis of microbial genomes, gene neighbourhood orthology searches
	antiSMASH [42]	Identification, annotation and comparative analysis of secondary metabolite biosynthesis gene clusters
	KEGG [43]	Database of organism specific collection of metabolite and metabolic pathway
Parts optimization and synthesis	RBS Calculator [44]	Automated design of RBSs based on a thermodynamic model of transcription initiation
	RBSDesigner [45]	Algorithm for prediction of mRNA translation efficiencies
	Gene Designer 2.0 [46], Optimizer [47],	Gene, operon and vector design, codon optimization and primer design
	DNAWorks [48], TmPrime [49]	Oligonucleotide design for PCR-based gene synthesis, with integrated codon optimization
	CloneQC [50]	Quality of sequenced clones by detecting errors in DNA synthesis
Pathway and circuit design	Biojade [51]	Software tool for design and simulation of genetic circuits
	Clotho [52]	Flexible interface for synthetic biological systems design; within the interface, a range of apps/plugins can be utilized to import, view, edit and share DNA parts and system designs
	GenoCAD [53]	CAD software that allows drag-and-drop drawing and simulation of biological systems
	Asmparts [54]	Computational tool that generates models of biological systems by assembling models of parts
	SynBioSS [55]	Designing, modelling and simulating synthetic genetic constructs
	CellDesigner [56]	Graphical drawing of regulatory and biochemical networks that can be stored in Systems Biology Markup Language (SBML)
Metabolic modelling	COBRA Toolbox [57]	Metabolic modelling and FBA
	SurreyFBA [58]	Constraint-based modelling of genome-scale networks
	CycSim [59], BioMet Toolbox [60]	Analysing genome-scale metabolic models; includes enzyme knockout simulations
	iPATH2 [61], GLAMM (genome-linked application for metabolic maps) [62]	Interactive visualization of data on metabolic pathways

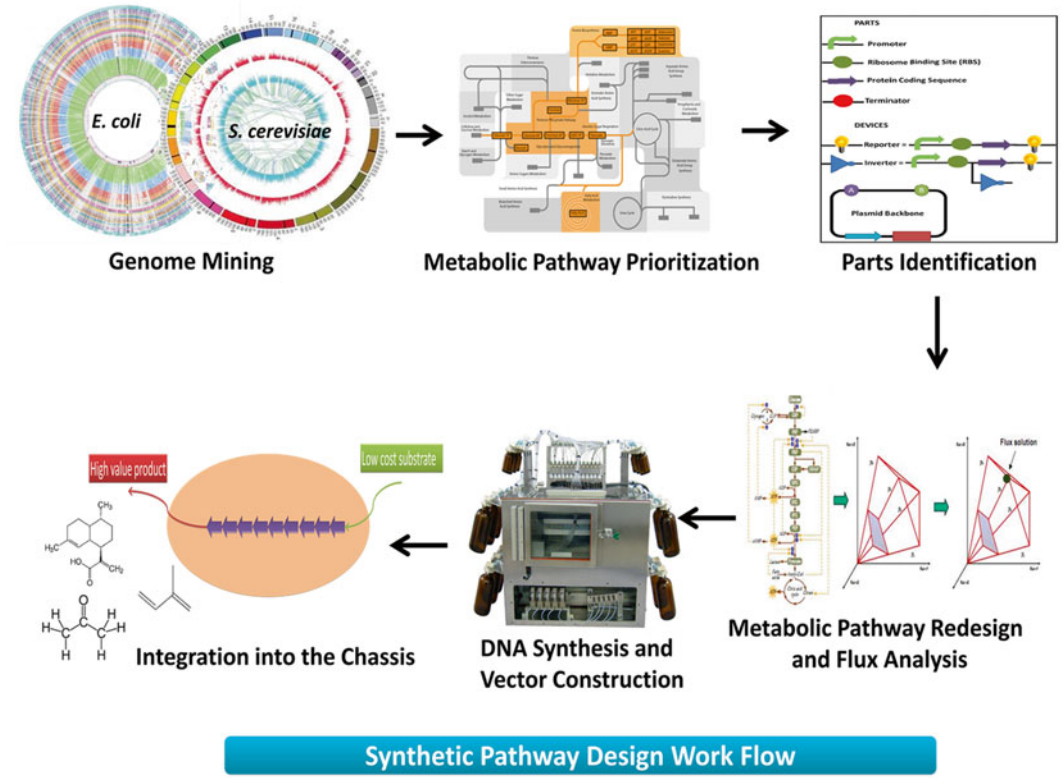

Fig. 1.1 Synthetic pathway design workflow

All the genetic manipulations involve the construction of a vector that contains all the enzymes required to reconstitute the novel metabolic pathway in the heterologous host. Therefore the cloning vector should be stable, have a consistent copy number, should replicate and express large sequences of DNA. The enzyme production rate from these vectors can be tuned to the desired levels by varying the promoter [65], ribosome binding strength [66] and stabilizing the half-life of the mRNA [67]. Of these, promoters are essential in controlling biosynthetic pathways that respond to a change in growth condition or to an important intermediary metabolite [68, 69]. These kinds of promoters allow inexpensive and inducer-free gene expression. Once a vector with all the desired properties is constructed the expression of the genes should be well coordinated, which can be done using a non-native RNA polymerase or transcription factor that can induce multiple promoters [70]; group related genes into operons; vary ribosome binding strength for the enzymes encoded in the operon [71]; or controlling mRNA stability of each coding region [72].

1.4 Important Breakthrough in Metabolic Engineering Using Synthetic Biology Approach

Though synthetic biology and construction of unnatural pathways is in its infancy, several pioneering experimental efforts in this direction have highlighted the immense potential of the field. In parallel, DNA sequencing has revealed a huge amount of information within the cellular level in terms of isozymes catalysing the same reaction in different organism. Alongside development of