Methods in
Molecular Biology 1073

Springer Protocols

Karen M. Polizzi
Cleo Kontoravdi *Editors*

Synthetic Biology

Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Synthetic Biology

Edited by

Karen M. Polizzi

Centre for Synthetic Biology and Innovation, Department of Life Sciences,
Imperial College London, London, UK

Cleo Kontoravdi

Centre for Process Systems Engineering, Department of Chemical Engineering,
Imperial College London, London, UK

Humana Press

Editors
Karen M. Polizzi
Centre for Synthetic Biology and Innovation
Department of Life Sciences
Imperial College London, London, UK

Cleo Kontoravdi
Centre for Process Systems Engineering
Department of Chemical Engineering
Imperial College London, London, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
ISBN 978-1-62703-624-5 ISBN 978-1-62703-625-2 (eBook)
DOI 10.1007/978-1-62703-625-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013945303

Printed on acid-free paper

Humana Press is a brand of Springer
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Synthetic Biology utilizes the design and construction principles of engineering to develop new biological components and systems or embed novel functions into existing ones, and standardize their behavior. This systematic approach to improving and increasing the programmability and robustness of biological components is expected to lead to the facile assembly of artificial biological components and integrated systems. While there has been considerable success in the field, it is still far from its full potential, with major challenges including standardization of parts so that they function reliably, with functional stability in face of mutations and other biophysical constraints such as noise, and integration of different parts.

The ambitious goals and interdisciplinary nature of this new research field have prompted the advancement of molecular biology techniques to meet the need for rapid development of biological building blocks as well as for their functional characterization and quality control. In parallel, researchers in the field of Systems Biology have recognized that the development of novel components necessitates advanced computational design tools that are capable of analyzing the behavior of parts and of constructing synthetic biological networks. This volume aims to review the latest developments in molecular biology techniques that find use in Synthetic Biology and to present some of the enabling computational tools that will aid in systematizing the design and construction of parts and systems. For a more comprehensive set of the latter, readers should look for our sister volume.

London, UK

Karen M. Polizzi
Cleo Kontoravdi

Contents

PART I INTRODUCTION

PART II DNA ASSEMBLY

PART III PART, PATHWAY, AND GENOME ENGINEERING

PART IV COMPUTATIONAL TOOLS FOR MODELLING BIOLOGICAL SYSTEMS

Contributors

Yuya Akeno • *Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Osaka University, Suita, Japan*

Travis S. Bayer • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK*

Imre Berger • *European Molecular Biology Laboratory (EMBL), BP 181, Polygone Scientifique, Grenoble, France; Unit of Virus Host Cell Interactions (UVHCI), Polygone Scientifique, Grenoble, France*

David Bikard • *Bacterial Genome Plasticity Unit, Institut Pasteur, Paris, France*

Anton Bryksin • *Department of Biochemistry, Center for Fundamental and Applied Molecular Evolution, Emory University School of Medicine, Atlanta, GA, USA*

James Chappell • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK*

Bing-Xue Dong • *State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou, China*

Tom Ellis • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK; Department of Bioengineering, Imperial College London, London, UK*

Carola Engler • *NOMAD BIOSCIENCE GMBH, Weinbergweg 22, Halle (Saale), Germany*

Paul Freemont • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK*

Matthias Haffke • *European Molecular Biology Laboratory (EMBL), BP 181, Polygone Scientifique, Grenoble, France; Unit of Virus Host Cell Interactions (UVHCI), Polygone Scientifique, Grenoble, France*

Mattheos A.G. Koffas • *Center for Biotechnology and Interdisciplinary studies, Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA*

Cleo Kontoravdi • *Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London, UK*

Sergei Kucherenko • *Department of Chemical Engineering and Chemical Technology, Imperial College London, London, UK*

Chang-Jie Li • *State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou, China*

Gang Li • *State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou, China*

Yu-Huan Liu • *State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou, China*

Sylvestre Marillonnet • *Department of Cell and Metabolic Biology, Leibniz-Institut für Pflanzenbiochemie, Weinberg 3, Halle, Germany*

Ichiro Matsumura • *Department of Biochemistry, Center for Fundamental and Applied Molecular Evolution, Emory University School of Medicine, Atlanta, GA, USA*

Didier Mazel • *Bacterial Genome Plasticity Unit, Institut Pasteur, Paris, France*

YAN NIE • *European Molecular Biology Laboratory (EMBL), BP 181, Polygone Scientifique, Grenoble, France; Unit of Virus Host Cell Interactions (UVHCI), Polygone Scientifique, Grenoble, France*

KAREN M. POLIZZI • *Centre for Synthetic Biology and Innovation, Department of Life Sciences, Imperial College, London, London, UK*

RUI T.L. RODRIGUES • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK*

HERBERT M. SAURO • *Department of Bioengineering, University of Washington, Seattle, WA, USA*

ZENGYI SHAO • *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

VANDANA SHARMA • *Department of Biomedical Engineering, University of California, Davis, CA, USA*

SEAN C. SLEIGHT • *Department of Bioengineering, University of Washington, Seattle, WA, USA*

CRISTINA VIOLA • *European Molecular Biology Laboratory (EMBL), BP 181, Polygone Scientifique, Grenoble, France; Unit of Virus Host Cell Interactions (UVHCI), Polygone Scientifique, Grenoble, France*

TIM WEENINK • *Centre for Synthetic Biology and Innovation, Imperial College London, London, UK; Department of Bioengineering, Imperial College London, London, UK*

PENG XU • *Center for Biotechnology and Interdisciplinary Studies, Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA*

BEI-WEN YING • *Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan*

YOHEI YOKOBAYASHI • *Department of Biomedical Engineering, University of California, Davis, CA, USA*

TETSUYA YOMO • *Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan; Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency, Suita, Japan*

LI-PING ZHANG • *State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou, China*

HUIMIN ZHAO • *Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Department of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

Part I

Introduction

Chapter 1

What Is Synthetic Biology?

Karen M. Polizzi

Abstract

Synthetic biology is a rapidly developing field that aims to engineer new biological systems that do not already exist in Nature or redesign existing systems from scratch. The emergence of synthetic biology has been supported by a number of enabling technologies and what has developed is a broad field that currently encompasses many activities. The aim of this chapter is to introduce the field and discuss some key examples to date. What these examples have in common is a set of underlying molecular biology techniques including DNA assembly and combinatorial diversity generation, as well as computational modelling to assist in designing the new biological systems.

Key words Synthetic biology, Metabolic engineering, Artificial life, Engineering design, Applications

1 Introduction: What Is Synthetic Biology?

Synthetic biology lies at the intersection of engineering, biological sciences, and computational modelling (Fig. 1), borrowing from each tools and concepts that help the synthetic biologist design new biological entities with user-defined properties. The evolution of the field has been enabled by a series of technological developments in each of the contributing disciplines. Faster and cheaper DNA sequencing technology has led to increased availability of DNA sequencing information and Moore's law style decreases in the cost of DNA synthesis have contributed a wealth of potential genes for synthetic biologists to work with and the ability to order these on demand instead of resorting to traditional molecular biology methods which are slow and tedious. Systems biology has supplied a greater understanding of the function and interconnectivity of natural biological systems as well as computational models that describe cellular and/or systems level processes. Computational power has increased to the extent that many of these models can run on ordinary desktop computers. It is the confluence of these events that has led to our ability to engineer biology.

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_1, © Springer Science+Business Media New York 2013

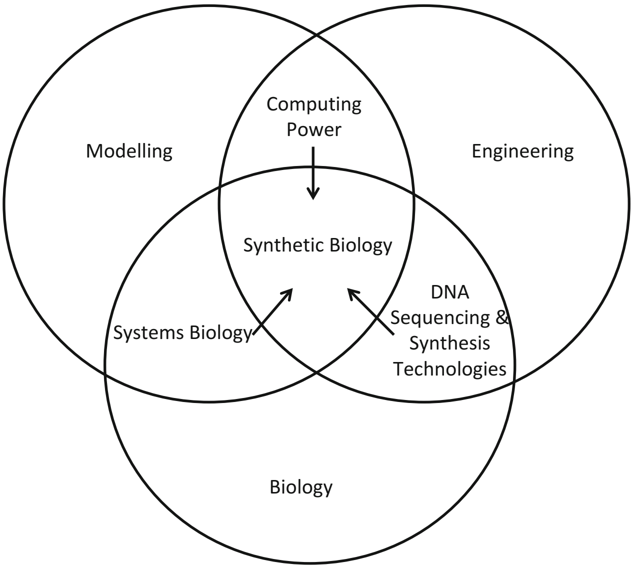


Fig. 1 Synthetic biology is a set of research activities at the intersection of engineering, computational modelling, and biological sciences. It builds on a variety of technologies and tools including improvements in DNA sequencing, cheaper gene synthesis technologies, increased computational power, and a better understanding of biological systems gained through systems biology

As with any new field, the exact definition of what constitutes "synthetic biology" is still murky and it will likely take years before a unified field emerges. In general, the field currently encompasses a broad set of ideas surrounding both meanings of the word synthetic. Some work adopts the definition of synthetic as artificial, unnatural, or not occurring in nature. For example, the work of Steve Benner's group to develop DNA that incorporates unnatural base pairs [1] uses such a definition, as does the goal to derive self-replicating systems based on chemical components [2]. Other work employs the use of the term synthetic as it is used in synthetic chemistry, the generation of compounds through a series of chemical reactions (in this case biochemical reactions). An example here would be computationally-assisted metabolic engineering to overproduce metabolites [3]. Clearly, these two definitions are related as the organisms which are constructed to do the synthetic chemistry are augmented beyond what would occur naturally. Table 1 summarizes some of the definitions of synthetic biology employed in the literature. Many research activities will encompass more than one of these definitions and the list is probably not exhaustive of all self-described synthetic biology activities.

As discussed above, some approaches to synthetic biology focus on creating artificial life. These include projects that create living organisms with nonnatural components such as noncanonical amino acids [4] or based on alternative genetic codes [5], as well as attempts

Table 1
Types of synthetic biology research

Type of research	Examples
Nonnatural systems	Expanded genetic code, orthogonal ribosomes, proteins containing noncanonical amino acids, biology of reversed chirality
Self-replicating chemical systems	Photochemical systems, self-replicating RNA systems, protocells
Minimal cells	Genome reduction, natural minimal cells, synthetic cells, vesicles harboring minimal genetic circuits
Advanced metabolic/genetic/protein engineering	Rational strategies for metabolic engineering, metabolic flux analysis, pathway design, computation modelling of whole cell metabolism, protein design
Engineering-based approaches	Forward engineering based on computational modelling, computer-aided design of pathways/organisms, parts/devices/systems, modular construction of pathways

to create self-replicating chemical systems or “protocells” [6]. In addition, work towards the creation of minimal organisms (i.e., cells that contain the bare minimum number of genes essential for life, thus representing blank slates for inputting new function) could also be classified as creating artificial life [7]. Other approaches attempt to use cells to perform synthetic chemistry. This includes approaches which could be thought of as “next generation” metabolic engineering utilizing metabolic flux analysis, whole genome scale modelling, and high throughput experimentation in the construction of new metabolic pathways [8, 9]. Straddling the two definitions of synthetic is what can loosely be described as the “engineering science”-based approaches. Engineering science approaches literally treat biology a substrate to be engineered and employ common engineering design techniques from other disciplines such as electrical or aerospace engineering including computer-aided design tools, predictive modelling, and prototyping of systems [10]. For researchers in this area, the focus is on tools and techniques and applications can come in under either definition of synthetic.

2 What Tools and Protocols Does a Synthetic Biologist Need?

Regardless of the type of research, the tools a synthetic biologist needs are broadly similar. They include efficient ways to clone single genes of interest, pathways/operons, and larger pieces of DNA into plasmids or genomes and ways to generate combinatorial libraries to diversify individual parts as well as examine the effects of spatial arrangement on the productivity of pathways and systems. Alongside this is the need to model biological systems for the purposes of design and optimization and to understand which parameters in the model are the most "sensitive" (i.e., exert the biggest control over the behavior of the system). These parameters are the ones most likely to need careful attention in the design phase. This collection of easy-to-follow, useful protocols has been selected to aid every synthetic biology in his/her goal of engineering new biological systems.

References

1. Yang Z, Hutter D et al (2006) Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. Nucleic Acids Res 34:6095–6101
2. Declue MS, Monnard P-A et al (2009) Nucleobase mediated, photocatalytic vesicle formation from an ester precursor. J Am Chem Soc 131:931–933
3. Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23:612–616
4. Noren CJ, Anthonycahill SJ et al (1989) A general-method for site-specific incorporation of unnatural amino-acids into proteins. Science 244:182–188
5. Neumann H, Wang K et al (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464:441–444
6. Dzieciol AJ, Mann S (2012) Designs for life: protocell models in the laboratory. Chem Soc Rev 41:79–85
7. Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2:45
8. Chung BKS, Selvarasu S et al (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast pichia pastoris for strain improvement. Microb Cell Fact 9:50
9. Lemuth K, Hardiman T et al (2008) Global transcription and metabolic flux analysis of Escherichia coli in glucose-limited fed-batch cultivations. Appl Environ Microbiol 74:7002–7015
10. Endy D (2005) Foundations for engineering biology. Nature 438:449–453

Part II

DNA Assembly

Chapter 2

Gene Synthesis Method Based on Overlap Extension PCR and DNAWorks Program

Gang Li, Bing-Xue Dong, Yu-Huan Liu, Chang-Jie Li, and Li-Ping Zhang

Abstract

Gene synthesis by chemical methods provides a powerful tool for modifying genes and exploring their structure, expression, and function in the post-genomic era. However, a bottleneck in recent gene synthesis technologies is the high cost of oligonucleotide synthesis and post-synthesis sequencing. Here, we describe a simple, rapid, and low-cost gene synthesis method based on overlap extension PCR (OE-PCR) and the DNAWorks program. This method enables DNA sequences with sizes ranging from 200 bp to 3 kb to be synthesized with few errors, and these errors can be easily corrected by site-directed mutagenesis. Thus, it is amenable to automation for the multiplexed synthesis of different genes and has a potential for high-throughput gene synthesis.

Key words Gene synthesis, DNAWorks, Overlap extension PCR, Site-directed mutagenesis

1 Introduction

In the post-genomic era, gene sequences from thousands of unknown proteins have to be obtained in order to explore the function of these proteins. In many cases, chemical synthesis of gene sequences may be the best choice because template DNAs are often not readily available or the natural DNA sequences may not be optimally expressed in heterologous systems, such as *Escherichia coli* (*E. coli*) and yeast. In recent years, the demand for the full synthesis of gene sequences has increased dramatically in order to study gene structure, expression, and function and gene synthesis technology has been widely used in many applications, including de novo synthesis of novel biopolymers [1], codon optimization [2–4], DNA vaccines construction [5], or simple gaining access to known DNA sequences when original templates are not available.

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_2, © Springer Science+Business Media New York 2013

To date, several strategies of gene synthesis have been reported including oligonucleotide ligation [6], the *Fok*I method [7], self-priming PCR [8–10], dual asymmetrical PCR (DA-PCR) [11], PCR-based assembly [12], and the template-directed ligation (TDL) [13]. Among these methods, PCR-based assembly approach (overlap extension PCR, OE-PCR) developed by Stemmer et al. is the most popular due to its simplicity and reproducibility. This involves generating overlapping oligonucleotides which, when assembled, form the template for the gene of interest. The oligonucleotides are then repetitively extended by PCR, to assemble the full-length gene in a single step [12]. However, the method does not work consistently for all genes [14, 15] and requires optimization on a case-by-case basis. To solve these problems, Young and Dong developed a two-step total gene synthesis method that combines DA-PCR and OE-PCR [16]. This method only needs one set of conditions for all oligonucleotides and eliminates the mutation problem occurring in almost all current gene synthesis approaches. However, the length of oligonucleotides used for this method is limited to less than 25 nt, thus the number of oligonucleotides required and the corresponding cost of oligonucleotide synthesis are greatly increased. Furthermore, the process of this method is tedious and time-consuming due to the addition of a DA-PCR step. In addition, the process of designing oligonucleotides for gene assembly is boring and confusing, especially when a large number of genes need to be synthesized. Therefore, an automated strategy for synthetic gene design is urgently demanded.

In this chapter, we describe a simple, reproducible method for low-cost gene synthesis based on an improved single-step OE-PCR protocol and DNAWorks software. This approach enables any DNA sequence to be synthesized quickly with few errors and low cost based on several improvements on Stemmer's method, including automated design of oligonucleotides using the DNA-Works software (instead of manual design), keeping the melting temperatures (Tm) of the overlapping regions similar to ensure primer specificity and normalized annealing conditions, using higher fidelity PrimeSTAR HS DNA polymerase (TaKaRa, China) to reduce error rate, designing special OE-PCR conditions, as well as adopting site-directed mutagenesis [17] to correct mutagenic sites accidentally produced in the gene synthesis process. Using this cost-effective method, several genes with lengths of 1,090 bp [18], 1,032 bp [18], 1,583 bp [18], 2,256 bp (Fig. 1), and 3,792 bp (Fig. 2, *see* **Note 1**) have been successfully synthesized.

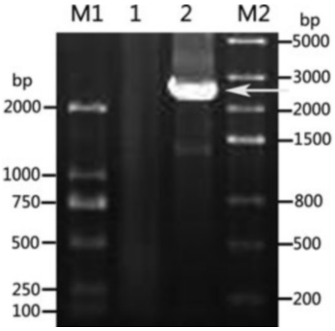


Fig. 1 Gel electrophoresis in 0.8 % agarose gel of a galactosidase gene (2,256 bp) of *Thermoanaerobacter mathranii* synthesized with our method. *Lane M1*, DL2000 DNA marker (TaKaRa); *Lane 1*, products from the first run of assembly PCR; *Lane 2*, products from the second run of assembly PCR; *Lane M2*, DL5000 DNA marker (TaKaRa)

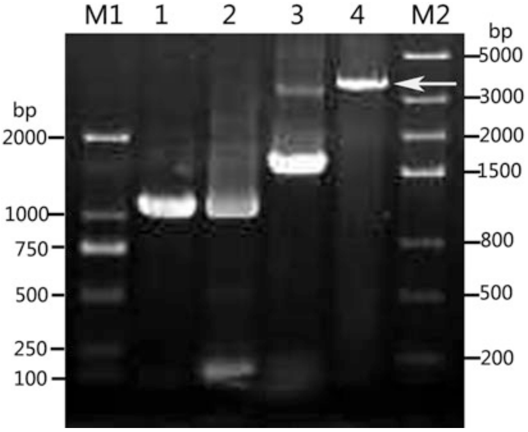


Fig. 2 Gel electrophoresis in 0.8 % agarose gel of a 3,792-bp fusion gene, assembled from three genes, including a manganese peroxidase gene of *Phanerochaete chrysosporium* (1,137 bp), a *cip1* peroxidase gene of *Coprinus cinereus* (1,092 bp), and a laccase gene of *Trametes versicolor* (1,563 bp). *Lane M1*, DL2000 DNA marker (TaKaRa); *Lane 1*, products of a manganese peroxidase gene of *Phanerochaete chrysosporium* from the second run of assembly PCR; *Lane 2*, products of a *cip1* peroxidase gene of *Coprinus cinereus* from the second run of assembly PCR; *Lane 3*, products of a laccase gene of *Trametes versicolor* from the second run of assembly PCR; *Lane 4*, the second run of PCR products of a 3,792-bp fusion gene, assembled from abovementioned three genes; *Lane M2*, DL5000 DNA marker (TaKaRa)

2 Materials

2.1 Reagents

1. Agarose and dNTPs.
2. Gel Extraction Kit (Omega, Guangzhou, China), TaKaRa MutanBEST Kit (TaKaRa, Dalian, China), QuikChange Lightning Site-Directed Mutagenesis Kit and Multi-site QuikChange Lightning Site-Directed Multi Mutagenesis Kit (Stratagene, La Jolla, CA, USA).

2.2 Biological Materials

1. *E. coli* DH5α competent cells.
2. Vector pUC19, DNA marker, PrimeSTAR HS DNA Polymerase (TaKaRa), PfuUltra High-Fidelity DNA polymerase (Stratagene).
3. Oligonucleotides, designed using DNAWorks program, were synthesized by Invitrogen Biotechnology Corp. Ltd. (Shanghai, China).
4. The synthetic genes in this chapter were a galactosidase gene of *Thermoanaerobacter mathranii* (2,256 bp, GenBank accession number AJ316558) and a fusion gene (3,792 bp in total), composed of a manganese peroxidase gene of *Phanerochaete chrysosporium* (1,137 bp, GenBank accession number M77513), a cip1 peroxidase gene of *Coprinus cinereus* (1,092 bp, GenBank accession number X69457), and a laccase gene of *Trametes versicolor* (1,563 bp, GenBank accession number AF414109).

2.3 Equipment

1. 9800 Fast Thermal Cycler (ABI, USA).
2. Agarose gel electrophoresis system (Biorad, USA).
3. Spectrophotometer UV 9600 (RayLeigh, Beijing, China).
4. Centrifuge MIKRO220R (Hettich, Germany).

3 Methods

The whole procedure of the gene synthesis method based on OE-PCR and DNAWorks program consists of six steps: oligonucleotide design, oligonucleotide synthesis, gene assembly, gene amplification, gene cloning and sequencing, and error correction of mutation sites (Fig. 3). Here we give a detailed description to the procedure step by step.

3.1 Oligonucleotide Design Using DNAWorks Program

1. Access the DNAWorks program (http://helixweb.nih.gov/dnaworks). The length of the gene for oligonucleotides design using DNAWorks program should be less than 3 kb. If the size of a gene is far more than 3 kb, it should be firstly separated into several 1–3 kb DNA fragments followed with 20–50 nt overlaps between

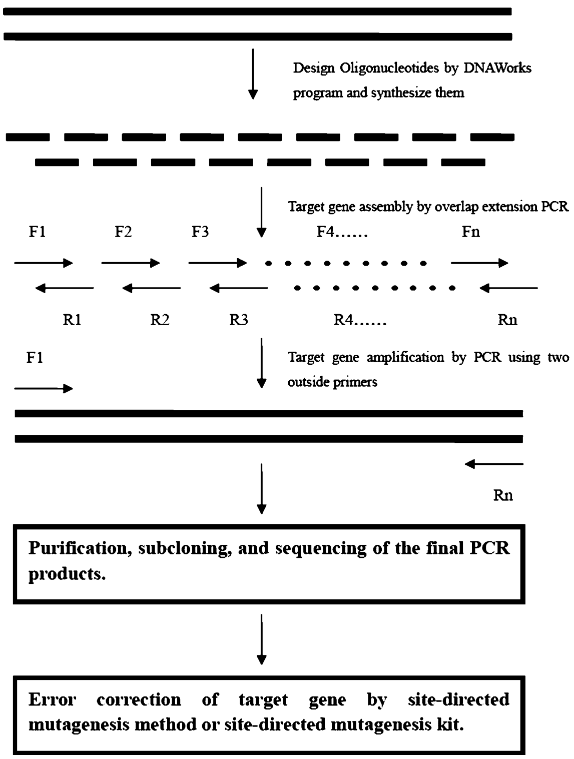

Design Oligonucleotides by DNAWorks program and synthesize them
Target gene assembly by overlap extension PCR
F1
F2
F3
F4......
Fn
R1
R2
R3
R4......
Rn
F1
Target gene amplification by PCR using two outside primers
Rn
Purification, subcloning, and sequencing of the final PCR products.
Error correction of target gene by site-directed mutagenesis method or site-directed mutagenesis kit.

Fig. 3 Schematic diagram of gene synthesis method based on overlap extension PCR and DNAWorks program

adjacent DNA fragments. Then, synthesize each DNA fragment respectively, and assemble them to a full gene (*see* **Note 1**).

2. Enter a job name and your email address, then choose the desired codon frequency table for expression optimization (if required).
3. Enter the desired annealing temperature, oligonucleotide length and codon frequency threshold, and oligonucleotide parameters (*see* **Note 2**).
4. Check the "TBIO" box and but do not choose "No gaps in assembly" (*see* **Note 3**).
5. Restriction sites or custom sites can be excluded from the protein coding region of the synthetic gene using "restriction site screen" and "custom site screen" function.
6. Finally, upload the sequence file or enter the sequence manually, and press "design oligos" button, the software will give a list of the sequences (5′–3′) of the oligonucleotides necessary to generate the synthetic gene. Furthermore, the final scores for the individual trials are also displayed in this section. In general, oligonucleotides with lower scores are better (*see* **Notes 4** and **5**).

7. Have the oligonucleotides synthesized at 50 nmol scale by your usual provider (*see* **Note 6**).

3.2 Gene Assembly and Verification

1. Dilute oligonucleotides to a concentration of 100 μM with double distilled water. Mix in equal volume and then dilute mixture to 2 μM for later use.
2. For assembly PCR, mix a 30 μL reaction mixture containing 2 μL mixed oligonucleotides (2 μM), 1× PCR buffer, 200 μM dNTPs, 0.5 U PrimeSTAR HS DNA polymerase.
3. Thermocycle for 30 cycles at 98 °C for 10 s, 60 °C for 1 min, and 72 °C for *n* min (extension time *n* should be determined according to the length of the target gene, every 1 kb of gene length corresponds to 1 min of extension time) (*see* **Notes 7–9**).
4. Use 1.5 μL of crude extension mixture for PCR amplification using 0.5 μM of the two outer primers in a final volume of 30 μL and the same PCR conditions from **step 2**.
5. Analyze the PCR product by agarose gel electrophoresis (Fig. 2) and purify the target DNA band by gel extraction.
6. Clone the target DNA into a pUC19 vector digested with *Sma*I. Screen the resulting white *E. coli* colonies after transformation for the presence of insert by colony PCR. Sequence several positive clones to check for fidelity (*see* **Note 10**).
7. If you cannot find a completely correct target gene sequence from 3 to 10 random selected clones, you should choose the target gene sequence with the least errors, and correct it using a site-directed mutagenesis kit (such as TaKaRa MutanB-EST Kit, Multi-site QuikChange Lightning Site-Directed Multi Mutagenesis Kit) following the manufacturer's protocol (*see* **Note 11**).

4 Notes

1. Our protocol is suitable to synthesize a DNA sequence with the length ranging from 200 bp to 3 kb. Because the lengths of most genes are lower than 3 kb, this method can be directly used to synthesize almost all of genes. If this protocol is applied to synthesize a DNA sequence far more than 3 kb, such as a long gene, a gene cluster, a phage genome, or a fusion DNA fragment, the target DNA sequence should firstly be separated into several 1–3 kb DNA fragments followed with 20–50 nt overlaps between adjacent DNA fragments. Then, each DNA fragment is synthesized using this protocol. Finally, several DNA fragments with completely correct DNA sequence can be easily assembled into a full-length gene, gene cluster, phage genome, or fusion DNA fragment by OE-PCR. In Fig. 2, we

assembled three genes with length of 1–2 kb into a 3,792-bp fusion gene using this modification.

2. Oligonucleotide parameters include annealing temperature (generally limited to be within 58–70 °C), oligo length, and codon frequency threshold. Aim to keep the length of each oligonucleotide below 50 nt which is the maximum synthetic length of oligonucleotide guaranteeing complete fidelity for most commercial DNA synthesis service providers, forty nucleotides is probably the most optimal based on cost performance. The oligonucleotide concentration and ion concentrations are used to calculate appropriate annealing temperatures. When designing oligonucleotides for our method begin with default values for the codon frequency threshold ("10 %"), oligonucleotide concentration ("1E-7"), Na^+/K^+ concentration ("0.05 M"), Mg^{2+} concentration ("0.002 M"), number of solutions ("1").
3. Checking the TBIO box is helpful in designing oligonucleotides that perform well in the later gene assembly steps, but restricting the oligonucleotides to no gaps will slow down the optimization and result in higher scores due to a higher probability of misprimes.
4. Make sure that none of the overlaps between oligonucleotides drop below 12 nt because a short overlap may allow mispriming. 15–25 nt overlaps at both 5′ and 3′ ends between adjacent oligonucleotides are recommended.
5. AT base pairs should be distributed equally in oligonucleotides designed by utilizing degeneracy of codon usage. It is also helpful to reduce base deletions during gene assembly.
6. We strongly recommend having oligonucleotides purified by HPLC (high performance liquid chromatography) or PAGE (polyacrylate gel electrophoresis), in order to reduce errors during the oligonucleotides assembly.
7. A "touchdown" PCR or "hot start" PCR is helpful to minimize mispriming during gene assembly and amplification. It is also beneficial to produce a single band in target gene amplification by PCR using two outside primers.
8. Optimizing the PCR conditions is always useful to minimize errors introduced during PCR. Annealing temperature is the most important parameter of PCR conditions for gene assembly, and a high annealing temperature (60–68 °C) is generally crucial for decreasing errors during gene assembly.
9. A high-fidelity DNA polymerase is very important to minimize errors introduced during gene assembly and amplification. PrimeSTAR HS DNA polymerase (TaKaRa) is recommended as DNA polymerase to assemble and amplify

gene due to its high cost-efficiency, but any other high-fidelity DNA polymerase, such as PfuUltra High-Fidelity DNA polymerase (Stratagene), will also work.

10. Any suitable cloning procedure can be substituted in this step.
11. Using our protocol, the frequency of errors in synthetic genes is generally 0–2 single base deletions or point mutations/kb, and it is often easy to find a completely correct target gene sequence from 3 to 10 random selected clones. If you cannot find it, a target gene sequence with the least errors should be selected and corrected using site-directed mutagenesis method [17] or - site-directed mutagenesis kit (TaKaRa MutanBEST Kit, Stratagene QuikChange Lightning Site-Directed Mutagenesis Kit, Stratagene QuikChange Lightning Multi Site-Directed Mutagenesis Kit is available).

Acknowledgements

This work was supported by the grant of the National Natural Science Foundation of China (No. 31170117, 31270156), the National Marine Research Special Funds for Public Welfare Projects of China (201205020), the Major Science and Technology Projects of Guangdong Province, China (2011A080403006), the Fundamental Research Fund for the Central Universities of Sun Yat-sen University (No. 11lgpy23), the key grant of Science and Technology Department of Henan Province (112102210385), and Science Foundation for the high-level talent of Nanyang Normal University (zx20110007).

References

1. Van Hest JCM, Tirrell DA (2001) Protein-based material, toward a new level of structural control. Chem Commun 19:1897–1904
2. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346–353
3. Xiong AS, Yao QH et al (2005) High level expression of a recombinant acid phytase gene in *Pichia pastoris.* J Appl Microbiol 98:418–428
4. Xiong AS, Yao QH et al (2006) High level expression of a synthetic gene encoding *Peniophora lycii* phytase in methylotrophic yeast *Pichia pastoris.* Appl Microbiol Biotechnol 72:1039–1047
5. Yang ZY, Kong WP et al (2004) A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature 428:561–564
6. Scarpulla RC, Narang S, Wu R (1982) Use of a new retrieving adaptor in the cloning of a synthetic human insulin A-chain gene. Anal Biochem 121:356–365
7. Mandecki W, Bolling TJ (1988) *Fok*I method of gene synthesis. Gene 68:101–107
8. Dillon PJ, Rosen CA (1990) A rapid method for the construction of synthetic genes using the polymerase chain reaction. Biotechniques 9:298–300
9. Chen GQ, Choi I, Ramachandran B, Gouaux JE (1994) Total gene synthesis: novel single step and convergent strategies applied to the construction of a 779 base pair bacteriorhodopsin gene. J Am Chem Soc 116:8799–8800
10. Hayashi N, Welschof M et al (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. Biotechniques 17:310–315

11. Sandhu GS, Aleff RA, Kline BC (1992) Dual asymmetric PCR: one-step construction of synthetic genes. Biotechniques 12:14–16
12. Stemmer WP, Crameri A et al (1995) Single step assembly of a gene and entire plasmid from large number of oligodeoxyribonucleotides. Gene 164:49–53
13. Strizhov N, Keller M et al (1996) A synthetic *cryIC* gene, encoding a *Bacillus thuringiensis* delta-endotoxin, confers *Spodoptera* resistance in alfalfa and tobacco. Proc Natl Acad Sci U S A 93:15012–15017
14. Lin Y, Cheng X, Clark TG (2002) The use of synthetic genes for the expression of ciliate proteins in heterologous systems. Gene 288:85–94
15. Gao X, Yo P et al (2003) Thermodynamically balanced inside-out (TBIO) PCR-based synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. Nucleic Acids Res 31:e143
16. Young L, Dong QH (2004) Two-step total gene synthesis method. Nucleic Acids Res 32:e59
17. An Y, Ji J et al (2005) A rapid and efficient method for multiple-site mutagenesis with a modified overlap extension PCR. Appl Microbiol Biotechnol 68:774–778
18. Dong BX, Mao RQ et al (2007) An improved method of gene synthesis based on DNAWorks software and overlap extension PCR. Mol Biotechnol 37:195–200

Chapter 3

BioBrick™ Assembly Using the In-Fusion PCR Cloning Kit

Sean C. Sleight and Herbert M. Sauro

Abstract

Synthetic biologists assemble genetic circuits from standardized biological parts called BioBricks™. BioBrick™ examples include promoters, ribosome binding sites, DNA or RNA-coding sequences, and transcriptional terminators. Standard BioBrick™ assembly normally involves assembly of two BioBricks™ at a time using restriction enzymes and DNA ligase. Here we describe an alternative BioBrick™ assembly protocol that describes the assembly of two BioBricks™ using the In-Fusion PCR Cloning Kit. This protocol can also be adapted to use similar recombination-based assembly methods, such as SLIC and Gibson assembly.

Key words BioBrick, DNA assembly, In-Fusion, Synthetic biology, Plasmids, Genetic circuits

1 Introduction

DNA assembly is a cornerstone of synthetic biology research and is used by many experimental synthetic biologists to assemble new genetic circuits. There are ongoing improvements for DNA assembly methods since DNA synthesis costs are still prohibitive for many large-scale circuit construction efforts. Traditional DNA assembly methods using restriction enzymes and DNA ligase [1–6] have their own distinct advantages and disadvantages compared to recombination-based methods such as Gibson assembly [7], In-Fusion [8–10], SLIC [11], and USER [12]. A detailed discussion of various DNA assembly methods used in synthetic biology has been offered by Ellis et al. [13]. This chapter focuses on DNA assembly using the In-Fusion method and specifically illustrates the protocol for adapting the In-Fusion PCR Cloning Kit for BioBrick™ standard assembly [8]. For examples of assembly and reengineering of BioBricks, including assembly of two parts, part swapping, insertions, deletions, and three-way assemblies, refer to [8].

BioBricks™ are standardized biological parts such as promoters, ribosome binding sites (RBSs), DNA- or RNA-coding sequences, and transcriptional terminators. The Registry of Standard Biological

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_3, © Springer Science+Business Media New York 2013

Parts (http://www.partsregistry.org) stores and distributes BioBricks™ as physical DNA sequences on plasmids in 384-well plates. Standard BioBrick™ assembly involves digestion of two plasmids with different restriction enzymes, whereby the restricted DNA fragments can be ligated together due to their compatible sticky ends. Since there are standardized sequences on BioBricks™, this allows for the assembly of two BioBricks™ in an idempotent fashion [1]. After assembly of two BioBricks™, a "scar" sequence is left between the parts that cannot be recognized by the same restriction enzymes used for digestion. This standardized method allows for this newly assembled BioBrick™ to be again assembled with other BioBricks™. One of the advantages of BioBrick™ standard assembly is that the same components can be used for every assembly reaction, but disadvantages include the laborious extraction of DNA from gels and lower assembly efficiency (at least in our experience).

Standard BioBrick assembly can also be performed using the In-Fusion PCR Cloning Kit from Clontech in the protocol described here. The In-Fusion method allows for the assembly of two or more PCR products into a circular plasmid that can be directly transformed into competent cells [8–10]. Since the proprietary In-Fusion enzyme has 3′–5′ exonuclease activity, a PCR-amplified insert will fuse to a PCR-amplified linearized vector when the fragments have at least 15 bp of homology on both ends. This is because the ssDNA that is exposed by the exonuclease on one fragment can bind to complementary ssDNA on another fragment. The gaps in the duplex DNA are then repaired after transformation into *E. coli* competent cells. This method can be used to assemble two or more BioBricks™, as described in the protocol below, or reengineer existing BioBricks™ [8]. In-Fusion allows for the flexibility of PCR-amplifying any sequence from any part, then assembling these DNA fragments in one step. This last point is not unique to In-Fusion, however, as other recombinase-based assembly methods such as Gibson [7] and SLIC [11] are also flexible in this regard.

Figure 1 shows the general scheme for In-Fusion assembly. In this example, Part A is PCR-amplified with the vector and Part B is PCR-amplified as the insert. Since the AF/BR and AR/BF primers have homology, the PCR products using these primers will fuse in the In-Fusion reaction into a circular plasmid. For the protocol below, the R0011 *lacI*-regulated promoter will be used as the example for Part A and the E0240 GFP generator will be used as the example for Part B (Figs. 2 and 3). When these two parts are assembled, they will form a functional genetic circuit, where GFP is expressed from the R0011 promoter. Since R0011 is much smaller (55 bp) than E0240 (876 bp), it is recommended to PCR-amplify the smaller part with the vector so that the molar ratios of the PCR products are as similar as possible. Also, it is possible to

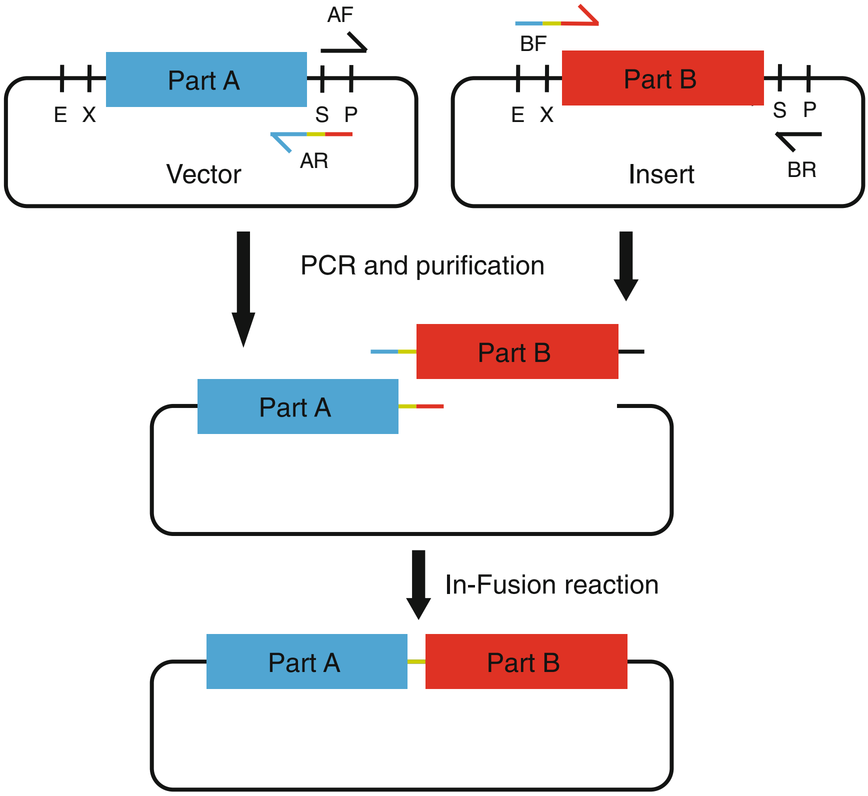


Fig. 1 In-Fusion assembly of two BioBricks. Three major steps are required after design of the assembly: PCR, purification, and an In-Fusion reaction. Parts A and B are PCR-amplified (in this example the vector is amplified with Part A) and purified without gel extraction. Each assembly requires four primers, where two are specific to the junction between parts (AR and BF) and two are general vector primers (AF and BR). BioBrick Part A (*blue*) and Part B (*red*) are on pSB1A2 vectors encoding ampicillin resistance. The pSB1A2 vector has a prefix sequence upstream of the part (indicated by the "E" for EcoR1 and "X" for XbaI restriction sites) and suffix sequence downstream of the part (indicated by the "S" for SpeI and "P" for PstI restriction sites). Primers are illustrated with a horizontal line and another shorter line (3′ end) intersecting at a 45° angle. Primer sequences are described in Subheading 2 and are color-coded to show their homology to other DNA fragments. The *thick black line* indicates BioBrick suffix homology on the pSB1A2 vector. The *yellow line* is the scar that is normally between parts after standard BioBrick assembly, if this is desired. The *blue line* is specific to Part A and the *red line* is specific to Part B. The purified PCR products are fused together in the In-Fusion reaction to create a circular plasmid. Restriction sites flanking the parts maintain the standard BioBrick format. This figure was adapted from [8]

amplify Part B with the vector instead of Part A, but this is not ideal for this particular assembly example. *See* ref. 8 for more details.

Figure 2 describes this assembly scheme at the individual part and genetic circuit levels. To amplify R0011 with the vector, we will use the AF primer as the forward primer (specific to the pSB1A2 vector) and create the AR reverse primer (*see* Subheading 3.1 below). To amplify E0240, we will create the BF forward primer (*see* Subheading 3.1 below) and use the BR reverse primer (specific

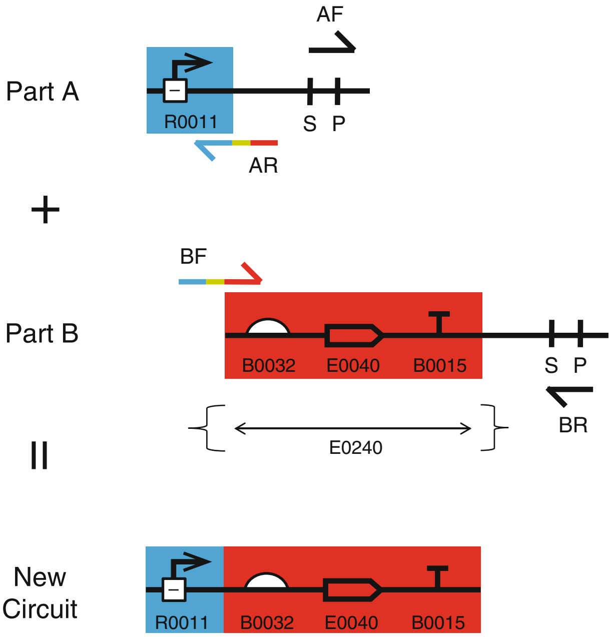


Fig. 2 In-Fusion assembly of two BioBricks at the genetic circuit level. Part A consists of the *lacI*-regulated promoter R0011. Part B is the E0240 GFP generator and consists of the B0032 ribosome binding site (RBS), E0040 GFP-coding sequence, and B0015 terminator. After the assembly reaction, a functional genetic circuit is constructed and will express GFP after transformation into competent cells. If the competent cells express LacI (as in a *lacI*^{q} strain), then IPTG is required for GFP expression. This figure was adapted from [8]

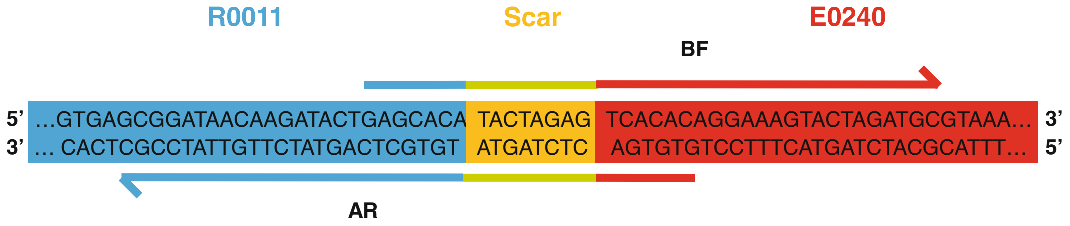

R0011
Scar
E0240
BF
5' ...GTGAGCGGATAACAAGATACTGAGCACA TACTAGAG TCACACAGGAAAGTACTAGATGCGTAAA... 3'
3' ... CACTCGCCTATTGTTCTATGACTCGTGT ATGATCTC AGTGTGTCCTTTCATGATCTACGCATTT... 5'
AR

Fig. 3 Primer design of the BF and AR primers. The BF primer is a forward primer specific to E0240 at the 3′ end and has a nonspecific region at the 5′ end that consists of part of R0011 and an 8-bp scar sequence. The AR primer is a reverse primer specific to R0011 at the 3′ end and has a nonspecific region at the 5′ end that consists of part of E0240 and an 8-bp scar sequence. These "junction" primer sequences overlap by 20 bp (as do the vector primer sequences by 26 bp), allowing for the PCR products to fuse in the In-Fusion reaction

to the pSB1A2 vector). Figure 3 illustrates the AR and BF primers at the DNA sequence level. We previously found that using higher lengths of homology (up to 48 bp) works better than shorter lengths on average [8]. However, for this example, we will use 20 bp of homology for illustration purposes.

2 Materials

1. Clontech In-Fusion PCR Cloning Kit (the dry down kit is recommended).
2. SOC medium (included in the In-Fusion PCR Cloning Kit).
3. Chemically competent cells (included in the In-Fusion PCR Cloning Kit).
4. Standard Desalted Primers (IDT primers are recommended).
5. Thermocycler.
6. 0.2 mL PCR tubes.
7. Phusion PCR Mastermix (or another high fidelity polymerase mastermix).
8. BioBrick™ template DNA for insert(s) and vector (mini-prepped plasmid diluted in water to 0.5 nmol).
9. Gel box with power supply.
10. Gel supplies (agarose gel with Sybr Safe or ethidium bromide, running buffer, loading dye, parafilm).
11. New England Biolabs 1 kb ladder (or an equivalent 1 kb ladder).
12. Transilluminator (Clare Chemical Dark Reader or UV transilluminator).
13. Gel imaging system.
14. Qiagen PCR Purification Kit.
15. Microvolume spectrophotometer (e.g., Nanodrop—not required, but very useful).
16. TE Buffer (pH 8.0).
17. Heat block or water bath (adjustable to 42 °C).
18. LB+Amp (100 μg/mL) agar plate.
19. LB+Amp (100 μg/mL) liquid medium.
20. Qiagen Miniprep Kit.
21. AF primer: 5′-TACTAGTAGCGGCCGCTGCAGGCTTC-3′ (specific to the pSB1A2 suffix+5 bases downstream of the suffix).
22. BR primer: 5′-GAAGCCTGCAGCGGCCGCTACTAGTA-3′ (reverse complement of AF primer).
23. VF2 primer: 5′-TGCCACCTGACGTCTAAGAA-3′ (~100 bp upstream of the part).
24. VR primer: 5′-ATTACCGCCTTTGAGTGAGC-3′ (~100 bp downstream of the part).

3 Methods

Perform all procedures at room temperature unless otherwise specified.

3.1 Genetic Circuit and Primer Design

1. Design a genetic circuit of interest. The example used in this protocol will be to assemble the R0011 promoter with the E0240 GFP generator to make a functional circuit (Figs. 1 and 2).
2. Obtain the template DNA that will be PCR-amplified (*see* **Note 1**). Dilute the Part A (R0011) and Part B (E0240) plasmid DNA to 0.5 nmol in molecular grade water or TE Buffer and store at −20 °C.
3. Create sequence files of Parts A and B to assemble using a DNA sequence visualization program (*see* **Note 2**). Save the files as "R0011" and "E0240."
4. Join the Part A and B sequence files to make the desired sequence to assemble. At the end of the R0011 DNA sequence, add the "scar" sequence ("TACTAGAG") if desired (*see* **Note 3**). Then, copy the E0240 DNA sequence and paste it after the scar sequence. Save the file as "R0011 + E0240."
5. Design and order the BF primer (Fig. 3). In the R0011 + E0240 sequence file, starting from the 5′ end (left-hand side) of the E0240 sequence, highlight nucleotides downstream (to the right) until you reach a T_m of ~58–65 °C (*see* **Note 4**). Using 24 nucleotides gives a T_m of ~59 °C. This is the sequence in the primer that is specific to the E0240 template DNA. Important: make sure there are no repeated sequences (e.g., scars, transcriptional terminators) at the very 3′ end of the primer, or you will get nonspecific amplification during the PCR reaction. In Fig. 3, this highlighted sequence corresponds to the red line in the BF primer. Next, highlight this 24 nucleotide sequence along with 20 nucleotides upstream (to the left) of E0240. This 20 nucleotide sequence upstream is the nonspecific "junction" sequence with homology to R0011/scar sequence and corresponds to the blue and yellow lines of the BF primer in Fig. 3. The total highlighted sequence should be a total of 44 nucleotides (20 nucleotides of homology to R0011/scar and 24 nucleotides that are specific to E0240). Copy this 44 nucleotide sequence into your favorite primer company's order form (*see* **Note 5**).
6. Design the AR primer (Fig. 3). In the R0011 + E0240 sequence file, starting from the 3′ end (right-hand side) of the R0011 sequence, highlight nucleotides upstream until you reach a T_m of ~58–65 °C. Using 24 nucleotides gives a T_m of ~58 °C. This is the sequence in the primer that will be specific to the R0011

template DNA. Important: make sure there are no repeated sequences (e.g., scars, transcriptional terminators) at the very 3′ end of the primer, or you will get nonspecific amplification during the PCR reaction. In Fig. 3, this highlighted sequence corresponds to the blue line in the AR primer. Next, highlight this 24 nucleotide sequence along with 20 nucleotides downstream of R0011. This 20 nucleotide sequence downstream is the nonspecific "junction" sequence with homology to the scar/E0240 sequence and corresponds to the yellow and red lines of the AR primer in Fig. 3. The total highlighted sequence should be a total of 44 nucleotides (24 nucleotides that are specific to R0011 and 20 nucleotides of homology to the scar/E0240). Copy this 44 nucleotide sequence into a new sequence file and *make the reverse complementary sequence*. Then copy this sequence into your favorite primer company's order form and make the purchase.

3.2 PCR, Purification, and Quantitation

1. After primers arrive and have been resuspended (*see* **Note 6**), perform a PCR reaction using Phusion PCR Mastermix. Follow the manufacturer's instructions for the exact protocol. PCR-amplify the R0011/vector with the AF and AR primers and E0240 with the BF and BR primers. Use a positive control with template DNA and primers you already know will give you a PCR product and a negative control that has no template DNA.
2. Run your PCR products out on an agarose gel (*see* **Note 7**). If both PCR products look to be the correct size and there are no secondary bands, then proceed to the next step (otherwise *see* **Note 8**).
3. Purify your PCR products using the Qiagen PCR Purification Kit. Follow the manufacturer's instructions. DpnI digestion is often recommended at this step, but is not required (*see* **Note 9**).
4. Quantify the purified PCR products in ng/μL (*see* **Note 10**).

3.3 Assembly Reaction and Verification

1. Calculate the amount of insert and vector DNA for the assembly reaction. Enter the ng/μL of the E0240 insert ("I") and R0011/vector ("V") DNA in an Excel (or other spreadsheet program), as shown in Table 1. For this example, we will use 50 ng/μL for the insert DNA and 25 ng/μL for the vector DNA. Next, calculate the size of the insert and vector DNA. For E0240, the size is 876 bp+20 bp homology to R0011/scar +26 bp (BR primer)=922 bp. For R0011/vector, the size is 55 bp (R0011)+2,079 (vector)+20 bp homology to scar/E0240=2,154 bp. In reality, this size calculation does not need to be this precise, but within 100 bp is ideal. Calculate the total ng of insert DNA to use using the formula in Table 1. This formula is two times the insert:vector DNA ratio,

Table 1
Assembly reaction calculation table

I (ng/μL)	*V* (ng/μL)	*I* (size)	*V* (size)	*I* (ng)	*V* (ng)	*I* (μL)	*V* (μL)	H_2O (μL)
50	25	922	2,154	2 × (922/2,154) × 100 = 85.6	100	85.6/50 = 1.7	100/25 = 4.0	10 − 1.7 − 4.0 = 4.3

This table illustrates an example of the calculations required for performing an In-Fusion reaction. The insert *I* and vector *V* concentrations (ng/μL), sizes (bp), total amounts (ng), and volumes (μL) for the In-Fusion reaction are shown below each heading

times 100 ng (the ideal vector DNA to use in the reaction). To calculate the insert μL to use in the reaction, divide the insert ng by the insert ng/μL. To calculate the vector μL to use in the reaction, divide the vector ng by the vector ng/μL. Then subtract the insert and vector μL by 10 (the total volume for the assembly reaction) to get the amount of molecular grade water to use.

2. Perform the assembly reaction. Mix the insert, vector, and water μL calculated from the last step into a 0.2 mL PCR tube. Then, mix this 10 μL volume with the lyophilized enzyme mixture in one of the dry-down tubes included in the In-Fusion kit by pipetting up and down (slowly to avoid bubbles) until the mixture is homogeneous. Transfer this entire volume back into the same 0.2 mL PCR tube. Put this tube in the thermocycler set at 37 °C for 15 min, 50 °C for 15 min, and a hold at 4 °C.
3. Thaw one tube of competent cells (provided with the In-Fusion kit) on ice about 15 min before the reaction will finish (*see* **Note 11**).
4. After the assembly reaction, add 30 μL of TE Buffer to the reaction tube and mix by pipetting up and down a few times.
5. Check that the competent cells are fully thawed by gently flicking on the tube.
6. Add 2.5 μL of the assembly reaction with TE Buffer to the tube of thawed competent cells and gently flick the tube to mix. Put the tube back on ice to incubate for 30 min.
7. While the cells/DNA are incubating, set your water bath or heat block to 42 °C and put one LB+Ampicillin (100 μg/mL) agar plate in the 37 °C incubator to pre-warm.
8. After the 30 min incubation, transform the DNA by incubating the competent cells tube at 42 °C for 45 s, then put the tube back on ice for 1 min.
9. Add 200 μL SOC medium to the tube.

10. Incubate the transformant cells tube for 1 h at 37 °C shaking at 250 RPM (*see* **Note 12**).
11. Spread 200 μL of the transformed cells on the pre-warmed LB+Ampicillin (100 μg/mL) agar plate and incubate plate upside down at 37 °C overnight (16–20 h is ideal), or until colonies are visible.
12. To test whether the assembly was successful or not, perform a functional screen or colony PCR reaction (*see* **Note 13**).
13. Grow three successful transformant colonies in 5 mL of LB+ Ampicillin (100 μg/mL) liquid medium overnight (16–20 h is ideal).
14. Miniprep at least one of the cultures to extract the plasmid using the Qiagen Miniprep Kit. Follow the manufacturer's instructions.
15. Submit the plasmid for DNA sequencing using the VF2 and VR primers (~100 bp on either side of the insert that are specific to the pSB1A2 vector). Using both primers will hopefully sequence the entire circuit.
16. Analyze the DNA sequencing results. This can be performed by aligning the reference R0011+E0240 file previously made with the sequence data and checking for the correct sequence. The trace (.ab1) file should always be checked to determine the quality of the sequence data.

4 Notes

1. Plasmid (template) DNA for PCR can be obtained in one of two ways. The first method is to transform R0011 and E0240 BioBrick™ DNA from the Registry plate into competent cells, grow overnight (16–20 h) at 37 °C shaking at 250 RPM, and extract the plasmid DNA using a miniprep kit. The second method is to dilute the BioBrick™ DNA from the Registry plate into water (either 1:10 or 1:100 dilution of the 10 μL DNA in TE Buffer usually works well). The second method is faster, but PCR results are not as consistent as the first method.
2. A free and easy-to-use DNA sequence visualization program is A Plasmid Editor (APE) that you can download here: http://biologylabs.utah.edu/jorgensen/wayned/ape/. Copy the DNA sequence from the Registry and paste this sequence into APE, then save the file. It will help to color code both parts with different colors by using the Feature Library function. This program also allows you to see the melting temperature (T_m) as you highlight DNA which will be useful in **steps 5** and **6** in Subheading 3.1.

3. This 8-bp scar sequence is an artifact of Standard assembly, but also provides spacing in between the promoter and RBS sequence to maximize expression levels. Unpublished results from our lab suggest that removal or lengthening of this scar to 20-bp severely decreases expression. All BioBrick™ scars are 8-bp except that there is a 6-bp scar in between the RBS and coding sequence that provides optimal spacing.

4. The kit also recommends the following for the template DNA-specific portion of the primer at the 3′ end: use a T_m of ~58–65 °C and length of 18–25 nucleotides, the T_m difference of the forward and reverse primers should be ≤4 °C, the GC content should be 40–60 %, there should not be identical runs of the same nucleotide at the 3′ end (e.g., AAAAA), and the last five nucleotides at the 3′ end of each primer should contain no more than two guanines (G) or cytosines (C). The kit also recommends to avoid complementarity within each primer to prevent hairpin structures and between primer pairs to avoid primer dimers. You can determine if the 3′ portion of each primer is unique and specific by performing a BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/). Clontech provides an online primer design tool (at http://bioinfo.clontech.com/infusion/) that may be useful. In most cases, as long as the 3′ end of the primer has sufficient homology to the template DNA and the 5′ end has at least 15 bp of homology to the PCR product to which it will be joined, the PCR product will amplify and assembly will be successful, but these recommendations will minimize failed assembly attempts.

5. I recommend IDT for primer and DNA synthesis orders.

6. I recommend using 100 pmol/μL in molecular grade water.

7. Using a 1 % agarose gel works fine, unless you are trying to resolve small fragments (<100 bp), in which case I recommend using a 3 % gel.

8. If you obtained the correct size bands, but have secondary bands, I recommend repeating the PCR reaction using a temperature gradient on the primer annealing step. A higher temperature should give you only the correct size band without secondary bands, but if the temperature is too high, then you may not get any PCR products. Alternatively, you can gel purify the correct size band using the QiaQuick Gel Extraction Kit, but this is not recommended due to the low yields obtained. The assembly reaction may work even if you have secondary bands, but the success rate will be lower.

9. It is often recommended to digest the template DNA with DpnI after the PCR reaction so that these plasmids do not later get transformed. In my experience, In-Fusion assembly has a sufficiently high success rate that this step is not required, but may be helpful for certain difficult assembly reactions. Also, using phosphatase on the vector DNA is not necessary and may even decrease efficiency.
10. I recommend using a Nanodrop to quantify DNA since it is very accurate. If a Nanodrop (or equivalent DNA quantitation machine) is not available, compare the PCR product band strength to the band strength in a 1 kb or 100 bp ladder, and quantify the DNA concentration in ng/μL.
11. You do not need to use the entire tube of competent cells unless you are trying to maximize the number of successful transformants. Normally 30 μL is a sufficient volume for assembly and make sure your assembly mixture volume:competent cell volume ratio is below 1:10.
12. Since this particular example uses a vector that encodes ampicillin resistance, this step is optional. If the antibiotic were kanamycin, or others that target the ribosome, this step is mandatory.
13. For this particular example, colonies that have the correctly assembled plasmid will fluoresce and therefore only a functional screen is necessary to determine successful transformants. The colonies can be visualized for fluorescence using a transilluminator. This is the ideal way to screen for successful transformants, but if this is not possible, a colony PCR reaction can also be performed to test for a successful assembly. In some cases, a colony PCR reaction is the only way to test for successful transformants. For this particular example, doing a colony PCR reaction using the VF2/VR primers (~100 bp on either side of the insert that are specific to the pSB1A2 vector) will yield a PCR product the size of 55 bp (R0011) + 8 bp (scar) + 876 bp (E0240) + ~200 bp (vector) = ~1,139 bp in the case of a successful transformant. If the transformant has a "background" plasmid (R0011 or E0240), then the PCR product size will be different. R0011 will have a very short size product, but E0240 will be nearly identical to the correctly assembled plasmid. Therefore, it will be difficult to distinguish a PCR product amplified from E0240 from the correctly assembled plasmid of R0011 + E0240. Of course, a clever colony PCR reaction could be devised, such as the use of a forward primer in the middle of R0011 and a reverse primer in the middle of E0240. The PCR reaction will work only if both parts are present.

Acknowledgements

The authors would like to thank Bryan Bartley, Jane Lieviant, Wilbert Copeland, and Alec Nielsen for useful comments that have contributed to this protocol. We would also like to thank other members of the Sauro and Klavins synthetic biology groups at the University of Washington for materials and useful discussions. Funding for In-Fusion assembly reactions has come from Sauro Lab start-up funds (University of Washington) and BEACON: An NSF Center for the Study of Evolution in Action (0939454).

References

1. Knight T (2003) Idempotent vector design for standard assembly of Biobricks. *DSpace* http://hdl.handle.net/1721.1/21168
2. Shetty RP, Endy D, Knight TF Jr (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5
3. Kelly JR, Rubin AJ, Davis JH et al (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng 3:4
4. Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol 26:787–793
5. Anderson JC, Dueber JE, Leguia M et al (2010) BglBricks: a flexible standard for biological part assembly. J Biol Eng 4:1
6. Weber E, Engler C, Gruetzner R et al (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765
7. Gibson DG, Young L, Chuang R-Y et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
8. Sleight SC, Bartley BA, Lieviant JA, Sauro HM (2010) In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res 38:2624–2636
9. Zhu B, Cai G, Hall EO, Freeman GJ (2007) In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. Biotechniques 43:354–359
10. Benoit RM, Wilhelm RN, Scherer-Becker D, Ostermeier C (2006) An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids. Protein Expr Purif 45:66–71
11. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
12. Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res 35:e55
13. Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol 3:109–118

Chapter 4

Overlap Extension PCR Cloning

Anton Bryksin and Ichiro Matsumura

Abstract

Rising demand for recombinant proteins has motivated the development of efficient and reliable cloning methods. Here we show how a beginner can clone virtually any DNA insert into a plasmid of choice without the use of restriction endonucleases or T4 DNA ligase. Chimeric primers encoding plasmid sequence at the 5′ ends and insert sequence at the 3′ ends are designed and synthesized. Phusion® DNA polymerase is utilized to amplify the desired insert by PCR. The double-stranded product is subsequently employed as a pair of mega-primers in a PCR-like reaction with circular plasmids. The original plasmids are then destroyed in restriction digests with Dpn I. The product of the overlap extension PCR is used to transform competent *Escherichia coli* cells. Phusion® DNA polymerase is used for both the amplification and fusion reactions, so both steps can be monitored and optimized in the same way.

Key words Overlap extension PCR cloning, Recombinant vector, Phusion, Restriction enzyme ligation independent

Abbreviations

B-ME	Beta-mercaptoethanol
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide triphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth
LiAc	Lithium acetate
LiB	Lithium borate
PCR	Polymerase chain reaction
Phusion HF buffer	Phusion DNA polymerase high fidelity buffer
SOC	Super optimal broth with catabolite repression
TAE	Tris–acetate–EDTA
TBE	Tris–borate–EDTA

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_4, © Springer Science+Business Media New York 2013

1 Introduction

In traditional ligation-dependent cloning, a target gene can be PCR-amplified using specifically designed primers that contain unique restriction sites. The resulting PCR product can then be purified, digested with specific restriction endonucleases and ligated into a vector with compatible cohesive ends [1]. This technique remains in widespread use, but can be time-consuming and unreliable in the hands of novice workers. PCR-mediated cloning methods [2–9] are advantageous because they are easy to learn and troubleshoot. The PCR-mediated recombination of insert and vector DNA [2] is conceptually similar to the QuikChange™ site directed mutagenesis procedure [10], except that the two strands of the PCR product are used as megaprimers in reactions with the plasmid template (Fig. 1). The PCR amplification of the insert (Step A) and the subsequent recombination with vector (Step B) can be catalyzed by the same thermostable polymerase. Both reactions are optimized in the same way, eliminating the need to master the idiosyncrasies of multiple restriction enzymes, polymerases, glycosylases, recombinases, and ligases.

The primary limitation of overlap extension PCR cloning technique is its maximum insert length of approximately 7 kb; it probably works better with smaller cloning vectors as well (*see* **Note 1**). Several factors could explain the observed size-dependence of cloning efficiency. First, transformation efficiency decreases with increasing plasmid size [11, 12]. Second, insert exposure to UV light during purification can damage the DNA, particularly fragments longer than 3 kb [13], beyond the repair ability of *E. coli* DNA repair machinery. Third, the molar quantity of the cohesive ends in the overlap extension PCR cloning reaction decreases with an increase in insert size. As a result, the maximum amount of the plasmid that can be achieved in the reaction also decreases. Finally, the phenomenon could also be related to our use of undigested circular vectors as templates (Part B). Larger inserts possess greater conformational entropy, so a decreased proportion will interact productively with the plasmid template. These obstacles could be overcome by electroporation, DNA stains excited by near-infrared (rather than ultra-violet) light, improved PCR yields and primers with higher T_m overlap regions, respectively. Alternative approaches described in this book, such as Golden Gate assembly [14], Gibson cloning [15], or DNA assembler [16], may be more effective for longer constructs.

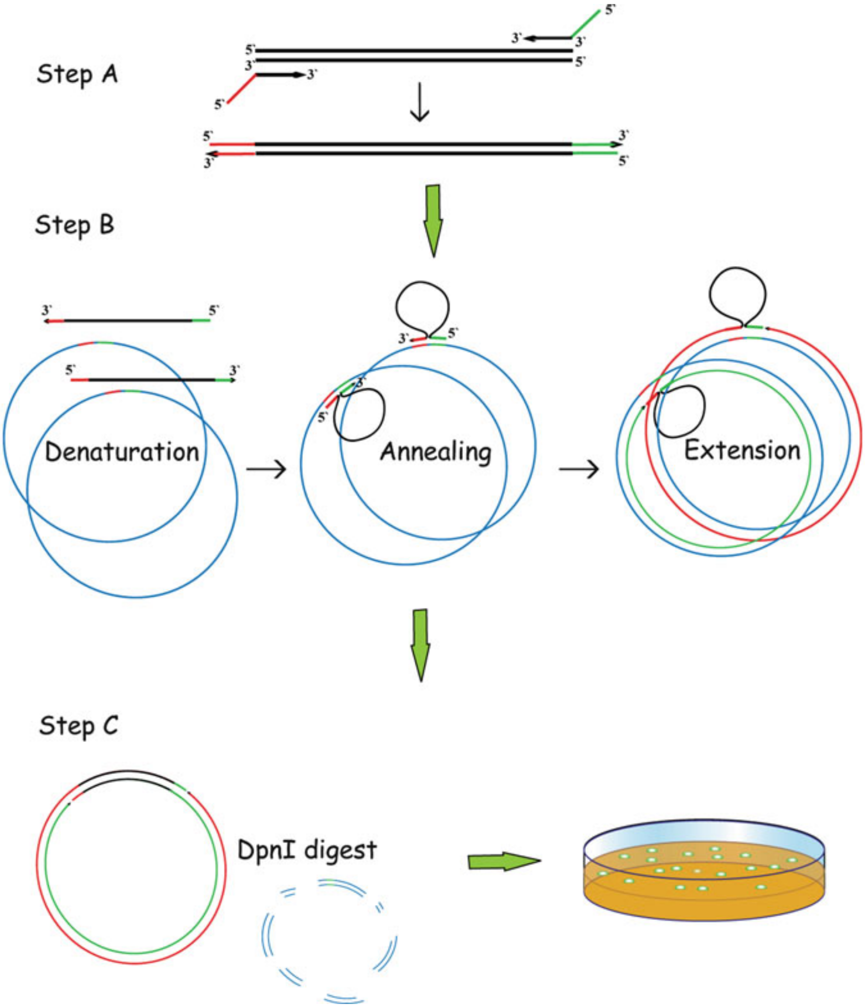


Fig. 1 Overview of overlap extension PCR cloning. (Step A) The insert is PCR amplified with the chimeric primers. The ends of final PCR product overlap regions of the vector. (Step B) Vector and insert are mixed, denatured, and annealed. The hybridized insert is then extended by Phusion® polymerase using the vector as a template until polymerase reaches the 5′-end of the insert. After several PCR cycles, the new plasmid with two nicks (one on each strand) accumulates as a product. (Step C) The parental plasmid is destroyed by DpnI digest, and new plasmid is used to transform *E. coli*

2 Materials

2.1 Equipment

1. Automated thermal cycler.
2. Agarose gel electrophoresis system (to perform analyses of the PCR reactions).
3. UV transilluminator (for DNA detection).
4. Tabletop centrifuge.

5. Silica-based DNA-purification mini columns.
6. 1.5 mL Eppendorf tubes.
7. Thin-walled PCR tubes.

2.2 Supplies and Reagents

1. Sterile deionized water.
2. Oligonucleotide primers 2 μM solution (*see* **Note 2**).
3. Template DNA (to PCR amplify the insert).
4. Destination vector plasmid.
5. Phusion® DNA polymerase.
6. 5× Phusion® HF buffer.
7. *Dpn*I restriction endonuclease.
8. DMSO.
9. dNTP mix (10 mM each in water).
10. 2-Mercaptoethanol, 10 % solution in water.
11. Chemically competent *E. coli* cells prepared according to a previously published protocol [17].
12. 0.8 % agarose gels in appropriate buffer (TAE, TBE, LiAc, LiB, or similar).
13. DNA intercalating dye, e.g., ethidium bromide (to visualize the results of agarose electrophoresis).
14. LB-agar plates with the appropriate antibiotics.
15. Reagents for DNA purification from agarose gel on silica mini columns.
16. DNA ladder.
17. SOC broth (0.5 % Yeast Extract, 2 % Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, autoclave, then add 20 mM Glucose [1]).

3 Methods

The overlap extension PCR cloning reaction must be monitored and optimized like any other long PCR protocol. Amplification yields are poor when the reaction conditions are too stringent (primers fail to anneal) or too relaxed (nonspecific priming). Both can lead to empty lanes in agarose gels, although the latter can also present as smears or undesired bands. Reaction stringency is controlled by altering reactant concentrations (template, primers), annealing temperature, buffer components (magnesium, pH, glycerol, DMSO), or the number of temperature cycles. We optimize PCRs by running multiple reactions in parallel at different annealing temperatures (in gradient-producing thermocyclers) and/or magnesium concentrations. The products are visualized by agarose gel electrophoresis; reactions are further optimized if necessary.

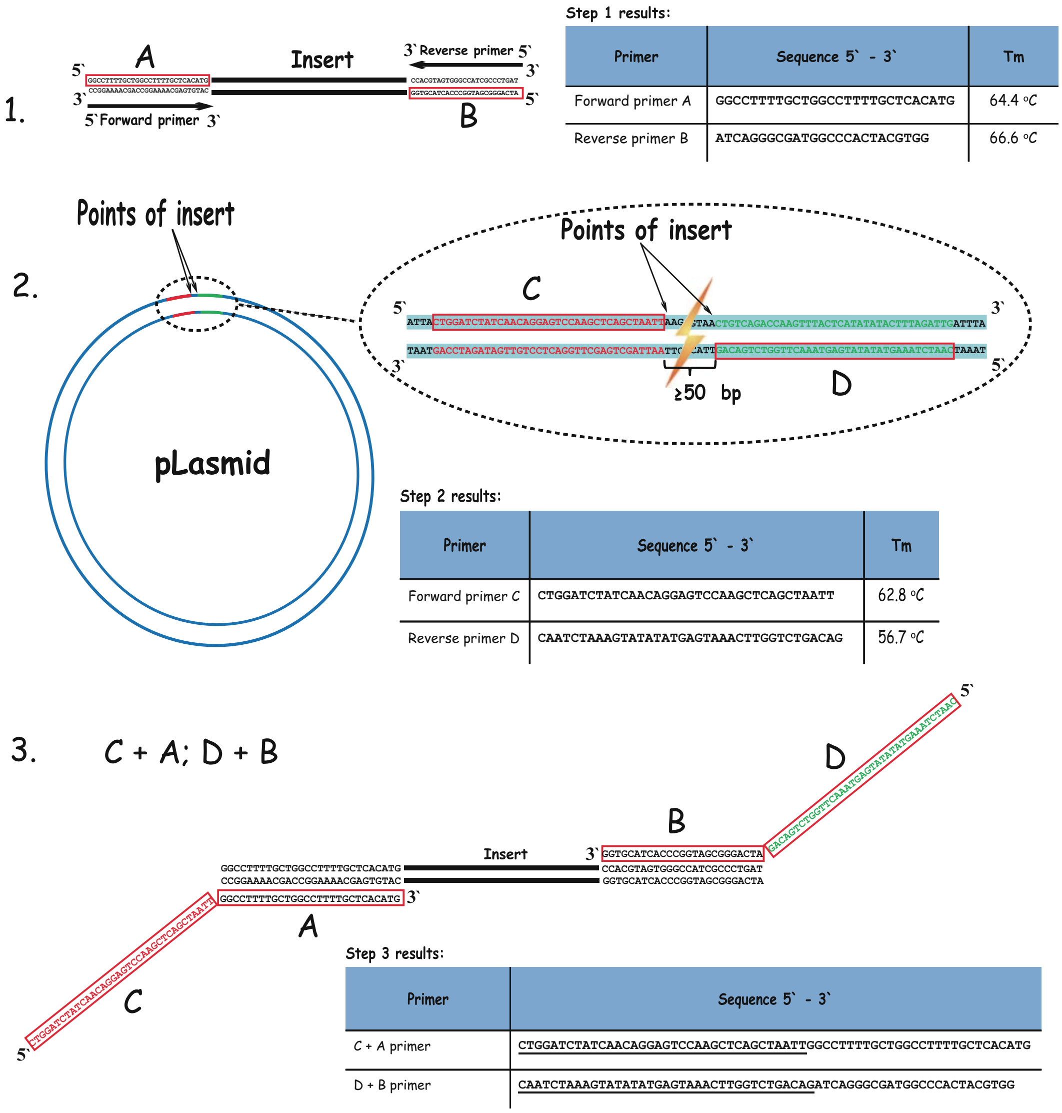
3.1 Primer Design

The primers for overlap extension PCR amplification of an insert (Fig. 1, Step A) must be designed to include 5′ sequence extensions identical to sequences in the vector. These extensions subsequently allow the PCR products to act as giant primers on the vector fragment (Fig. 1, Step B). Based on our experience with PCR-mediated fusion of long fragments [18], we recommend that the plasmid-specific part of the primer be designed to anneal at the high melting temperature (65–68 °C or higher) (*see* **Note 3**). The easiest way to design primers for overlap extension PCR cloning is to use the online "Restriction Free"-cloning tool developed by Steve Bond [19]. We rely upon the following protocol, as it teaches beginners how to design chimeric primers (Fig. 2).

1. Design appropriate primers A and B to PCR amplify the insert using Web-based tools (e.g., Primer3 [20]; Primer Design [21]; or Primer-Blast [22]).
2. Select the desirable insert points on the plasmid; they could be either in close proximity to each other or (preferably) 50 to several hundred base pairs apart. Then select 30–40 bp upstream of the left point of insert on the top strand of the plasmid.
3. Copy this sequence and estimate its T_m using an online tool (Oligo Analyzer [23]). If the T_m is satisfactory, save the sequence as primer C.
4. Select 30–40 bp downstream of the right point of insert on the bottom strand of the plasmid. Copy the reverse sequence and analyze its T_m using an online tool (Oligo Analyzer [23]); if T_m parameter is satisfied, save the sequence as primer D.
5. Attach the sequence of primer C to the 5′ end of primer A. Attach the sequence of primer D to the 5′ end of primer B.

3.2 Insert Preparation (Step A)

1. To PCR amplify the insert, mix the following components in a thin-wall PCR tube:
   - DNA 0.1 ng/μL (to amplify the insert)—1 μL
   - 5× reaction buffer for Phusion® DNA polymerase—20 μL
   - Oligo1 (2 μM solution)—15 μL
   - Oligo2 (2 μM solution)—15 μL
   - dNTP mix (10 mM each)—2 μL
   - DMSO—2 μL
   - H_2O—44 μL
   - Phusion® DNA polymerase (*see* **Note 4**)—0.5 μL
2. Divide reaction between two tubes (50 μL in each tube for a better thermal transfer). Set up the thermal cycler using the following parameters:


1.
A
Insert
5`
3`
GGCCTTTTGCTGGCCTTTTGCTCACATG
CCGGAAAACGACCGGAAAACGAGTGTAC
5`Forward primer 3`
3`Reverse primer 5`
3`
5`
CCACGTAGTGGGCCATCGCCCTGAT
GGTGCATCACCCGGTAGCGGGACTA
B
2.
Points of insert
pLasmid
Points of insert
C
5`
3`
ATTACTGGATCTATCAACAGGAGTCCAAGCTCAGCTAATTAAG GTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTA
TAATGACCTAGATAGTTGTCCTCAGGTTCGAGTCGATTAATTC CATTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTAAAT
3`
5`
≥50 bp
D
3.
C + A; D + B
D
5`
GACAGTCTGGTTCAAATGAGTATATATGAAATCTAAC
B
Insert
3`
GGTGCATCACCCGGTAGCGGGACTA
GGCCTTTTGCTGGCCTTTTGCTCACATG
CCACGTAGTGGGCCATCGCCCTGAT
CCGGAAAACGACCGGAAAACGAGTGTAC
GGTGCATCACCCGGTAGCGGGACTA
GGCCTTTTGCTGGCCTTTTGCTCACATG
3`
A
CTGGATCTATCAACAGGAGTCCAAGCTCAGCTAATT
C
5`

Step 1 results:

Primer	Sequence 5` - 3`	Tm
Forward primer A	GGCCTTTTGCTGGCCTTTTGCTCACATG	64.4 °C
Reverse primer B	ATCAGGGCGATGGCCCACTACGTGG	66.6 °C

Step 2 results:

Primer	Sequence 5` - 3`	Tm
Forward primer C	CTGGATCTATCAACAGGAGTCCAAGCTCAGCTAATT	62.8 °C
Reverse primer D	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG	56.7 °C

Step 3 results:

Primer	Sequence 5` - 3`
C + A primer	CTGGATCTATCAACAGGAGTCCAAGCTCAGCTAATTGGCCTTTTGCTGGCCTTTTGCTCACATG
D + B primer	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGATCAGGGCGATGGCCCACTACGTGG

Fig. 2 Design of the primers for overlap extension PCR cloning. (1) Design simple primers A and B that could hypothetically be used to amplify the insert in a normal PCR. (2) Design simple primers C and D that could in principle be employed to amplify a 50–200 bp region in the parent plasmid that will ultimately be replaced in the insert. (3) Fuse the 3′ end of primer C to the 5′ end of primer A. Similarly, attach the 3′ end of primer D to the 5′ end of primer B

Number of cycles	PCR step	Temperature (°C)	Time
1 cycle	Initial denaturation	100	2 min
30–32 cycles	Denaturation	94	30 s
	Annealing	60	30 s
	Extension	68	1.5 min/kb
1 cycle	Final extension	68	10 min

3. After PCR amplification is complete, assess the DNA fragment size and quality by running 2 μL of the PCR sample on 0.8 % agarose gel. If the expected PCR product is not observed, refer to **Note 5**. If the PCR product is the right size, but the primer–primer dimer band is also present (typically, between 100 and 150 bp), perform agarose gel purification of the PCR product before proceeding to the next step. If the PCR product is the right size and present as a single band, perform silica-column purification of the product. In either case, elute the product from the silica column with the minimum amount of sterile deionized water (typically 25–50 μL) to maximize the concentration of the eluted product.

3.3 Overlap Extension PCR (Step B) and Restriction Digest of Original Plasmid (Step C)

During this step, the insert prepared in Step A is extended over the acceptor plasmid backbone during PCR. Overlap extension PCR (Step B) requires a high GC content in the vector-specific part of the chimeric primers (5′ end). The thermocycler should be set for relatively low annealing temperature (*see* **Note 3**).

1. Mix the following components in thin-wall PCR tube:
   - Vector DNA—30 ng
   - 5× reaction buffer for Phusion® DNA polymerase—20 μL
   - Gel purified insert—30 μL (*see* **Note 6**)
   - dNTP mix (10 mM each)—2 μL
   - DMSO—2 μL
   - H_2O—44 μL (if vector is 1 μL)
2. Divide reaction between *two tubes—negative control* and *reaction* (49.5 μL each); add 0.5 μL of Phusion® DNA polymerase to the reaction tube. Set up the thermal cycler using the following parameters (*see* **Note 7**):

Number of cycles	PCR step	Temperature (°C)	Time
1 cycle	Initial denaturation	100	1 min
18–22 cycles	Denaturation Annealing Extension	95 60 68 (*see* **Note 8**)	50 s 50 s 12 min
1 cycle	Final extension	68	12 min

The program takes approximately 5 h to complete.

3. After the end of the program, add 10 units (typically 0.35 μL) of DpnI restriction enzyme directly to the reaction and the negative control tubes; incubate for 1 h at 37 °C.

The *Dpn*I endonuclease works well in Phusion® HF buffer. Restriction endonuclease *Dpn*I targets methylated DNA sequences and can cleave the DNA template isolated from most *E. coli* strains, but not the PCR product [24].

3.4 Cell Transformation and Analysis of Clones

Ordinary chemically competent *E. coli* (transformation efficiency of 10^6–10^7 cfu/μg) is usually sufficient for gene cloning procedures. Consider high efficiency chemically competent or electrocompetent cells for the construction or transfer of gene libraries.

1. Mix the following components in thin-wall PCR tube:
   - Chemically competent *E. coli* cells—20 μL
   - 10 % B-ME—0.88 μL
2. Incubate on ice (or the thermal cycler at 4 °C) for 10 min
3. Add *Dpn*I treated PCR product from Step B or negative control—1 μL
4. Incubate on ice or in the thermal cycler at 4 °C for 10 min
5. Heat shock for 30 s at 42 °C; then incubate for 2 min at 4 °C
6. Add SOC broth—222 μL
7. Incubate for 1 h at 37 °C
8. Plate 200 μL from each tube on LB/agar plates with the appropriate antibiotic and 100 mM IPTG (if necessary).
9. Leave plates at 37 °C (or at 30 °C for toxic products) for 12–36 h. If there are no colonies on a plate after this interval, *see* **Note 9**.
10. Colonies do not ordinarily grow on the negative control plate (Step B reaction without Phusion® polymerase, *see* **Note 10**). For cloning projects with an insert ≤2 kb, pick two colonies (*see* **Note 11**) and grow them in 5 mL of LB with an appropriate antibiotic overnight. Purify plasmids from the 5 mL cultures

and analyze them by sequencing. For larger inserts, pick at least five colonies and perform the restriction digest of the purified plasmids to confirm that the insert is the correct size before sequencing.

4 Notes

1. We previously determined that the number of colonies produced by overlap extension PCR cloning is inversely proportional to the size of the insert (Fig. 3a), where the maximal possible insert size is 7 kb.
2. Primers can be purchased from major manufacturers such as IDT (Coralville, IA), Operon, Invitrogen, and Sigma without size purification (gel extraction or HPLC).
3. We recommend that the plasmid-specific part of the primer be designed to anneal at high melting temperatures (65–68 °C or higher) because the concentration of megaprimers (PCR product of Step A) is much lower than that of synthetic primers in regular PCRs. Furthermore, the overlapping region could be shorter, and lower in melting temperature, than expected due to the exonuclease activity of residual DNA polymerase. A 250× molar excess of insert DNA (from Step A) over acceptor plasmid template (3–30 ng) is recommended.
4. Theoretically, the first PCR (Step A) can be performed with any thermostable DNA polymerase. We use Phusion® DNA polymerase throughout the entire cloning procedure (Steps A and B) in order to keep things simple. Use of the Phusion® polymerase is absolutely crucial during the second PCR (Step B, Table 1).
5. We optimize PCRs by changing the cycling parameters, amount of DMSO and the concentration of magnesium ions. We also analyze the quality of the template DNA by agarose gel electrophoresis. Phusion® DNA polymerase is reliable over the wide range of PCR parameters in our hands, but sensitive to the quality of the template.
6. We gel purify inserts with the Qiagen Gel Extraction protocol (using a QIAcube) with the elution volume of 50 μL. In our hands, 30 μL of purified insert (out of 100 μL final reaction volume) is the maximum amount of purified insert (from Step A) that can be used during Step B without reaction inhibition.
7. We empirically found that 18–22 PCR cycles in Step B result in the maximum number of colonies formed on the plate.
8. Our original publication in “Biotechniques” [9] mistakenly states the extension temperature as 98 °C instead of 68 °C.

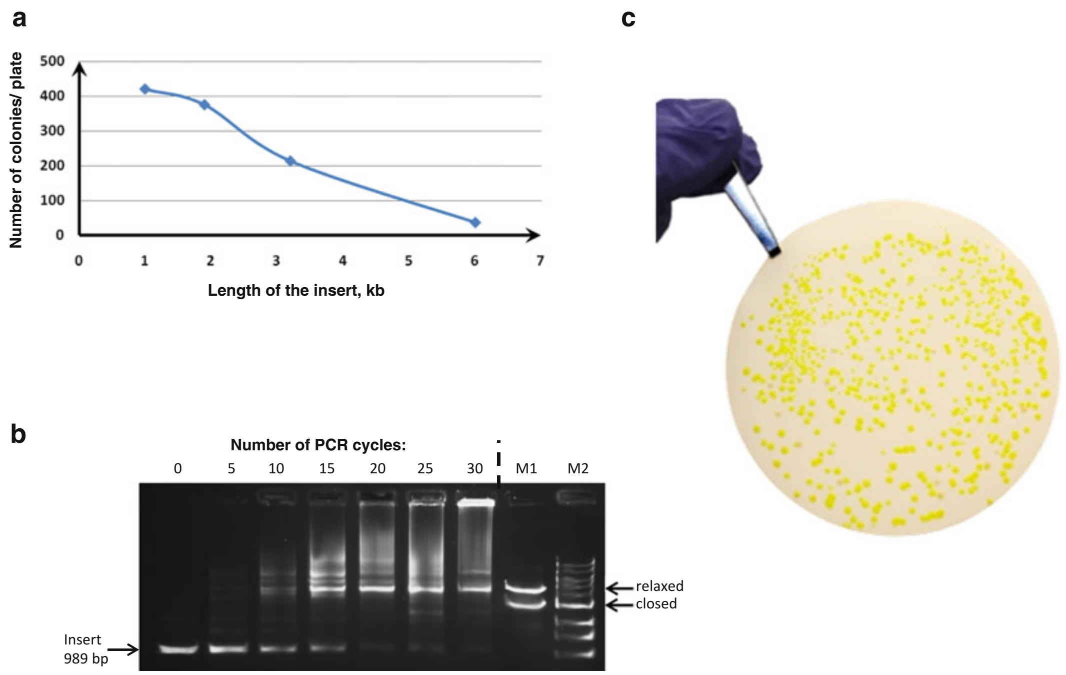


Fig. 3 (**a**) Cloning efficiency of overlap extension PCR is inversely proportional to insert length. Phusion® DNA polymerase was used to PCR amplify DNA encoding the following products: green fluorescent protein (*gfp*, 1 kb), beta-D-glucuronidase (*gusA*, 1.9 kb), β-galactosidase (*lacZ*, 3.2 kb), and the bacterial Lux operon (*luxABCDE*, 6 kb). The PCR products were gel purified and separately combined with the pQE30 vector in the overlap extension PCRs (3 ng of the pQE30 vector was mixed with 175–500 ng of the insert in a total reaction volume of 10 μL and subjected to 18 cycles of PCR). The original plasmid was digested with restriction enzyme DpnI, and the overlap extension PCR products were used to transform competent *E. coli* cells. (**b**) The overlap extension PCR products after 0, 5, 10, 15, 20, 25, and 30 cycles were analyzed by agarose gel electrophoresis (3 ng of the pQE30 vector were reacted with 175 ng of insert, 250-fold molar excess, in a total reaction volume of 10 μL). Four microliters of each reaction were separated on a 0.8 % agarose gel and compared with 1 kb DNA ladder (M2) or the product plasmid (in two forms—closed circular and relaxed circular) purified from *E. coli* (M1). (**c**) Colonies transformed with products of overlap extension PCR. Three nanograms of the pQE30 vector were mixed with 175 ng of the *gfp* insert in a total reaction volume of 10 μL and subjected to 18 cycles of PCR. The original plasmid was digested with DpnI, and the overlap extension PCR products were used to transform competent *E. coli* cells. The transformed cells were spread on LB agar plates supplemented with chloramphenicol. Colonies formed on the plate were lifted using nitrocellulose (for better contrast) to show that the PCR reactions did not introduce many deleterious mutations

9. PCR amplification of some vectors may be challenging. If overlap extension PCR cloning produces no colonies, and if the desired PCR product band is not visible on an agarose gel (Fig. 3b), consider PCR amplification of the overlapped part of the plasmid [25]. This step improves control by eliminating the uncertainty about acceptor plasmid backbone amplification. We recommend purification of both the vector and the insert PCRs before the overlap extension PCR (Step B), in spite of claims to the contrary, for better performance. In our

Table 1
Comparison of different PCR systems in overlap extension PCR cloning efficiency

DNA polymerase	Number of colonies expressing GFP/plate	Number of white colonies/plate
KOD polymerase	14	0
Phusion polymerase	417	8
Expand long template DNA polymerase mix	12	2
Deep vent DNA polymerase	0	0
Pfu polymerase	9	1
Taq polymerase	Not tested	Not tested

Three nanograms of pQE30 vector were reacted with 175 ng of PCR amplified *gfp* insert (250 molar excess) in a total reaction volume of 10 μL. One microliter of the reaction was used to transform 20 μL of *E. coli* competent cells, which were spread on LB agar plates supplemented with chloramphenicol. The green colonies on each plate were counted the next day

experience, the presence of the primer–primer dimers in the original PCR mixture may result in a significant accumulation of the (plasmid–primer dimer) clones.

10. If you see colonies on the negative control plate, check the genotype the strain of *E. coli* used as a source of the acceptor plasmid to see whether it is capable of DNA methylation. If so, test the performance of the *Dpn*I enzyme. Typically, 30 min incubation at 37 °C is sufficient to destroy the entire parental DNA in the mixture.
11. Pick a small number of transformed colonies. In our experience, practically all the colonies for small inserts carry the desired construct (Fig. 3c). Picking the second colony precludes the unnecessary delay if the plasmid in the first colony turns out to be incorrect.

Acknowledgements

This work was supported by the NIH (1 R01 GM086824) and NSF (MCB-0951076).

References

1. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Zuo P, Rabie BM (2009) One-step DNA fragment assembly and circularization for gene cloning. Curr Issues Mol Biol 12:11–16

3. Shuldiner AR, Scott LA, Roth J (1990) PCR-induced (ligase-free) subcloning: a rapid reliable method to subclone polymerase chain reaction (PCR) products. Nucleic Acids Res 18:1920
4. Shuldiner AR, Tanner K et al (1991) Ligase-free subcloning: a versatile method to subclone polymerase chain reaction (PCR) products in a single day. Anal Biochem 194:9–15
5. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441
6. Geiser M, Cebe R, Drewello D, Schmitz R (2001) Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. *Biotechniques* 31:88–90, 92
7. van den Ent F, Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 67:67–74
8. Unger T, Jacobovitch Y et al (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol 172:34–44
9. Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48:463–465
10. QuickChange http://www.stratagene.com/manuals/210513.pdf. Stratagene
11. Froger A, Hall JE (2007) Transformation of plasmid DNA into E. coli using the heat shock method. J Vis Exp 2007:253
12. Yoshida N, Sato M (2009) Plasmid uptake by bacteria: a comparison of methods and efficiencies. Appl Microbiol Biotechnol 83:791–798
13. Shevchuk NA, Bryksin AV (2007) Construction of long DNA molecules from multiple fragments using PCR. In: Hughes S, Moody A (eds) PCR (Methods Express Series). Scion Publishing, London, pp 197–216
14. Engler C, Marillonnet S (2011) Generation of families of construct variants using golden gate shuffling. Methods Mol Biol 729:167–181
15. Gibson DG, Young L et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
16. Shao Z, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37: e16
17. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of Escherichia coli with plasmids. Gene 96:23–28
18. Shevchuk NA, Bryksin AV et al (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res 32:e19
19. Restriction Free Cloning (http://www.rf-cloning.org/)
20. http://primer3.sourceforge.net/
21. http://www.bioinformatics.org/jambw/5/2/index.html
22. http://www.ncbi.nlm.nih.gov/tools/primer-blast/
23. http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/
24. Gomez-Eichelmann MC, Lark KG (1977) Endo R DpnI restriction of Escherichia coli DNA synthesized in vitro. Evidence that the ends of Okazaki pieces are determined by template deoxynucleotide sequence. J Mol Biol 117:621–635
25. Li C, Wen A et al (2011) FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnol 11:92

Chapter 5

One-Step Isothermal Assembly of DNA Fragments

Rui T.L. Rodrigues and Travis S. Bayer

Abstract

The One-Step Isothermal DNA Assembly method allows for the efficient assembly of DNA constructs using fragments up to several hundred kilobases in as little as 15 min. Applications of this method range from the addition of promoters to expression constructs to the assembly of bacterial genome fragments. The production of circularized DNA using this method also enables the direct transformation of target organisms, bypassing intermediate transformations for plasmid propagation in those species where expression could lead to toxicity and cell death. Variations of the method allow for specific cloning tasks to be performed, as well as the use of microarray slides as a source of DNA. The level of precision and simplicity of this method makes it a valuable tool for most cloning efforts and all levels of proficiency in molecular biology.

Key words Gibson assembly, Cloning method, Isothermal assembly, Chew-back and anneal

1 Introduction

The creation of the first chemically synthesized genome at the J. Craig Venter Institute was an impressive technical feat and an advance toward whole genome engineering projects [1]. However, the impact of the cloning and DNA assembly strategies developed during that research project is potentially more far reaching than the resulting functional synthetic *M. mycoides* genome [2]. In the assembly of DNA fragments scale up to a few hundreds of kilobases, the One-Step Isothermal DNA Assembly (also known as Gibson Assembly) method has provided researchers with a simple and efficient method for the creation of DNA molecules from simple expression constructs to fractions of small bacterial genomes and complete mitochondrial genomes [3].

By using an enzyme mix containing a exonuclease, a DNA polymerase and a DNA ligase it is possible to (a) produce ssDNA overhangs in DNA fragments by nucleotide excision at their termini, (b) prime the re-polymerization of the ssDNA by designing homologous overhangs, and (c) ligate the nicked DNA resulting

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_5, © Springer Science+Business Media New York 2013

from the complete re-polymerization of the previously excised nucleotides, obtaining a closed circular dsDNA molecule ready for transformation.

All of the steps in this method are performed at the same temperature (50 °C) making this method extremely simple and reliable—which is made possible by the design of the enzyme mix and reaction conditions. The choice of the T5 exonuclease is due to its ability to create 5′ overhangs in linear dsDNA through its exonuclease activity, without cleaving closed circular DNA (reaction product) and its inactivation after 15 min at 50 °C; Phusion DNA polymerase was selected to fill in the gaps due to its proofreading capability, removing noncomplementary sequences; while the Taq DNA ligase has high activity and fidelity at the reaction temperature (allowing for an efficient termination of the reaction) and does not catalyze the ligation of blunt ends avoiding incorrect assembly that would result of other enzymes with the same function.

Preparation of fragments for assembly can be done by PCR—target sequences are amplified with primers containing an extra 5′ sequence that overlaps with the fragment to be assembled adjacently. This overlap should be of at least 40 bp to obtain good efficiency, but shorter overlaps from 20 bp can be used.

Although this method has been designed for the generation of genome cassettes, it can also be used for the insertion of smaller fragments into plasmids, such as promoters [4]. In addition, the method does not generate DNA sequence "scars" such as when cloning via restriction sites. Another useful application of the method is for the assembly of expression constructs containing proteins that would have deleterious effects when expressed in *E. coli* [5]—these plasmids can be transformed immediately into the target organism as the high efficiency of the method produces large amounts of double-stranded and circularized product (Fig. 1).

2 Materials

1. *Isothermal reaction buffer* (*3 mL*): Combine 1.5 mL of 1 M Tris–HCl pH 7.5, 75 μL of 2 M MgCL, 30 μL of 100 mM dATP, 30 μL of 100 mM dCTP, 30 μL of 100 mM dGTP, 30 μL of 100 mM dTTP, 150 μL of 1 M DTT, 0.75 g of PEG-8000 (*see* **Note 1**), 150 μL of 100 mM NAD, and ultrapure water up to 3 mL. Aliquot 100 μL of the buffer into 0.5 mL tubes (30 aliquots) and store at −20 °C.
2. *33× Assembly Master Mix* (*375 μL*): Combine 2,000 units of Taq DNA ligase, 12.5 units of Phusion DNA polymerase, and 2 units of T5 exonuclease when using overlaps shorter than

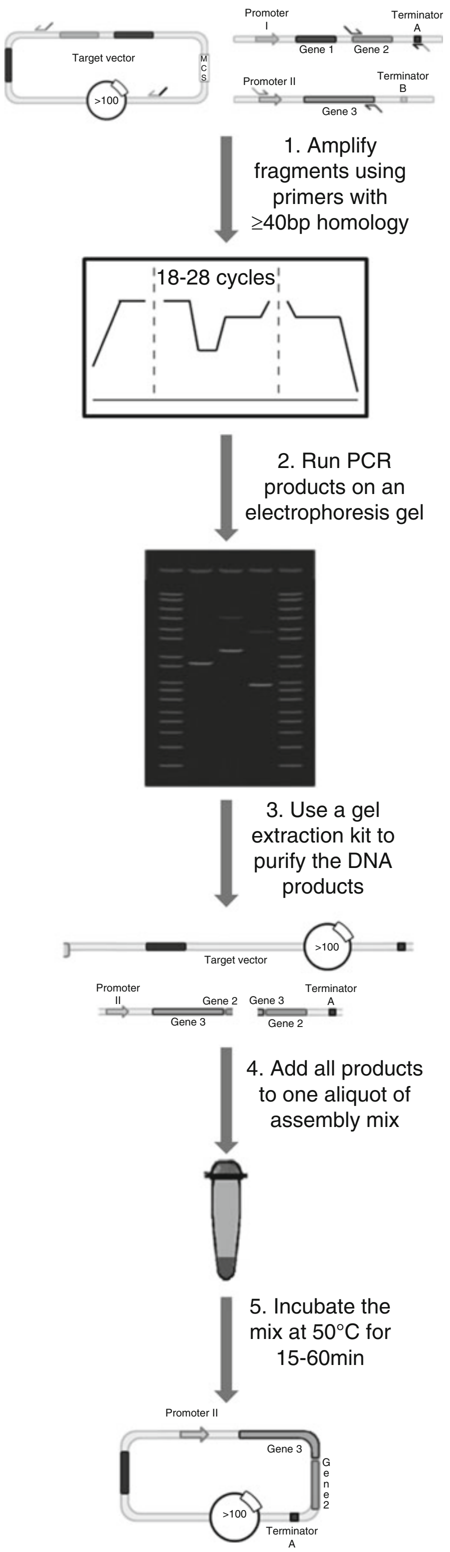

Target vector
M C S
>100
Promoter I
Gene 1
Gene 2
Terminator A
Promoter II
Gene 3
Terminator B
1. Amplify fragments using primers with ≥40bp homology
18-28 cycles
2. Run PCR products on an electrophoresis gel
3. Use a gel extraction kit to purify the DNA products
>100
Target vector
Promoter II
Gene 2
Gene 3
Gene 3
Terminator A
Gene 2
4. Add all products to one aliquot of assembly mix
5. Incubate the mix at 50°C for 15-60min
Promoter II
Gene 3
Gene 2
>100
Terminator A

Fig. 1 Example: assembly of an expression operon from three parts

150 bp or 10 units if longer, ultrapure water up to 375 μL. Aliquot 15 μL into 0.2 mL tubes (25 aliquots) and store at −20 °C (*see* **Note 2**).

3 Methods

1. Amplify the target sequences using the appropriate primers (*see* **Note 3**).
2. Purify the PCR products and quantify the amount of DNA obtained (*see* **Notes 4** and **5**).
3. Mix the fragments to be assembled in an equimolar solution with a total volume of 5 μL (*see* **Notes 6** and 7).
4. A 15 μL aliquot should be thawed on ice and a water bath or thermocycler unit set to 50 °C.
5. Add the DNA mix to the thawed aliquot and transfer the tube containing the complete assembly mixture to the water bath or thermocycler (*see* **Note 8**).
6. Wait 15–60 min (*see* **Note 9**).
7. Transform into the desired cells using up to 1 μL of the sample if transforming by electroporation and up to 10 μL if using chemically competent cells.

4 Notes

1. Addition of PEG-8000 should be done slowly to avoid formation of clumps.
2. 15 μL aliquots can be stored at −20 °C for up to 1 year. The master mix will still be active after ten freeze–thaw cycles.
3. A number of cycles between 18 and 25 is appropriate for this method. Use 25–50 μL of PCR mix for fragments up to 3 kb, scale up volumes for larger products.
4. Gel purification should be used to guarantee that the expected products were produced—PCR purification will suffice as an alternative but digestion with *Dpn*I may be required to avoid empty vector carrying colonies.
5. A relatively high DNA concentration is required for the method to be efficient, consider eluting the sample in two steps—first in a small volume (15–30 μL), followed by the complete elution of the sample (50–80 μL). 10 ng per 6 kb of DNA can be used as a reference for the amount of DNA to be added.

6. Larger amounts of DNA (50–100 ng per 6 kb fragment) can be used to increase the efficiency of the cloning procedure—higher concentrations will help offset parameters such as cell competence or assembly of large number of fragments.
7. Due to the small volume available for this mix consider precipitating the purified DNA or concentrating the sample by eluting samples in small volumes after using silica column-based PCR clean-up protocols.
8. Large DNA fragments (>6 kb) will frequently be purified at low concentrations, resulting in the need for a large part of the reaction volume to be taken up by them. Total reaction volume can be adjusted to accommodate a higher volume of the DNA mix and will sometimes be preferable to pursuing precipitation or concentration of the samples.
9. 60 min is optimal for Gibson assembly; however, assemblies can be done efficiently within 15 min. When assembling libraries, the mixture can be left reacting for longer periods of time to guarantee a more extensive assembly.

References

1. Gibson DG, Benders GA et al (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319:1215–1220
2. Gibson DG, Glass JI et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56
3. Gibson DG, Smith HO et al (2010) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7:901–903
4. Krishnakumar R, Assad-Garcia N et al (2010) Targeted chromosomal knockouts in Mycoplasma pneumoniae. Appl Environ Microbiol 76:5297–5299
5. Widmaier DM, Voigt CA (2010) Quantification of the physiochemical constraints on the export of spider silk proteins by salmonella type III secretion. Microb Cell Fact 9:78–90

Part III

Part, Pathway, and Genome Engineering

Chapter 6

Creation and Characterization of Component Libraries for Synthetic Biology

Tim Weenink and Tom Ellis

Abstract

Large numbers of well-described components are essential for advanced synthetic biology and model-guided design of pathways and regulatory networks. Here a method is presented for the creation of libraries of novel control elements. From these libraries, parts with well-defined properties can be selected and used in construction of finely tuned synthetic systems. The example of the PFY1 promoter in *S. cerevisiae* is used to describe library creation using degenerate synthetic oligos and the circular polymerase extension cloning (CPEC) method. Additionally the workflow of screening the raw library for functional parts is included to provide a full overview of the process of creating and characterizing a component library for synthetic biology.

Key words Components, Parts, Libraries, Screening, Synthetic biology, Characterization, Regulation, Promoter, Yeast, Engineering

1 Introduction

Developments in the field of synthetic biology have led to an increasing number of available genetic devices. However, the available biological parts and ways to connect them often limit the design and realization of devices and systems. Here, a method for the creation of novel control elements is presented. These elements are created as libraries of biological parts that vary only with respect to a limited set of parameters. These libraries thus facilitate tuning of devices guided by mathematical modelling of the system.

Here, the synthesis and screening steps are described for the example of the constitutive PFY1 promoter in *S. cerevisiae*, generating a library with diversity only in DNA sequence and the promoter output parameter. Promoters are crucial components in genetic networks and are often used in the rewiring of control circuits. The library synthesis approach described here is not limited to promoters or to *S. cerevisiae*, but is broadly applicable to all genetic parts following appropriate protocol adaptations.

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_6, © Springer Science+Business Media New York 2013

Three main methods exist for producing component libraries: (1) combinatorial shuffling of short DNA sequences [1–4], (2) targeted mutation and selection of an existing component [5, 6], and (3) targeted nucleotide replacement using synthetic oligos [7, 8]. Here, a work-through of the third method is described, and is recommended as it is rapid and compatible with rational design of new components. This protocol was adapted from a previously documented protocol [9, 10].

2 Materials

2.1 DNA Amplification

1. NEBuffer 2 (NEB, Beverly, MA).
2. 10× BSA (NEB, Beverly, MA).
3. DNA polymerase I, large (Klenow) fragment (NEB, Beverly, MA).
4. 0.25 mM dNTP mix (NEB, Beverly MA).
5. 0.04 M NaOH solution (0.02 M if working directly with colony material).
6. Phusion DNA polymerase (NEB, Beverly, MA), or other non-strand displacing polymerase.

2.2 Transformations

1. High-efficiency competent *DH5α E. coli*, prepared fresh [11] or commercially acquired (NEB, Beverly, MA).
2. LB medium (sigma). Dissolve mixture according to manufacturer specifications or prepare the following in 1 L of deionized water: 10 g tryptone, 5 g yeast extract, and 10 g NaCl. Autoclave or filter-sterilize.
3. Ampicillin (sigma) dissolve to 100 mg/mL and filter sterilize for a 1,000× stock solution.
4. YPD medium: 1 % yeast extract, 2 % peptone in purified water (autoclaved) supplemented with 2 % glucose. Glucose has to be separately autoclaved (usually as a 50 % w/v solution).
5. Auxotrophic marker selection plates: prepare 900 mL deionized water with 6.7 g yeast nitrogen base (YNB without aminoacids) and 1.4 g dropout mix (lacking His, Tryp, Leu, and Ura). Set the pH of this solution to 5.6 and stir on a hot plate until all components are dissolved. Add 20 g agar and autoclave. Separately prepare 1 % solutions of L-Leucine, L-Histidine HCL, and L-Tryptophan and filter sterilize. Add 10, 2, and 2 mL (respectively) of these solutions to the autoclaved bottle where needed. Finally, add 40 mL filter-sterilized 50 % glucose and add sterile water to get a final volume of 1 L. For more information, *see* [12].

3 Methods

3.1 Design of Component Library

For diversity-inherent component library synthesis, it is recommended to base the component design on an existing well-characterized component, where essential sequence motifs are understood and consensus sequences are available. Additional sequence motifs (e.g., for regulation) can be introduced to this component, and if designed correctly, should behave as intended. It is also possible with this technique to create an entirely new hybrid component by linking together several characterized motifs in a rational design [13].

For the design, the core sequences of essential motifs are specified and the base sequences between these are unspecified (i.e., defined as "N"). It is recommended that the precise nucleotide distance between motifs is maintained; otherwise, their function can be lost. More "N" sequence in the design (including some within motifs) ensures more variation in the final component library but will require more screening as a higher percentage of components risk being nonfunctional.

To avoid the constraints imposed by restriction enzyme cloning, circular polymerase extension cloning (CPEC) is used for the insertion of the component library into the selected vector [14, 15]. This is a sequence-independent cloning method, which avoids any need to add restriction sites to the middle of the promoter.

3.2 *In Vitro* Synthesis of Component Library

Three methods exist for physically assembling the DNA for the component; (1) custom synthesis order from a commercial vendor, (2) *Taq* polymerase extension from one long primer [16], and (3) *Klenow* polymerase extension of two annealed long oligonucleotides [8]. Described here is the third method, which allows for up to 200 bp of diversified sequence from two affordable oligonucleotides. A schematic is shown in Fig. 1.

1. Design library oligonucleotides, defining each unspecified nucleotide as "N" (*see* **Note 1**). To allow annealing of oligos, a region of 10–20 conserved base pairs is required to fall about halfway in the component design and each oligo will end with part or all of this sequence. This can usually be made to coincide with a fixed motif such as a transcription factor binding site. Take care to conserve the distance between fixed motifs.
2. Design vector amplification oligos that overlap with the library oligos and will anneal to the vector allowing extension away from the library fragment. *See* Fig. 1.
3. Purchase oligonucleotides (*see* **Note 2**) and upon receipt, dilute oligos in sterile water to 10 μM.

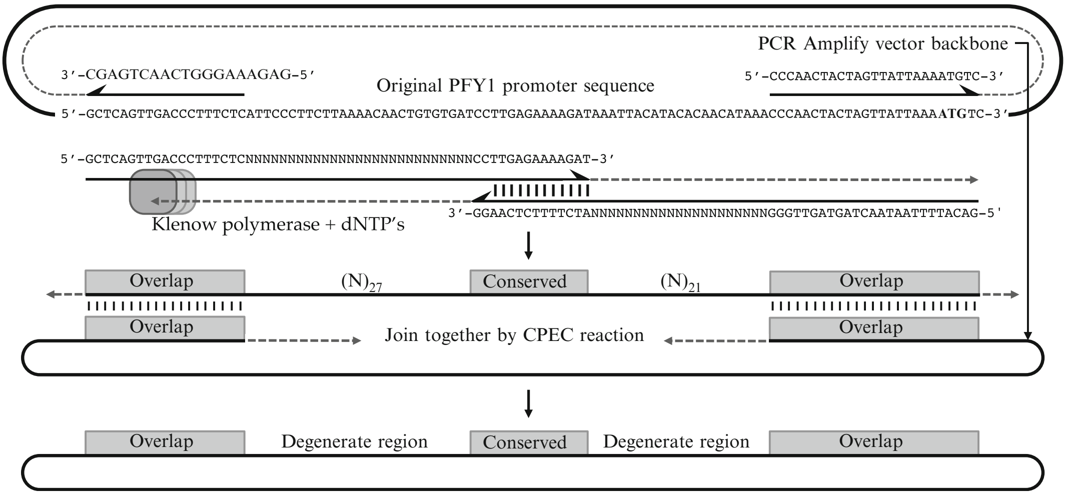

PCR Amplify vector backbone
3′-CGAGTCAACTGGGAAAGAG-5′
Original PFY1 promoter sequence
5′-CCCAACTACTAGTTATTAAAATGTC-3′
5′-GCTCAGTTGACCCTTTCTCATTCCCTTCTTAAAACAACTGTGTGATCCTTGAGAAAAGATAAATTACATACACAACATAAACCCAACTACTAGTTATTAAAATGTC-3′
5′-GCTCAGTTGACCCTTTCTCNNNNNNNNNNNNNNNNNNNNNNNNNNNCCTTGAGAAAAGAT-3′
Klenow polymerase + dNTP's
3′-GGAACTCTTTTCTANNNNNNNNNNNNNNNNNNNNNGGGTTGATGATCAATAATTTTACAG-5′
Overlap
$(N)_{27}$
Conserved
$(N)_{21}$
Overlap
Overlap
Join together by CPEC reaction
Overlap
Overlap
Degenerate region
Conserved
Degenerate region
Overlap

Fig. 1 Design of primers and oligos for a synthetic promoter library based on the PFY1 promoter. A pair of short primers amplifies the vector backbone. The corresponding annealing sites also form the overlap with the degenerate fragments. Two long oligos with degenerate regions ("N") anneal at the center and are extended by Klenow polymerase to form a double-stranded insert fragment. This is then joined to the amplified vector backbone through the CPEC reaction to create a synthetic promoter library

4. In a PCR tube, mix 2 μL of each oligo, 10 μL of NEBuffer 2, 10 μL of 10× BSA, and 75 μL of sterile water. Float the tube in a bath of fast-boiling water for 5 min and then leave the water bath and tube to cool down to slowly room temperature for ~2 h (*see* **Note 3**).
5. Add 1 μL DNA polymerase I, large (Klenow) fragment, and 1 μL 0.25 mM dNTP mix and incubate for 1 h at 37 °C. Inactivate the enzyme with 10 min at 75 °C and cool slowly to room temperature.
6. Run gel electrophoresis (*see* **Note 4**) and purify the correct size double-stranded DNA band using agarose gel electrophoresis and QIAquick Gel Purification kit.

3.3 Design of Characterization Vector

To screen and characterize the component in the host system requires a vector to introduce and maintain the component in the system as well as any other genetic material needed to assess the component.

For the example of a promoter library, the vector requires (1) site for library promoter insertion, (2) a reporter gene coding sequence downstream of the library promoter cloning site, and (3) the usual genes and sequences for selecting transformants and vector propagation in bacteria.

Specifically for working in *S. cerevisiae* an auxotrophic marker and site for integration into the yeast genome are required and it is also advisable to have terminator sequences following the reporter gene.

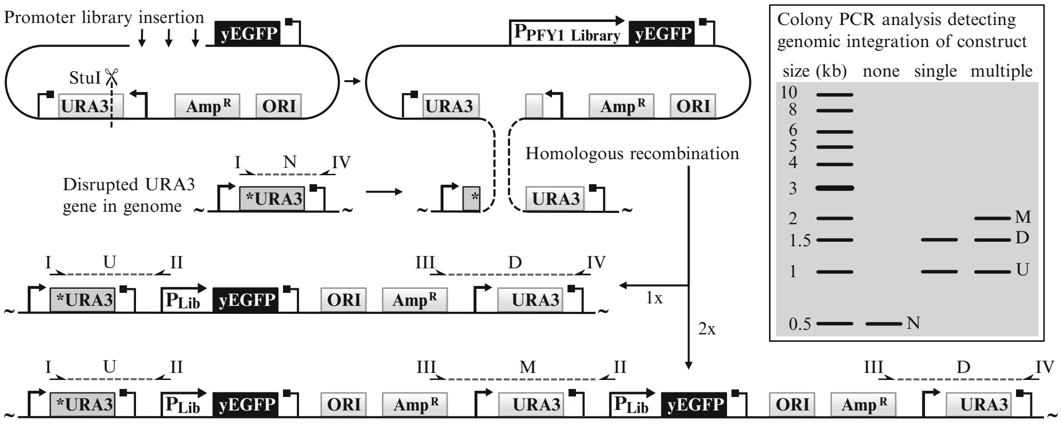


Fig. 2 Screening for multiple genomic integration of the library construct. Screening vector (*upper left*) contains yeast enhanced green fluorescent protein (yEGFP), a terminator, an origin of replication (ORI) for bacterial propagation, the ampicillin resistance gene (Amp^R), and a copy of the yeast URA3 auxotrophic marker gene. After library fragment insertion, the plasmid is linearized with StuI and yeast is transformed with this construct. Homologous recombination will target the construct to a disrupted URA3 gene. *Asterisk* denotes a disrupting Ty transposon element to which primer I anneals. Primer IV anneals downstream on the genome and when no construct is integrated will produce a short fragment N with primer I. When a single construct is integrated, primer I and IV will produce an upstream (U) and downstream (D) fragment with primer II and III, respectively. When multiple constructs have been integrated, a third fragment (M) will be produced by primers II and III. When it is observed, the corresponding colony should be discarded. The *inset* shows the relative sizes of the fragments simulated on an agarose gel

An example vector is shown in Fig. 2 which used yeast enhanced green fluorescent protein (yEGFP).

3.4 Assembly of Library and Introduction into Host System

The following procedure describes the final assembly of the library. However, following this procedure once will result in a library that contains a relatively large number of original promoter sequences. This is because the template used for the amplification of the characterization vector cannot be selectively removed and will therefore still contain the original promoter sequence. It is therefore recommended that the following steps are first followed on a small scale to create a small library of 8–16 clones. These can then be screened (*see* Subheading 3.6, **steps 1** and **2**) for a *null-output* promoter (*see* Fig. 3). When a characterization vector with this promoter is used as template in the second (large scale) iteration of the library creation, all clones with this sequence will be discarded in **step 2** of the screening process because they have low output.

1. In preparation of the CPEC reaction, amplify the characterization vector in a PCR reaction with the primers that overlap with the library component (*see* Subheading 3.3 and **Note 5**). Perform gel electrophoresis on the product and purify using QIAquick Gel Purification.

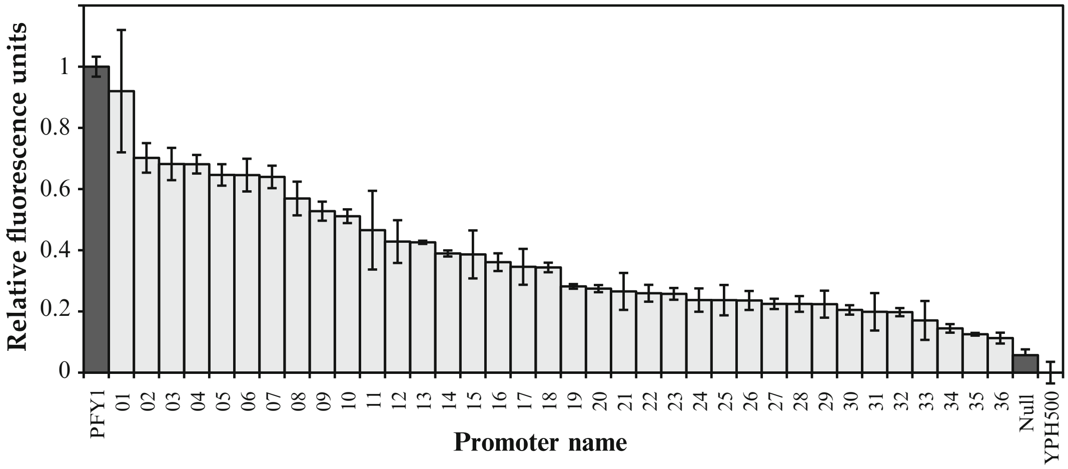


Fig. 3 Characterization of PFY1-derived synthetic promoter library. Fluorescence output of library promoters driving expression of a single chromosomal copy of yEGFP as determined by flow cytometry in triplicate. Fluorescence represents the mean average of geometric mean values as a proportion of the original PFY1 sequence. *Error bars* show 1 standard deviation from the mean

2. Run the CPEC reaction using the purified library piece and vector backbone. Use the fragments in a 1:1 M ratio with a total concentration of 10–20 ng/μL. For the CPEC reaction, only non-strand displacing DNA polymerases can be used. Phusion DNA polymerase is recommended. Run the reaction for 10 cycles under the following conditions. 30 s initiation and denaturation at 98 °C, 30 s annealing at 55 °C (*see* **Note 6**), 15 s/kb extension at 72 °C. And finally an extra extension period equivalent to 1–2 times the normal extension step at the end.

3. Large-scale transform *E. coli* cells with 1 μL unpurified CPEC product (*see* **Note 7**). Electrocompetent or chemically competent cells are both usable, but the highest efficiency possible is desired to get a diverse library. High-efficiency chemically competent *DH5α E. coli* are recommended, using 4 reaction aliquots per library and plating out transformation products over 16 LB + Amp agar 90 mm plates (four plates per transformation). Follow the transformation protocol accompanying the competent cells, and also include a negative control, plating 1/4 of this transformation on a single plate. Leave for around 12 h at 37 °C until colonies are clearly visible but still small (less than 1 mm diameter).

4. If transformation is successful (i.e., number of colonies on each plate is significantly greater than that on the control plate), then these plates form the raw bacteria library and if intended for use in *E. coli*, then the clones are ready to be screened.

5. For work in other systems, collect all of the bacteria by directly scraping-off all colonies on each plate into 1 mL LB using a sterile plate spreader. Pool the colonies in a single tube, spin down the cells, extract all plasmid into sterile water using a QIAprep Spin Miniprep Kit using one column per 4 plates and then determine the yield and DNA concentration.
6. The pooled plasmid library can be used to transform the host system. Yeast transformation is described in detail elsewhere [11]. For genomic integration, a restriction enzyme digestion at the integration site is performed for 1 μg of pooled plasmid. Purify this digest reaction, eluting into water using a QIAquick purification column and use this linearized DNA to transform yeast following a standard LiAc protocol but scaled up by 10 times and plating out on ten auxotrophic marker selective agar plates. This can be done on 20 normal (90 mm) plates, 8 large (135 mm) plates, or two 250 × 250 mm plates. To check transformation efficiency, a single negative control transformation with pure water can be performed and plated onto a single 90 mm plate. Incubate plates for 48 h at 30 °C.
7. If transformation is successful (i.e., thousands of colonies on the library plates and very few on the negative control), then these plates form the raw yeast library and clones are ready to be screened.

4 Screening of Component Library

Screening the component library requires taking individual colonies from the library plates, growing these and assaying them for intended function. By using 96-well format methods, a large number of colonies can be screened rapidly.

For the yeast PFY1 constitutive promoter library, the procedure is as follows:

1. In each well of two to four 96-well plates, inoculate 300 μL of growth media with a single colony from the raw library and incubate for 22 h at 30 °C. If there are no further requirements for selection or induction, YPD can be used.
2. Estimate output of the raw library of promoters by parallel measurement of reporter gene output. For yEGFP, cell fluorescence is measured at 450 nm using a high throughput flow cytometer or FACS machine. If this equipment is not available, a fluorescence-capable Plate Reader can also be used. In that case, use 96-well plates with a transparent bottom. For all measurements, include a negative control not expressing

yEGFP and a positive control with the original promoter sequence. Discard any colonies that do not show significant fluorescence above the negative control. This step typically eliminates 60–80 % of the transformants.

3. Perform an integration-check PCR on the remaining colonies using appropriate PCR oligos (*see* **Note 8**). For colony PCR of yeast, take 10 μL of culture and add it to 10 μL 0.04 M NaOH. Boil in a thermocycler for 10 min at 98 °C. From this mixture, take 1 μL per 25 μL PCR master mix and perform PCR, resolving products with a standard agarose gel. Discard colonies identified as multiple-integrants, typically eliminating 20–30 % of the library. *See* Fig. 2 for a schematic representation of this process and primer design.
4. While taking samples for the colony PCR, simultaneously inoculate a new 96-well plate with the selected colonies.
5. Using the lysed cells produced in **step 3**, PCR amplify the created library components confirmed to be single integrants. Have these components Sanger sequenced through the putative promoter regions of the construct and discard any duplicates (of the original sequence or within the library) or samples that fail sequencing. This typically eliminates 10–20 % of the library. The remaining colonies form the candidate library.
6. Choose your final library (e.g., 24 clones) from the candidates, selecting to have as diverse as possible range of output levels (*see* **Note 9**). Inoculate 96-well plate with the selected colonies.

4.1 Characterization of Component Library

Characterizing the selected component library is a scaled-up, parallel equivalent to characterization of a regular component. A library designed to have diversity in one or two properties should be assayed carefully for these, with biological repeats to ensure quality data. Theoretically, other properties should be conserved throughout the library, so only one member needs to be characterized for these.

For the example of yeast constitutive promoters, the procedure is as follows:

1. In 12 mL culture tubes, inoculate 3 mL YPD medium for each member of the library. Incubate for 20 h at 30 °C with orbital shaking, to an approximate optical density at 600 nm (OD_{600}) of 1.00.
2. Measure the expression for each promoter with a high-resolution assay. For fluorescent reporters, flow cytometry is ideal.
3. Repeat **steps 1** and **2** to get biological repeat data. For each promoter in the library, calculate the mean expression intensity.

Include a positive and negative control as before. An example of the final result for the PFY1 promoter is shown in Fig. 3.

4. Characterize other required properties of the component as appropriate, either just for a single example member or for as many as necessary.

5 Notes

1. In order to reduce the heterogeneity of the library, the mixture of the “N” bases can be adjusted during synthesis by “spiking” to match the original base 70 % of the time and have a 10 % chance of each of the remaining bases. Note, however, that this will significantly increase the price of the oligos.
2. Above 100 nt length it is essential to get PAGE-purification and usually required to order a higher yield.
3. Allow the air inside the tube to heat up before closing it, or use a CapLock to prevent the tube from popping open while in the water bath.
4. In case the library fragment is small (<80 bp), it may be necessary to use a higher percentage agarose gel (e.g., 2.0 %).
5. In this reaction, use as little template DNA as possible (0.1–0.01 ng). This will avoid the original promoter sequence from being overrepresented in the library.
6. Reaction conditions are dependent on the specific primers and DNA polymerase that are used. Follow the instructions in the manual of the used DNA polymerase to determining the correct reaction conditions.
7. If the transformation efficiency is low, make sure that the amplified vector and library piece are not exposed to UV radiation during the purification step, because this lowers transformation efficiency significantly.
8. When integrating constructs into the genome, a fraction of transformants will contain more than one copy of the construct. This will affect observed promoter strength and should therefore be avoided. Using this PCR check, multiple copies of the construct can be detected.
9. If the data from the initial screen is low resolution, it may be worthwhile to repeat the measurements with a higher resolution technique (such as flow cytometry) before selecting the final candidates.

References

1. Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Syst Biol 3:145
2. Gertz J, Siggia ED, Cohen BA (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature 457:215–218
3. Ligr M et al (2006) Gene expression from random libraries of yeast promoters. Genetics 172:2113–2122
4. Murphy KF, Balazsi G, Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. Proc Natl Acad Sci U S A 104:12726–12731
5. Alper H et al (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683
6. Alper H et al (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565–1568
7. Jensen PR, Hammer K (1998) Artificial promoters for metabolic optimization. Biotechnol Bioeng 58:191–195
8. Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64:82–87
9. Ellis T (2009) Synthesis and screening of regulatory component libraries for synthetic biology. Protoc Exch. doi:10.1038/nprot.2009.79
10. Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol 27:465–471
11. Warren DJ (2011) Preparation of highly efficient electrocompetent Escherichia coli using glycerol/mannitol density step centrifugation. Anal Biochem 413:206–207
12. Amberg DC, Burke D, Strathern JN (2005) Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual. CSHL, New York
13. Jeppsson M et al (2003) The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. Yeast 20:1263–1272
14. Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242–251
15. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441
16. Solem C, Jensen PR (2002) Modulation of gene expression made easy. Appl Environ Microbiol 68:2397–2403

Chapter 7

In Vivo and In Vitro Characterization of σ^{70} Constitutive Promoters by Real-Time PCR and Fluorescent Measurements

James Chappell and Paul Freemont

Abstract

The characterization of DNA regulatory elements such as ribosome binding sites and transcriptional promoters is a fundamental aim of synthetic biology. Characterization of such DNA regulatory elements by monitoring the synthesis of fluorescent proteins is a commonly used technique to resolve the relative or absolute strengths. These measurements can be used in combination with mathematical models and computer simulation to rapidly assess performance of DNA regulatory elements both in isolation and in combination, to assist predictable and efficient engineering of complex novel biological devices and systems.

Here we describe the construction and relative characterization of *Escherichia coli* (*E. coli*) σ^{70} transcriptional promoters by monitoring the synthesis of green fluorescent protein (GFP) both in vivo in *E. coli* and in vitro in a *E. coli* cell-free transcription and translation reaction.

Key words Synthetic biology, σ^{70} transcriptional promoters, DNA regulatory elements characterization, Relative promoter units, Cell-free transcription and translation reaction, green fluorescent protein and real-time PCR

1 Introduction

The use of characterized “biological parts” (BioParts) for engineering novel complex biological devices and systems is a key principle of synthetic biology [1, 2]. Collaborative and international efforts have been made to physically store and document collections of BioParts, including the BioBricks registry (parts.mit.edu) and BIOFAB collection (biofab.org). However, many BioParts within these registries have yet to be quantitatively characterized and the development of high-throughput characterization methods is still a major challenge of synthetic biology [3].

Traditional techniques for the characterization of DNA regulatory elements such as promoters and ribosome binding sites have used indirect measurements of strength by monitoring the expression of reporter genes such as fluorescent proteins, luciferases, and

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_7, © Springer Science+Business Media New York 2013

β-galactosidase [3–5]. However, a major drawback with these methods is the context dependence of such measurements limits the comparability of independently collected characterization data. Context variations include the exact measurement conditions, the make and model of instrumentation used and the reporter genes expressed [5]. A progression of this method as suggested by Kelly et al. has been the use of relative measurements, whereby a designated standard reference BioPart is measured in parallel to all characterizations and used to create a ratio of BioPart strength to that of a standard. This approach provides an internal control for any context variation and has been shown to reduce the context dependence of such measurement [5].

Here we describe the construction of a σ^{70} transcriptional promoter library and the standard reference promoter (BBa_J23101 from *parts.mit.edu*) and its characterization by expression of green fluorescent protein (GFP) in vivo in *E. coli* [5]. We describe the measurement of GFP synthesis and hence promoter strength by both real-time PCR to determine relative mRNA levels and by fluorescent measurements to determine the relative GFP levels.

In recent years, it has been shown that properties such as strengths and rate of abortive transcription to be comparable between in vivo and in vitro reactions [6–8]. The potential to perform characterization of BioParts in vitro could be advantageous for several reasons. Firstly, in vitro systems offer an ability to ascertain characteristics that can be difficult to study in vivo, for example abortive transcription rate [6]. Secondly, much of context and complexity of the cell is removed from the measurement for example in vitro there is no cell division, reduced or no cell metabolism, no environmental responses, and no DNA replication. Thirdly, in vitro techniques are amenable to high-throughput library generation and screening methods such as liquid handling robots and microfluidics [9]. Here we describe the use of an S12 cell-free *E. coli* transcription and translation reaction for the characterization of the σ^{70} transcriptional promoter library characterized in vivo.

2 Materials

Prepare all solutions using double distilled water (ddH_2O). Prepare and store all reagents at room temperature unless otherwise indicated. Follow all waste disposal regulations when disposing of culture and waste materials.

2.1 Oligonucleotide Library Construction

1. Design and order a pair of oligonucleotides for each member of the library of σ^{70} constitutive promoters to be tested. Design so that once annealed oligonucleotides will form a complementary double-stranded helix (Fig. 1). For data analysis, it is

	-35 Box -10 Box
σ^{70} Constitutive Promoter	TTGACANNNNNNNNNNNNNNNNNTATAATNNNNNN AACTGTNNNNNNNNNNNNNNNNNATATTANNNNNN
Reference Promoter	TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGC AAATGTCGATCGAGTCAGGATCCATAATACGATCG
Example Promoters	
BBa_J23101	TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGC
BBa_J23100	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC
BBa_J23106	TTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGC
BBa_J23150	TTTACGGCTAGCTCAGTCCTAGGTATTATGCTAGC
BBa_J23102	TTGACAGCTAGCTCAGTCCTAGGTACTGTGCTAGC
BBa_J23151	TTGATGGCTAGCTCAGTCCTAGGTACAATGCTAGC

Fig. 1 Sigma 70 constitutive promoter sequences with the −10 box (*Pribnow box*) and the −35 box highlighted in *blue*. The standard reference promoter BBa_J23101* sequence and the constitutive promoters from the "Anderson library" from the BioBricks registry*. *BioBrick registry (parts.mit.edu)

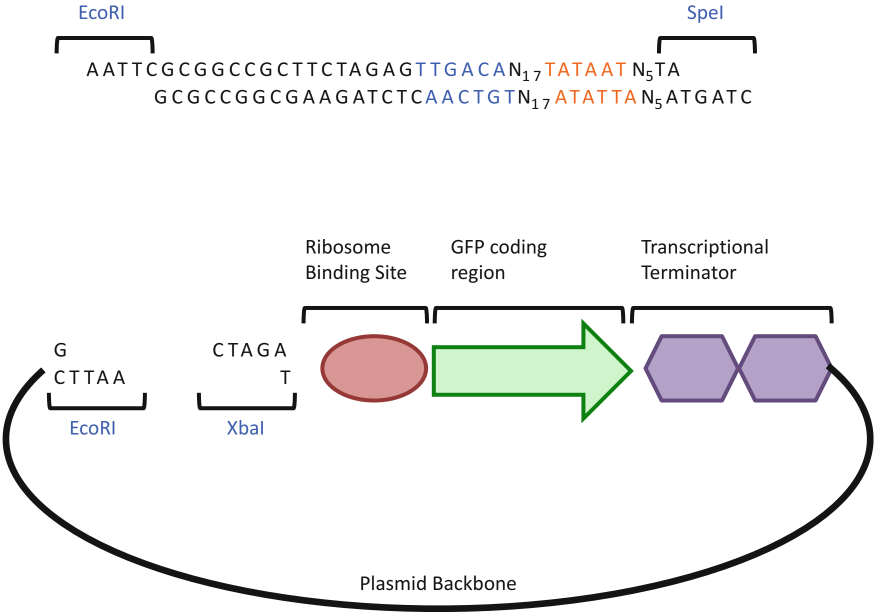

EcoRI
SpeI
AATTCGCGGCCGCTTCTAGAGTTGACAN_{17}TATAATN_5TA
GCGCCGGCGAAGATCTCAACTGTN_{17}ATATTAN_5ATGATC
Ribosome Binding Site
GFP coding region
Transcriptional Terminator
G
CTTAA
EcoRI
CTAGA
T
XbaI
Plasmid Backbone

Fig. 2 Sequence of annealed oligonucleotides of a constitutive promoter containing single-stranded overhangs complementary to that of the digested vector to be used, in this case the vector pSB1A2* containing the GFP open reading frame (BBa_I72504*) digested with EcoRI and XbaI. *BioBrick registry (parts.mit.edu)

important to include the reference promoter (BBa_J23101) as stated by Jason Kelly (Fig. 1) [5]. In addition, the oligonucleotides should be designed so that once annealed they contain single-stranded overhangs that are complementary to that of a digested GFP expression vector to be used (Fig. 2). The GFP expression vector used here is the BioBrick vector pSB1A2 containing BBa_I73504 available from the BioBricks Registry (*parts.mit.edu*), which was digested with the restriction enzymes EcoRI and XbaI (Fig. 2). In the example used here, the oligonucleotides contained single-stranded overhangs that resemble an EcoRI and SpeI digest and that once ligated into BBa_I73504 would adhere to the BioBrick registry standard cloning sites (*see* **Note 1**).

2. Annealing buffer (×10): 100 mM Tris–HCl pH 7.5, 1 M sodium chloride (NaCl), and 10 nM ethylenediaminetetraacetic acid (EDTA) (*see* **Note 2**).

2.2 In Vivo Culturing and Measuring Components

1. Minimal media: M9 salts (×1), 1 mM thiamine hydrochloride, 0.4 % glycerol, 0.2 % casamino acids, 2 mM magnesium sulfate, 0.1 mM calcium chloride, and a suitable antibiotic for the vector used. For the BBa_I73504, 100 μg/ml of ampicillin was used (*see* **Note 3**).
2. Cuvettes and spectrometer for optical density measurements at 600 nm.
3. Clear or black 96-well microplate transparent with clear base (*see* **Note 4**).
4. Plate reader for fluorescent measurements (excitation of 485 nm and emission of 520 nm) and optical density measurements at 600 nm.

2.3 mRNA Purification and Two-Step Real-Time PCR

1. Minimal media.
2. Cuvettes and spectrometer for optical density measurements at 600 nm.
3. RNeasy protect mini kit (Qiagen).
4. TE buffer: 10 mM Tris–HCl pH 8.0, 1 mM EDTA, containing 1 mg/ml lysozyme.
5. β-mercaptoethanol.
6. 96–100 % ethanol.
7. RNase-free H_2O.
8. RNase/DNase-free 1.5 ml tubes.
9. Quantitect reverse transcription kit (Qiagen).
10. QuantiFast SYBR green PCR kit (Qiagen).
11. Twin.tec PCR 96-well plate semi-skirted (Eppendorf).
12. Mastercycler Realplex 4 system (Eppendorf) or equivalent real-time instrumentation.
13. Reverse transcription primer mix: A 40 μM of the reverse transcription primers for 16s rRNA and GFP (Table 1).
14. Masterclear real-time PCR film (Eppendorf).

Table 1
Primers required for the reverse transcription and real-time PCR of GFP mRNA and 16s rRNA

Primer name	Sequence (5′–3′)	Product size (bp)
Reverse transcription primers		
16 s rRNA	TAAGGAGGTGATCCAAC	NA
GFPmut3b	TTATTATTTGTATAGTTCATC	NA
Real-time PCR primers		
16s rRNA Fwd	GCTACAATGGCGCATACAAA	101
16s rRNA Rev	TTCATGGAGTCGAGTTGCAG	
GFPmut3b Fwd	ATGGCCCTGTCCTTTTACC	85
GFPmut3b Rev	ATGTGGTCTCTCTTTTCGTTG	

2.4 Cell-Free Transcription and Translation Reaction Components

1. 2xYT: 16 g of bacto tryptone, 10 g Bacto Yeast Extract, 5 g NaCl, pH 7.0.
2. Buffer A: 10 mM Tris–acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium glutamate, and 1 mM dithiothreitol (DTT) with 0.05 % 2-mercaptoethanol.
3. Buffer B: 10 mM Tris–acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium glutamate.
4. Water soluble amino acids: To prepare from L-amino acid powder, prepare 500 mM of L-arginine, L-glycine, L-histidine, L-lysine, L-proline, L-serine, L-threonine, and L-valine in ddH$_2$O.
5. Acid soluble amino acids: To prepare from L-amino acid powder, prepare 500 mM of L-asparagine, L-aspartic acid, L-cystine, L-glutamine, L-glutamic acid, L-leucine, L-methionine, L-tryptophan, and L-tyrosine in 1 M hydrochloric acid.
6. Base soluble amino acids: To prepare from L-amino acid powder, prepare 500 mM of L-isoleucine and L-phenylalanine in 1 M potassium hydroxide.
7. Amino acid solution: Mix equal volumes of the water soluble, acid soluble, and base soluble amino acid solutions. Each amino acid should be in a concentration of 166 mM. Aliquot into suitable volumes and store at −80 °C.
8. Premix 1 (×5): 16.6 mM DTT, 5.3 mM cAMP, 283 μg/ml folinic acid, and 2,916 μg/ml of *E. coli* tRNA mixture (Sigma-Aldrich) and store at −80 °C.
9. Premix 2 (×3): 300 mM HEPES-KOH (pH 8.2), 400 mM ammonium acetate, 1 M potassium glutamate, and 15 mM sodium oxolate and store at −20 °C.
10. Premix 3: Premix 1 (×1), Premix 2 (×1), 2.4 mM rATP, 1.6 mM rCTP, 1.6 mM rUTP, 1.6 mM rGTP, 3.2 % (w/v) PEG-8000, 19.2 mM magnesium acetate, 3.2 mM of amino acid solution. Freshly prepare on same day of experiment.

11. For the cell-free transcription and translation reaction, a suitable DNA template is required. We have found that using minipreps or midi-preps of the *E. coli* cultures transformed with the promoter libraries to work best. A minimum concentration of 250 ng/μl is required to maintain overall volume of cell-free transcription and translation reaction at 50 μl (*see* **Note 5**).
12. Black 384-well microplate.

3 Methods

Carry out all procedures at room temperature unless otherwise specified. All enzymes should be stored on ice and ddH_2O used to prepare all solutions.

3.1 Preparation of Constitutive Promoter Library

1. If the purchased oligonucleotides are lyophilized, suspend to a stock concentration of 1 μg/μl and further to a working stock of 100 ng/μl in ddH_2O (*see* **Note 6**).
2. For each promoter of the library, prepare the annealed oligonucleotides by mixing 100 ng of each complementary oligonucleotides, annealing buffer (×1) and ddH_2O to 10 μl. Heat reaction to 95 °C for 5 min and cool to room temperature.
3. Prepare ligation by mixing 0.5 μl of annealed oligonucleotides, 25–50 ng of digested vector containing a GFP open-reading frame (Fig. 2) (*see* **Note 7**), 1 μl of T4 ligase (NEB), T4 ligase buffer (×1) (NEB), and ddH_2O upto 10 μl. Incubate for 2 h at room temperature.
4. Take ligation reaction and transform into *E. coli BL21 (DE3)* competent cells according to the manufactures protocol resulting in single colonies of transformed *E. coli* on an LB agar plate for each constitutive promoter (*see* **Note 8**). In addition, as a negative control transform *E. coli BL21 (DE3) E. coli* with undigested GFP expression vector.

3.2 In Vivo Characterization of Constitutive Promoter Library by Fluorescent Measurements

1. For each promoter of the library and the empty GFP expression vector, pick a single colony of transformed *E. coli* from the LB agar plate and inoculate 2 ml of M9 minimal media and grow for 16 h at 37 °C shaking at 255 rpm.
2. Re-dilute 50 μl of each *E. coli* culture into 5 ml of pre-warmed M9 minimal media and incubate for 3 h at 37 °C shaking at 225 rpm (*see* **Note 9**).
3. Remove 1 ml of each culture into a cuvette and measure the optical density at 600 nm (O.D.600) and prepare 1 ml of diluted culture to an O.D.600 of 0.07 in pre-warmed M9 minimal media (*see* **Note 10**).
4. Incubate the diluted cultures for 1 h at 37 °C shaking at 225 rpm.

5. For each member of the library, transfer 200 μl of *E. coli* culture in triplicate into a 96-well plate. For fluorescence and O.D.600 background measurements include 200 μl of *E. coli* culture containing empty GFP expression vector and 200 μl of M9 media.
6. Measure fluorescence (excitation at 485 nm, emission at 520 nm) and absorbance at 600 nm for 30 min at approximately 5 min intervals whilst incubating at 37 °C and shaking at 225 rpm.
7. For data analysis, remove the background fluorescence and absorbance from the empty vector cultures and M9 media, respectively. Average all triplicate measurements of each promoter for every time point. Calculate the relative promoter units (RPU) for each time point as below and plot the RPU against time:

$$\mathrm{RPU}(X) = \frac{\mathrm{Fl}_X/\mathrm{ABS}_X}{\mathrm{Fl}_{\mathrm{J23101}}/\mathrm{ABS}_{\mathrm{J23101}}}$$

where Fl_X and ABS_X is the fluorescence and O.D.600 of promoter tested and $\mathrm{Fl}_{\mathrm{J23101}}$ and $\mathrm{ABS}_{\mathrm{J23101}}$ is the fluorescence and O.D.600 of the reference promoter tested.

As the fluorescence and O.D.600 measurements are directly proportional to amount of GFP in culture and cell number, respectively, the graph should show the RPU to be at a steady-state across the time measured. Average the RPU from each time point to give the final value of RPU for each promoter.

3.3 In Vivo Characterization of Constitutive Promoter Library by Real-Time PCR

1. For each promoter of the library and the empty GFP expression vector pick a single colony of transformed *E. coli* from the LB agar plate and inoculate 2 ml of M9 minimal media and grow for 16 h at 37 °C shaking at 255 rpm (*see* **Note 11**).
2. Re-dilute 50 μl of each *E. coli* culture into 5 ml of pre-warmed M9 minimal media and incubate at 37 °C shaking at 225 rpm.
3. After 3 h remove 1 ml of each culture into a cuvette and measure the O.D.600. If culture density has reached an O. D.600 of 1 remove cultures from incubation, if below this O. D.600 continue to incubate and measure O.D.600 at regular intervals.
4. Once at an O.D.600 of 1 remove 600 μl of culture and add to 1.2 ml of RNA protect bacterial reagent, immediately vortex for 5 s and incubate at room temperature for 5 min (*see* **Note 12**).
5. Centrifuge for 10 min at 5,000 × *g* and carefully decant the supernatant from the bacterial pellet, ensuring residual supernatant is removed by gently dabbing the inverted tube on a paper towel.

6. Resuspend the bacterial pellet in 200 μl TE buffer containing 1 mg/ml of lysozyme and vortex for 10 s. Incubate at room-temperature for 5 min and every 2 min vortex for 10 s (*see* **Note 13**).
7. Add 10 μl β-mercaptoethanol per ml of RLT buffer before proceeding to next step (*see* **Note 14**).
8. Add 700 μl of buffer RLT and vortex for 10 s. Then add 500 μl of 96–100 % ethanol and mix by pipetting (*see* **Note 15**).
9. To purify the mRNA transfer 700 μl of the lysate to the RNeasy spin column placed in a 2 ml collection tube and centrifuge at 8,000 × *g* for 15 s and discard the flow-through. Repeat step for remaining lysate.
10. Wash the RNeasy spin column with 700 μl of buffer RW1 and centrifuge for 15 s at 8,000 × *g* and discard the flow-through.
11. Add 500 μl of Buffer RPE to the RNeasy spin column and centrifuge for 15 s at 8,000 × *g* and discard the flow-through. Repeat this wash.
12. Place RNeasy spin column in a new 2 ml collection tube and centrifuge 8,000 × *g* for 1 min.
13. To elute the mRNA place the RNeasy spin column in a RNase/DNase free 1.5 ml collection tube and elute with 50 μl of RNase-free H_2O and centrifuge for 1 min at 8,000 × *g*. If purified mRNA will not be used immediately it can be stored at −20 °C or at −80 °C for long-term storage.
14. Prepare the genomic DNA elimination reaction to remove any DNA from the purified mRNA. For each mRNA sample add 2 μl of gDNA Wipeout Buffer (7×), 1 μl of mRNA and 11 μl of RNase-free water. Incubate the reaction for 2 min at 42 °C and immediately place on ice.
15. Prepare the reverse transcription reaction on ice. To the 14 μl of genome wipeout reaction add: 1 μl of Quantiscript Reverse Transcriptase, 4 μl Quantiscript RT Buffer (5×) and 1 μl of the reverse transcription primers mixture. Incubate for 15 min at 42 °C, followed by incubation at 95 °C for 3 min to cease reaction. The final cDNA can be stored at −20 °C.
16. To quantify the 16s rRNA and GFP cDNA for each member of the promoter library real-time PCR is performed. Two separate master mixes for both 16s rRNA and GFP mRNA are prepared with their respective qPCR primer sets (Table 1). The composition of a reaction is shown in (Table 2) (*see* **Note 16**). The number of reactions needed for each master mix is three for each cDNA sample to allow triplicate measurements, three for a non-template control to check for contamination and three for pipetting buffer. Once master-mix is prepared it can be transferred into a twin.tec® PCR 96-well plate semi-skirted and 1 μl

Table 2
Composition of the real-time PCR reaction using the QuantiFast SYBR green PCR kit (Qiagen)

		Volume per reaction (μl)	
Reaction component	**Stock**	**10 μl reaction**	**25 μl reaction**
Fwd primer	100 μM	0.03	0.75
Rev primer	100 μM	0.03	0.75
SYBR green MM	2×	5	12.5
H_2O		3.94	9.85
Total volume		9	24

of cDNA sample for each reaction added. For the non-template control 1 μl of RNase/DNase free H_2O (*see* **Note 17**).

17. Seal the plate with a masterclear real time PCR film (*see* **Note 18**).
18. Centrifuge plate at 500 × *g* for 1 min (*see* **Note 19**).
19. The following real-time PCR program should be run: 95 °C for 5 min, 30× (95 °C for 15 s, 50.5 °C for 15 s and 72 °C for 20 s). In addition a melting curve should be run on the following program: 95 °C for 15 s, 60 °C for 15 s, 10 min gradient to 95 °C and 95 °C for 15 s.
20. Once complete, average the Ct value of the GFP and the 16 s rRNA for every member of the promoter library and the empty GFP expression vector. It is important to check the non-template controls for both 16 s rRNA and GFP reactions to determine that no contamination was present (*see* **Note 20**).
21. To analyze the real-time PCR data the $2^{-\Delta\Delta \mathrm{Ct}}$ method is used [10]. For every member of the promoter library and the empty GFP expression vector the amount of GFP mRNA added to the reverse transcription is normalized using the 16s rRNA as an internal control. Where the $\mathrm{Ct}_{\mathrm{GFP}x}$ and $\mathrm{Ct}_{\mathrm{ref}x}$ represent the Ct values of $\mathrm{Promoter}_X$ for the GFP and 16s rRNA, respectively (*see* **Note 21**).

$$\mathrm{Ct}_{\mathrm{GFP}x} - \mathrm{Ct}_{\mathrm{ref}x} = \Delta \mathrm{Ct}_x$$

The relative levels of GFP mRNA are then determined for each member of the promoter library by comparing to a calibrator. The calibrator used here is the empty GFP expression vector as this represents the basal expression of GFP in pSB1A2 vector. Each member of the promoter library is processed as a relative fold change from this value, where the $\Delta \mathrm{Ct}_{\mathrm{cal}}$ is the normalized

ΔCt of the empty GFP expression vector and N_x is the relative level of GFP mRNA for each member of the promoter library.

$$\Delta \mathrm{Ct}_x - \Delta \mathrm{Ct}_{\mathrm{cal}} = \Delta\Delta \mathrm{Ct}_x$$

$$2^{-\Delta\Delta \mathrm{Ct}_x} = N_x$$

Finally, the RPU can be calculated by comparing the relative GFP mRNA levels (N) of each promoter to that of J23101.

$$\mathrm{RPU} = \frac{N_x}{N_{\mathrm{J23101}}}$$

3.4 S12 Cell-Free Production and Reaction

1. Inoculate 20 ml of 2xYT media with a single colony of *E. coli Rosetta* strain and incubate for 16 h at 37 °C and shaking at 225 rpm (*see* **Note 22**).
2. Transfer the 20 ml of culture into 1 l of 2-YT media and grow at 37 °C shaking at 225 rpm until mid-log phase is reached (*see* **Note 23**).
3. Centrifuge at 4,000 × *g* for 15 min at 4 °C. Wash the cell-pellet in buffer A using 20 ml per gram of wet cell mass and centrifuge at 4,000 × *g* for 15 min at 4 °C. Repeat the wash step three times.
4. Resuspend washed pellet in 1.27 ml of buffer B per gram of wet cell mass. Lyse the suspension with a French press using a constant pressure of 20,000 psi (*see* **Note 24**).
5. Centrifuge this lysate at 12,000 × *g* for 10 min and recover the supernatant. Incubate at 37 °C for 30 min, aliquot and store at −80 °C.

3.5 In Vitro Characterization of Constitutive Promoter Library

1. Prepare premix solution 3 as described in methods and defrost the S12 cell-free extract and the promoter DNA library templates on ice.
2. Create a master mix of 30 μl of Premix 3 and 15 μl of S12 cell-free extract per promoter to be tested. Remember to include two additional reactions, one for a negative control of empty GFP vector and one to buffer any pipetting error.
3. For each promoter to be tested add 1 μg of the DNA template to 45 μl of the master mix and ddH_2O upto 50 μl. For fluorescence background measurements add 1 μg of empty GFP vector to a reaction for a negative control.
4. Quickly pipette each 50 μl of solution into a black 384 well microplate being careful to avoid bubbles.
5. Measure fluorescence (excitation at 485 nm, emission at 520 nm) at approximately 20–30 min intervals for 5 h whilst incubating at 30 °C (*see* **Note 25**).

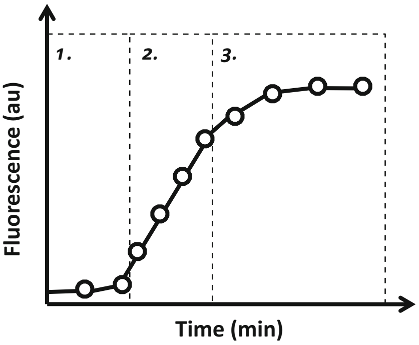


Fig. 3 A typical expression profile from an in vitro cell-free transcription and translation reaction with a constitutive expression of GFP. The graph is composed of three phases; 1. lag phase, 2. exponential phase, and 3. linear phase when energy and resources are limiting

6. For data analysis, remove the background fluorescence of the empty vector reaction for each time point. Plot the fluorescence against time to display the expression profile. It should show a sigmoidal curve with an initial lag phase, exponential phase of expression and a linear phase as energy and resources become limiting (Fig. 3). We use the exponential phase to calculate promoter strength as rate of transcription from the promoter strength is maximal. Take 2–3 time points in the exponential phase and calculate a rate of fluorescence increase:

$$\mathrm{Rate}(X) = \mathrm{Fl}_X t_3 - \mathrm{Fl}_X t_1$$

To convert this to a relative promoter units (RPU) divide the rate of transcription of the promoter tested by the rate of transcription of the J23101 reference promoter:

$$\mathrm{RPU}(X) = \frac{\mathrm{Rate}_X}{\mathrm{Rate}_{\mathrm{J23101}}}$$

4 Notes

1. The BioBricks registry is an open source collection of genetic BioParts known as BioBricks. The majority of BioBricks within the BioBrick registry have been submitted to adhere to the BioBricks assembly standard as described in the request for comments (RFC) 10 (*see* parts.mit.edu or biobricks.org for details).

2. In our experience instead of preparing annealing buffer, commercial restriction digestion buffers such as New England Biolabs (NEB) Buffer 2 (10 mM Tris-HCl pH 7.9, 50 mM sodium chloride, 1 mM DTT, 10 mM magnesium chloride) can be

used as a substitute. The variation in salt composition and the absence of EDTA does not seem to effect efficiency of later steps.

3. For preparing the M9 minimal media, it is important to prepare the thiamine hydrochloride freshly each time and sterilize using a 0.22 μm filter. The remaining solutions can be autoclave and stored at 4 °C. All components and a suitable antibiotic can be mixed together and stored at 4 °C. It is important to store M9 minimal media in the dark or to cover in foil because of the sensitivity of thiamine hydrochloride to light.
4. We use black 96-well microplate with clear base because the black plate minimizes background on fluorescent measurements whilst the clear base permits absorbance measurements. However, from experience the high level of fluorescence produced by the GFP expressing cultures negates any background from all clear plates.
5. For the final elution of mini- or midi-preps, we use ddH_2O to prevent excess salts within the S12 cell-free transcription and translation reaction as this can have an inhibitory effect. In addition, if DNA precipitation is required for concentration use sodium acetate and not ammonium acetate.
6. To ensure full suspension of oligonucleotides, apply ddH_2O and leave at room temperature for 30 min, vortexing every 10 min.
7. If non-phosphorylated oligonucleotides are used for the ligation, then do not dephosphorylate the vector. To minimize intra-ligation of vector gel-purify digested product to ensure only double digested plasmid is purified and used for ligation.
8. In the original paper by Kelly et al., TOP10-DH5α were used. From our experience BL21 (DE3) are also a suitable strain.
9. This first dilution of overnight culture is important to allow dilution of accumulated GFP by division and secondly to return cells to the log phase of growth.
10. In our experience *E. coli* expressing GFP under different strength promoters can have different growth rates. The second dilution is to ensure all *E. coli* cultures of the promoter library are at the same O.D.600 and that following the 1 h incubation will be in logarithmic growth phase. To calculate the volume of *E. coli* culture required for dilution in 1 ml to an O.D.600 0.07, use the following calculation:

$$\text{Volume} = \frac{0.07}{\text{O.D.}600_{\text{Culture}}} 1\ \text{ml}$$

11. If cultures are available from Subheading 3.2, then allow to grow to an O.D.600 of 1 and skip to **step 4**.

12. 1 ml of an *E. coli* culture at O.D.600 of 1 is a suitable number of cells to use for mRNA purification. It is important to use the correct number of cells, too few will result in low mRNA yields and too many will saturate the RNA binding capacity of the RNeasy spin columns which is 100 μg RNA.
13. Alternative to the lysozyme lysis method described here are compatible with the RNeasy Mini Kit (Qiagen), please see manufactures protocol for more details.
14. Once the β-mercaptoethanol is added to the RLT this should be stable for 1 month at 4 °C.
15. If any particulate material visible pellet this by centrifugation and ensure that only the supernatant is used and particulate may block column.
16. When using the QuantiFast SYBR Green PCR Kit (Qiagen), it is recommended that each reaction is 25 μl; however, in our experience we have reduced reaction volume to 10 μl without loss of efficiency.
17. It is important to carefully plan the plate layout before to help avoid pipetting errors. Generally, triplicates are placed together, if possible spaces are placed between triplicates and non-template controls placed away from cDNA containing samples, usually at the bottom or top of a plate, to avoid spilling over the wells. In our experience the 1 μl of cDNA can be added on the side of the wells to help keep track of where template has been added.
18. Be careful not to make any marks on the seal as this may interfere with the measurements.
19. The centrifugation step ensures that all reactions are mixed and collected at the bottom of the plate wells. This step is essential if cDNA is added onto the side of the 96 well-plate. After centrifugation check the wells for bubbles as this might disrupt the quantification.
20. Typically from a melting-curve analysis each amplicon amplified will have a unique dissociation peak and so as each sample only has one amplicon then a single peak should be present. However, it is possible that primer dimers will form, causing an increase in florescence during the real-time PCR and a unique dissociation peak in the melting curve analysis usually at 70 °C. If primer dimers are present, these will be present in the non-template control also. It is important to check the GFP or the 16 s rRNA amplicon are not present in the non-template control as this would indicate contaminations.
21. There are a variety of different data analysis methods that can be used. The delta delta CT method explained here is the simplest but only holds true if PCR efficiencies between the target and the reference gene are close to 100 % and not differ

by more than 10 %. PCR efficiencies are determined by preparing a standard curve [9, 10]. For the conditions and primers described here, the delta delta CT method is valid.

22. The growth media 2xYT is used because of the high densities of *E. coli* culture it can support. Typically in a 1 l flask an O.D.600 of 9 can be reached with mid-log phase of growth at O.D.600 of 4.5. For the cell-free extract preparation, the S12 method was used because of its time and cost-effectiveness when compared to the S30 method. The strain of *E. coli* used was Rossetta 2 cells (Invitrogen) as it has been previously shown that this produces a highly productive and less variable S12 cell-free extract [11].

23. From experience the mid-log phase of *E. coli* Rossetta strain in 2xYT media is typically at an O.D.600 of 1.5.

24. From experience it is best to run the cell-free suspension three times through the French press always maintaining the lysate on ice.

25. Depending on the specifications of the plate reader used, it might be required to use a transparent lid to prevent evaporation from the 384 micro well plate.

References

1. Endy D (2005) Foundations for engineering biology. Nature 438:449–453
2. Arkin A (2008) Setting the standard in synthetic biology. Nat Biotechnol 26:771–774
3. Martin L, Che A, Endy D (2009) Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. PLoS One 4:e7569
4. Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol 26:787–793
5. Kelly JR, Rubin AJ et al (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng 3:4
6. Goldman SR, Ebright RH, Nickels BE (2009) Direct detection of abortive RNA transcripts in vivo. Science 324:927–928
7. Patwardhan RP, Lee C et al (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol 27:1173–1175
8. Chappell J, Jensen K, Freemont PS (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. Nucleic Acids Res 41: 3471–3483
9. Gulati S, Rouilly V et al (2009) Opportunities for microfluidic technologies in synthetic biology. J R Soc Interface 6(Suppl 4): S493–S506
10. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3:1101–1108
11. Kim TW, Keum JW et al (2006) Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. J Biotechnol 126:554–561

Chapter 8

In Vivo Screening of Artificial Small RNAs for Silencing Endogenous Genes in *Escherichia coli*

Vandana Sharma and Yohei Yokobayashi

Abstract

Bacterial noncoding small RNAs (sRNAs) modulate expression of numerous genes through antisense interactions with mRNAs. This chapter describes an in vivo screening strategy to engineer artificial sRNAs that can posttranscriptionally regulate desired endogenous genes in *Escherichia coli*. Artificial sRNA libraries are constructed by randomizing the antisense domain of natural sRNAs and screened for gene silencing activity using a cotransformed reporter vector. These small synthetic riboregulators can be used in synthetic gene circuits to control cell functions by directly targeting endogenous genes.

Key words Noncoding RNA, Small RNA, Gene regulation, Gene silencing, RNA engineering, Translational regulation

1 Introduction

For synthetic biological circuits to be useful, they must produce one or more "outputs" that generate detectable signals or control cellular behavior. Most circuits express reporter genes (e.g., GFP or luciferase) or other exogenous proteins as circuit outputs. Alternatively, cellular behavior can be modulated by silencing an endogenous gene. Metabolic engineering may also benefit from dynamic modulation of endogenous gene expression levels to optimize metabolic flux. Bacterial noncoding small RNAs (sRNAs) present an attractive platform on which to design synthetic riboregulators capable of controlling endogenous gene expression.

First discovered in 1984 [1], bacterial sRNAs typically repress gene expression in *trans* by hybridizing with the targeted mRNAs in the 5′ untranslated region (UTR) and/or translation initiation region [2]. Most sRNAs contain a well-defined antisense domain and an auxiliary domain that is believed to interact with cellular components (e.g., Hfq RNA chaperone) or enhance sRNA stability. The antisense domain can be as small as six bases or as large as several dozen bases, and often form imperfect base-pairing with

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_8, © Springer Science+Business Media New York 2013

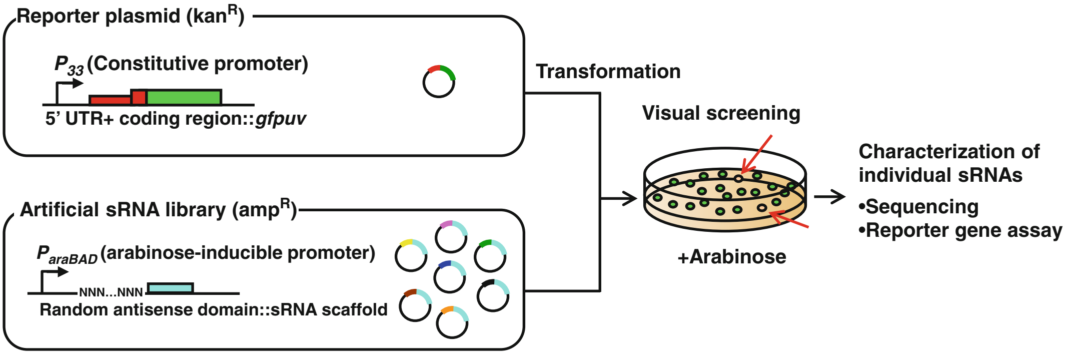

Reporter plasmid (kan^R)
P_{33} (Constitutive promoter)
5' UTR+ coding region::*gfpuv*
Artificial sRNA library (amp^R)
P_{araBAD} (arabinose-inducible promoter)
NNN...NNN
Random antisense domain::sRNA scaffold
Transformation
Visual screening
+Arabinose
Characterization of individual sRNAs
•Sequencing
•Reporter gene assay

Fig. 1 Artificial sRNA screening strategy. Reporter plasmid expresses GFPuv fused to the 5′ leader sequence from the targeted mRNA from a synthetic constitutive promoter. Artificial sRNA library expresses sRNAs with a randomized antisense domain fused to an sRNA scaffold. The two plasmids are sequentially transformed into a suitable *E. coli* strain. Active artificial sRNAs are visually screened by isolating colonies that display weak fluorescence. The artificial sRNA clones are further characterized individually. Reproduced from [6] with permission from American Chemical Society

the targeted mRNA [3–5]. We sought to exploit the modular architecture of bacterial sRNAs by randomizing and screening for active antisense domains while conserving the auxiliary domains (sRNA scaffolds) to engineer artificial sRNAs capable of regulating desired endogenous genes.

The overall strategy is illustrated in Fig. 1. First, the 5′ leader sequence (5′ UTR and early protein-coding sequence) of the targeted mRNA is fused to the coding sequence of a reporter gene (GFPuv) to construct a reporter plasmid. Second, artificial sRNA libraries are constructed by randomizing the antisense domain of various natural sRNAs. Third, the reporter plasmid and the sRNA libraries are cotransformed into an *E. coli* strain and screened for active sRNAs. Finally, the active sRNA clones are individually characterized. Synthetic gene circuits can incorporate the artificial sRNAs obtained by the method outlined here as circuit outputs to directly interface with the endogenous metabolic pathways or gene networks.

2 Materials

2.1 Cell Culture Media

1. LB: 1 % bacto tryptone, 0.5 % yeast extract, 1 % NaCl.
2. Ampicillin and kanamycin were used at the final concentrations of 100 and 50 μg/mL, respectively, where appropriate.
3. Arabinose was added at 0.1 % w/v where necessary.
4. Agar plates were prepared with 1.5 % w/v additional agar.

2.2 Equipment

1. UV Transilluminator (320 and 360 nm).
2. Thermal cycler.

3. Incubator.
4. Incubator-shaker.
5. Microcentrifuge (plasmid miniprep and DNA purification).
6. Agarose gel electrophoresis apparatus (chamber and power supply).
7. Microplate reader (cell fluorescence measurement).
8. UV spectrophotometer (for DNA quantification).

2.3 Plasmids and Primers

1. pKP33-Esp3I-GFPuv (*see* **Note 1**) or a similar host reporter plasmid.
2. pBAD-XH or a similar artificial sRNA expression plasmid (*see* **Note 2**).
3. Sequencing primers for the above plasmids.
4. PCR primers (sequence design considerations are given in Methods).

2.4 Reagents and Cells

1. Phusion DNA polymerase (New England Biolabs).
2. dNTPs.
3. Restriction enzymes (Esp3I is available from Fermentas).
4. Antarctic Phosphatase (New England Biolabs) or other alkaline phosphatase.
5. Agarose gel electrophoresis reagents.
6. DNA Clean & Concentrator-5 (Zymo) or other silica column DNA purification kit.
7. Zyppy Plasmid Miniprep Kit (Zymo) or other plasmid miniprep kit.
8. Quick Ligation Kit (New England Biolabs) or other T4 DNA ligase.
9. TOP10 *E. coli* competent cells (Invitrogen) (*see* **Note 3**).
10. Z-Competent *E. coli* Transformation Kit (Zymo)

3 Methods

3.1 Design and Construction of Reporter Plasmid

1. In principle, any reporter plasmid that is compatible with the plasmid used to construct the artificial sRNA library (i.e., replication origin and antibiotic resistance) could be used. In addition to GFPuv described here, other reporter genes such as β-galactosidase (LacZ) whose expression can be visualized on colonies should also work. Note that some optimizations such as promoter strength or plasmid copy number (i.e., replication origin) may be necessary depending on the target mRNA. The procedure described here uses pKP33-Esp3I-GFPuv

(kan^R, p15a ori) as an example. The plasmid is available from the authors upon request.

2. Obtain the target mRNA sequence through literature search or databases such as EcoCyc (http://ecocyc.org/). Identify the 5′ UTR and the first 10–20 codons of the coding sequence (5′ leader sequence) (*see* **Note 4**).
3. Amplify the 5′ leader sequence by PCR from genomic DNA with added Esp3I restriction sites for cloning (Fig. 2, *see* **Note 5**). Alternatively, synthesize the 5′ leader sequence from synthetic oligo DNAs or through commercial gene synthesis service.
4. Digest the PCR product with Esp3I following the manufacturer's instructions.
5. Digest pKP33-Esp3I-GFPuv with Esp3I. Dephosphorylate the digested plasmid with Antarctic Phosphatase following the manufacturer's instructions.
6. Purify the PCR product and the digested plasmid by column purification (e.g., DNA Clean & Concentrator-5).
7. Ligate the purified DNA fragments using T4 DNA ligase following the manufacturer's instructions.
8. Transform the ligation solution into competent *E. coli* cells and spread on a LB-kanamycin plate.
9. Miniprep the plasmids from several clones. Confirm the sequence and retransform the plasmid into a strain to be used for screening. The colonies should be visibly and uniformly fluorescent (when GFPuv is used) when placed over an UV transilluminator (360 nm) (*see* **Note 6**).
10. Pick one of the freshly transformed colonies and prepare competent cells using Z-Competent *E. Coli* Transformation Kit, following the kit instructions. Prepare at least 10 tubes, each containing 50 μL cells and store at −80 °C.

3.2 Design and Construction of Artificial sRNA library

1. We have previously designed artificial sRNAs based on Spot42, DsrA, MicF, and GcvB scaffolds [6]. All of these libraries contain a randomized antisense domain (20 or 30 bases) in the 5′ end followed by a 3′ putative sRNA scaffold. These scaffolds were identified based on biochemical and genetic characterizations that have been reported in the literature. We believe 10–20 randomized bases to be sufficient for the antisense domain, but other sRNA library configurations are certainly possible.
2. Clone the sRNA scaffold sequence into an expression plasmid that is compatible with the reporter plasmid constructed above. We use arabinose-inducible pBAD-XH which was derived from pBAD-His2B (Invitrogen) [6]. pBAD-XH contains XhoI and HindIII sites for cloning an sRNA scaffold sequence.

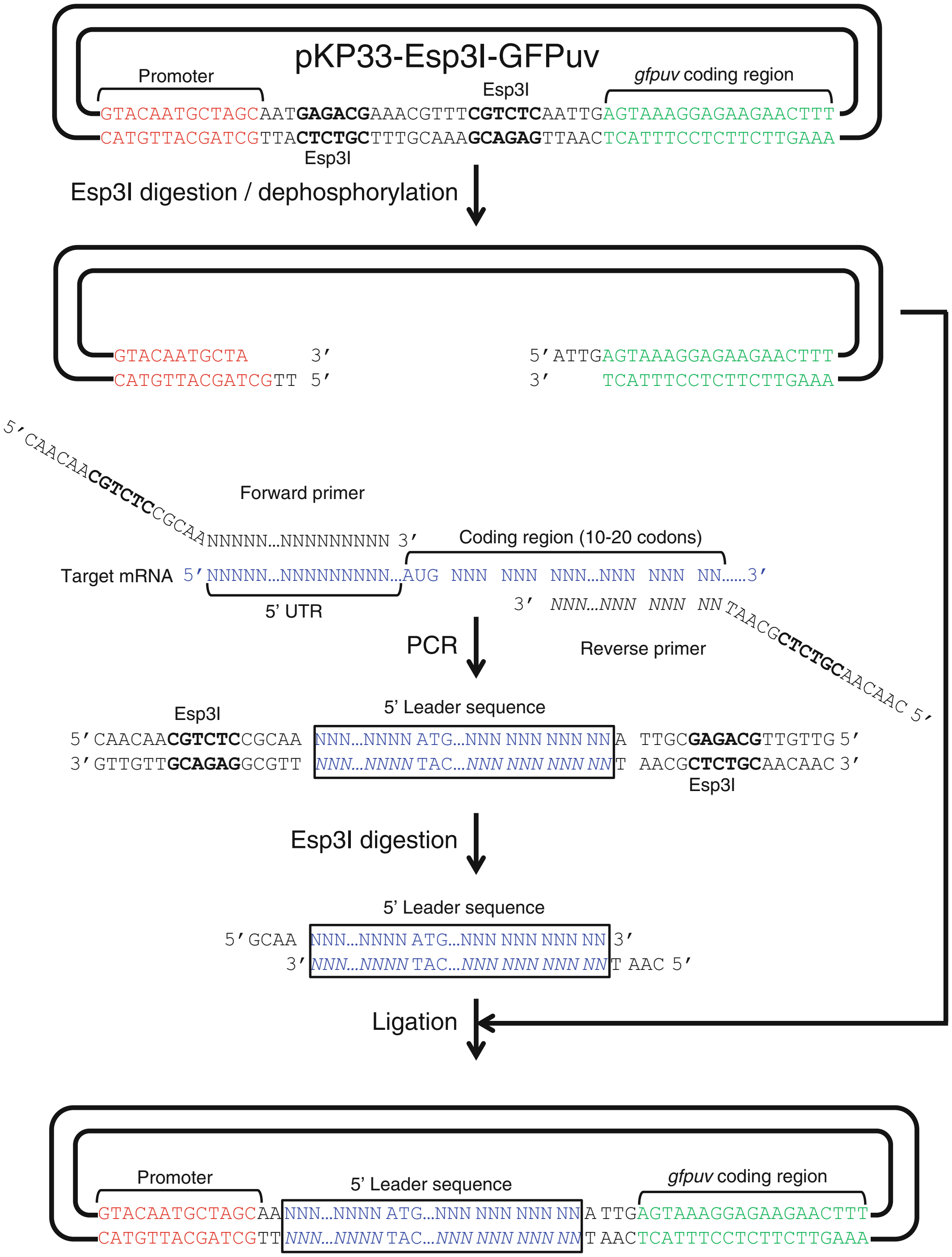

pKP33-Esp3I-GFPuv
Promoter
Esp3I
gfpuv coding region
GTACAATGCTAGCAATGAGACGAAACGTTTCGTCTCAATTGAGTAAAGGAGAAGAACTTT
CATGTTACGATCGTTACTCTGCTTTGCAAAGCAGAGTTAACTCATTTCCTCTTCTTGAAA
Esp3I
Esp3I digestion / dephosphorylation
GTACAATGCTA 3′
CATGTTACGATCGTT 5′
5′ATTGAGTAAAGGAGAAGAACTTT
3′ TCATTTCCTCTTCTTGAAA
5′CAACAACGTCTCCGCAANNNNNN...NNNNNNNNN 3′
Forward primer
Coding region (10-20 codons)
Target mRNA 5′NNNNN...NNNNNNNNN...AUG NNN NNN NNN...NNN NNN NN......3′
5′ UTR
3′ NNN...NNN NNN NNTAACGCTCTGCAACAAC 5′
PCR
Reverse primer
Esp3I
5′ Leader sequence
5′CAACAACGTCTCCGCAA NNN...NNNN ATG...NNN NNN NNN NN A TTGCGAGACGTTGTTG 5′
3′GTTGTTGCAGAGGCGTT NNN...NNNN TAC...NNN NNN NNN NN T AACGCTCTGCAACAAC 3′
Esp3I
Esp3I digestion
5′ Leader sequence
5′GCAA NNN...NNNN ATG...NNN NNN NNN NN 3′
3′ NNN...NNNN TAC...NNN NNN NNN NN T AAC 5′
Ligation
Promoter
5′ Leader sequence
gfpuv coding region
GTACAATGCTAGCAA NNN...NNNN ATG...NNN NNN NNN NN A TTGAGTAAAGGAGAAGAACTTT
CATGTTACGATCGTT NNN...NNNN TAC...NNN NNN NNN NN T AACTCATTTCCTCTTCTTGAAA

Fig. 2 Cloning strategy for the reporter plasmid. A type IIS restriction enzyme (Esp3I) was used to allow cloning of a 5′ leader sequence with minimal nonnative sequence. Only partial sequences of the synthetic promoter and the *gfpuv* coding region are shown

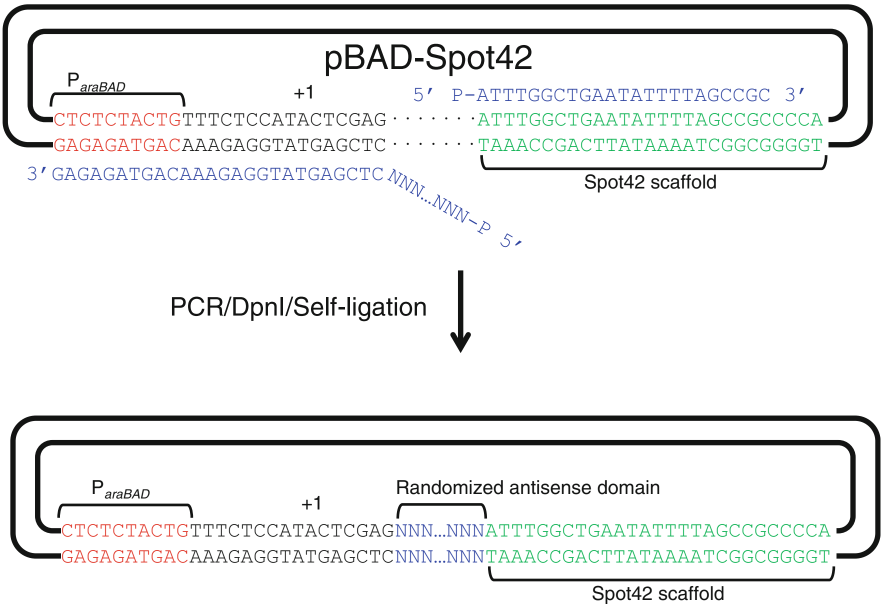

pBAD-Spot42
P araBAD
+1
5′ P-ATTTGGCTGAATATTTTAGCCGC 3′
CTCTCTACTGTTTCTCCATACTCGAG·······ATTTGGCTGAATATTTTAGCCGCCCCA
GAGAGATGACAAAGAGGTATGAGCTC·······TAAACCGACTTATAAAATCGGCGGGGT
3′ GAGAGATGACAAAGAGGTATGAGCTC NNN...NNN-P 5′
Spot42 scaffold
PCR/DpnI/Self-ligation
P araBAD
+1
Randomized antisense domain
CTCTCTACTGTTTCTCCATACTCGAGNNN...NNNATTTGGCTGAATATTTTAGCCGCCCCA
GAGAGATGACAAAGAGGTATGAGCTCNNN...NNNTAAACCGACTTATAAAATCGGCGGGGT
Spot42 scaffold

Fig. 3 Artificial sRNA library construction. PCR is used to insert a randomized antisense domain to a plasmid containing an sRNA scaffold. Only partial sequences of the P_{araBAD} promoter and the Spot42 scaffold are shown

3. Design and order a pair of partially degenerate primers to insert a randomized antisense domain by PCR. For a 3′ scaffold (e.g., Spot42), one primer should hybridize with the 3′ end of the promoter in the reverse direction while the other primer should hybridize with the 5′ end of the scaffold in the forward direction (*see* Fig. 3). One or both of the primers should contain consecutive degenerate bases (N) for the randomized antisense domain. The 5′ ends of the primers should be phosphorylated during synthesis.
4. Amplify the entire artificial sRNA expression plasmid by PCR using the above primers and Phusion DNA polymerase following the manufacturer's instructions (25 μL total volume).
5. Confirm the PCR product by agarose gel electrophoresis.
6. Add 1 μL (10 units) DpnI directly to the PCR solution. Mix thoroughly by pipetting and incubate at 37 °C for 90 min to digest the template plasmid.
7. Purify the PCR product by column purification (e.g., DNA Clean & Concentrator-5). Quantify the DNA concentration by UV spectroscopy. *See* **Note 7** for **steps 8–11**.
8. Dilute 10 ng of the above PCR product in 3 μL of deionized water and mix with 3 μL of 2× Quick Ligase Buffer. Add 0.3 μL of Quick Ligase, mix well, and incubate at 25 °C for 5 min. Scale up or down as appropriate.

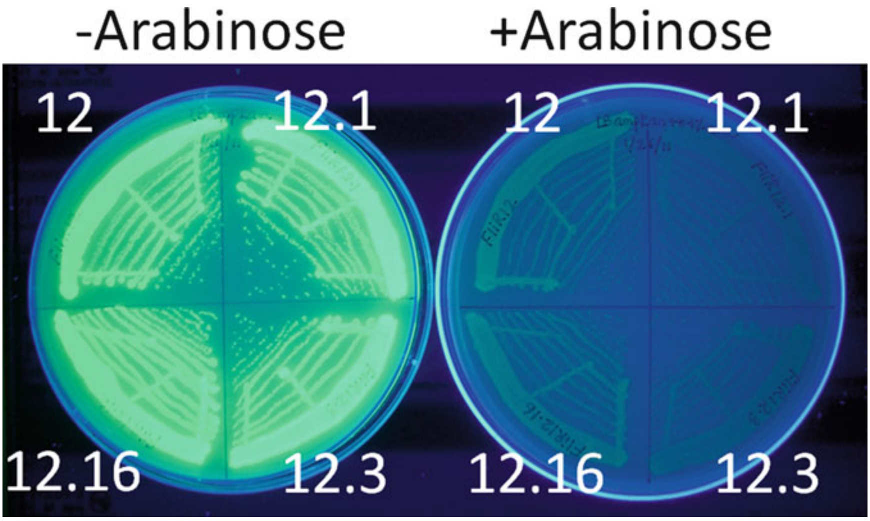


Fig. 4 Example of secondary screening of sRNA clones. Promising artificial sRNA clones (anti-*fliC* shown) were streaked on agar plates without and with arabinose, and imaged over a UV transilluminator (360 nm). Reproduced from [6] with permission from American Chemical Society

9. Transform the ligation mixture into appropriate high-efficiency *E. coli* competent cells. Plate ~10 μL cells on a LB-ampicillin plate to estimate the transformation efficiency. Culture the rest of the transformed cells in LB medium overnight at 37 °C.
10. Isolate the artificial sRNA library plasmid using a miniprep kit.
11. Count the number of colonies on the LB-ampicillin plate to calculate the library size. A library size of at least 10^4 is recommended.

3.3 Colony-Based Screening

1. Transform the artificial sRNA library plasmid obtained in Subheading 3.2 (**step 10**) into the competent cells harboring the reporter plasmid (Subheading 3.1, **step 10**) and plate the cells on LB-ampicillin-kanamycin agar plates supplemented with 0.1 % arabinose (*see* **Note 8**). As a negative control (no sRNA activity), transform an empty sRNA expression vector (e.g., pBAD-XH). This step should be done in a small scale first to optimize the transformation efficiency and the colony density on each plate. Each plate should contain no more than 800–1,000 colonies for ease of screening.
2. Observe each plate over a UV transilluminator (360 nm). Using the negative control (empty vector) plate as a guide, pick colonies with diminished fluorescence using sterile toothpicks and inoculate them into 500 μL LB medium in separate tubes. Wear appropriate UV protection and minimize the exposure of the cells to UV.
3. Perform secondary screening by streaking the isolated colonies on LB plates with and without arabinose (Fig. 4). Incubate the

plates overnight at 37 °C. Streak the negative control cells (empty sRNA expression plasmid) as well.

4. Observe the streaked cells over a UV transilluminator (360 nm). Confirm that GFPuv fluorescence is diminished on the arabinose-containing plates compared to arabinose-lacking plates. If fluorescence is diminished on both plates, it is possible that the cells contain a mutation in the reporter plasmid. Also look for any variations in the colony fluorescence within the streaked clones. If some cells are more or less fluorescent than others, it is possible that more than one colony was picked during screening. Repeat streak assay to confirm sRNA activity and to isolate single clones.
5. If desired, the clones can be used to start liquid cultures to quantify cellular fluorescence using a microplate reader.

3.4 Characterization of Artificial sRNAs

1. Inoculate a colony from the above screening into 1 mL LB medium and culture the cells overnight.
2. Isolate the plasmid using a miniprep kit following the manufacturer's instructions.
3. Selectively digest the reporter plasmid using an appropriate restriction enzyme (e.g., NheI).
4. Retransform the solution from **step 3** above into competent cells.
5. Isolate the artificial sRNA expression plasmids using a miniprep kit.
6. Retransform the artificial sRNA plasmids into the competent cells harboring the reporter plasmid (Subheading 3.1, **step 10**). The transformed cells should be characterized as described above (Subheading 3.3, **steps 2–5**) to rule out any mutations in the reporter plasmid or the host cell.
7. Analyze the sequences of the artificial sRNA plasmid clones. For example, the isolated antisense domains can be manually aligned for possible antisense interactions against the target mRNA leader sequence. Alternatively, software such as IntaRNA [7] can be used to suggest plausible antisense interactions (*see* **Note 9**).
8. Other arbitrary reporter plasmids containing endogenous or synthetic 5′ leader sequences may be used to examine the specificity of the artificial sRNAs. In theory, it is possible for an artificial sRNA to nonspecifically downregulate gene expression by affecting a major cellular factor. Such a possibility may be ruled out by testing multiple reporter constructs. Additionally, the antisense model obtained in **step 7** above can be tested by targeted mutations that disrupt the predicted sRNA–mRNA hybridization, as well as compensating mutations to restore the hybridization.

9. If the goal is to control an endogenous gene, the sRNA function should be tested by detecting the endogenous target by Western blot or other assays. It may be useful or necessary to subclone the artificial sRNAs into other expression vectors to optimize the sRNA expression level relative to the endogenous gene expression.

4 Notes

1. This plasmid is available from the authors upon request.
2. pBAD-XH and its derivatives in which natural sRNAs are cloned [6] (pBAD-Spot42, pBAD-GcvB, pBAD-DsrA, pBAD-MicF) are available from the authors upon request.
3. Other host strains may be used, but some choices of the plasmids may require specific host genotypes to be used. For example, the pBAD vector as described here requires the host strain to be compatible with arabinose induction.
4. In principle, 5′ UTR or intergenic region of an operon may be cloned and targeted as described here. However, how artificial sRNAs affect the expression of multiple genes within an operon has not been fully investigated.
5. Make sure to design the reverse primer with a correct 5′ appendage so that the coding sequence will be cloned in-frame with *gfpuv* in pKP33-Esp3I-GFPuv. If the 5′ leader sequence contains an Esp3I recognition site (5′-CGTCTC-3′), then other type IIS endonucleases such as BsaI may be used.
6. If the colony fluorescence is too weak, it may be difficult to perform visual screening of artificial sRNAs. If that is the case, optimization of the reporter plasmid may be necessary. For example, use of stronger promoters, shorter or longer 5′ leader sequences, different host strains, or other variants of fluorescent proteins such as superfolder GFP [8] may be considered.
7. It is recommended to perform **steps 8**–**11** in small scale first (one tube of competent cells) to optimize the transformation efficiency.
8. Make plates on the day of transformation and pour 12–15 mL medium per plate (100 mm Petri dish). In our hands, thin agar plates allow for easier screening of sRNA activity as ON cells appear much brighter on thin plates probably due to better penetration of the UV light through the agar. Such plates should be screened within 1 week of transformation. Arabinose is necessary only when using an arabinose-inducible plasmid.

9. If further improvement in the artificial sRNA activity is desired, a secondary artificial sRNA library may be designed and the screening steps can be repeated. For example, a short, recurring sequence was used to create a partially randomized secondary library in our recent work to develop anti-*fliC* artificial sRNAs [6].

References

1. Mizuno T, Chou MY, Inouye M (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci U S A 81:1966–1970
2. Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harbor Perspect Biol. doi:10.1101/cshperspect.a003798
3. Kawamoto H, Koide Y, Morita T, Aiba H (2006) Basepairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol Microbiol 61:1013–1022
4. Schmidt M, Zheng P, Delihas N (1995) Secondary structures of Escherichia coli antisense micF RNA, the 50-end of the target ompF mRNA, and the RNA/RNA duplex. Biochemistry 34:3621–3631
5. Møller T, Franch T, Udesen C et al (2002) Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. Genes Dev 16:1696–1706
6. Sharma V, Yamamura A, Yokobayashi Y (2012) Engineering artificial small RNAs for conditional gene silencing in Escherichia coli. ACS Synth Biol 1:6–13
7. Smith C, Heyne S, Richter AS et al (2010) Freiburg RNA tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res 38:W373–W377
8. Pedelacq JD, Cabantous S, Tran T et al (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 24:79–88

Chapter 9

Construction and Engineering of Large Biochemical Pathways via DNA Assembler

Zengyi Shao and Huimin Zhao

Abstract

DNA assembler enables rapid construction and engineering of biochemical pathways in a one-step fashion by exploitation of the in vivo homologous recombination mechanism in *Saccharomyces cerevisiae*. It has many applications in pathway engineering, metabolic engineering, combinatorial biology, and synthetic biology. Here we use two examples including the zeaxanthin biosynthetic pathway and the aureothin biosynthetic gene cluster to describe the key steps in the construction of pathways containing multiple genes using the DNA assembler approach. Methods for construct design, pathway assembly, pathway confirmation, and functional analysis are shown. The protocol for fine genetic modifications such as site-directed mutagenesis for engineering the aureothin gene cluster is also illustrated.

Key words DNA assembler, In vivo homologous recombination, Pathway engineering, Synthetic biology, Metabolic engineering, Gene cluster characterization and engineering

1 Introduction

Methods that enable rapid construction and engineering of biochemical pathways are invaluable in pathway engineering, metabolic engineering, combinatorial biology, and synthetic biology [1–6]. In all these studies, the conventional multistep sequential cloning method is typically used, which includes primer design, PCR amplification, restriction digestion, in vitro ligation, and transformation [7, 8]. In many cases, multiple plasmids are often required [7, 8]. This method is not only time-consuming and inefficient but also relies on unique restriction sites that become limited for large recombinant DNA molecules.

Thanks to its high efficiency and ease to work with, in vivo homologous recombination in yeast has been widely used for gene cloning, plasmid construction, and library creation [9–12]. Recently, we developed a new method, called “DNA assembler,” which enables design and rapid construction of large biochemical pathways in a one-step fashion by exploitation of the in vivo

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_9, © Springer Science+Business Media New York 2013

homologous recombination mechanism in *S. cerevisiae* [13]. This method is highly efficient and circumvents many potential problems associated with conventional cloning method such as requirement of time-consuming, multistep procedures, low efficiency, and dependence on unique restriction sites. Therefore, it represents a versatile approach for construction of biochemical pathways for synthetic biology, metabolic engineering, and functional genomics studies.

In addition to assembling biochemical pathways, DNA assembler also offers unprecedented flexibility and versatility in characterizing and engineering natural product gene clusters [14]. Microorganisms and plants have evolved to produce a myriad array of complex molecules known as natural products or secondary metabolites that are of biomedical and biotechnological importance [15–17]. Sequenced genomes and metagenomes represent a tremendously rich source for discovery of novel pathways involved in natural product biosynthesis [18, 19]. Over the last 2 decades, the complete genome sequences of more than 1,900 organisms have been determined, with more than 10,000 organisms in the pipeline (http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi). However, only a tiny fraction of the biosynthetic pathways from these organisms have been characterized, and discovery and sustainable production of natural products are often hampered by our limited ability to manipulate the corresponding biosynthetic pathways. Using DNA assembler, we developed a genomics-driven, synthetic biology-based strategy for rapid characterization and engineering of natural product biosynthetic pathways [14].

Here we use the zeaxanthin biosynthetic pathway and the aureothin biosynthetic cluster as two examples to illustrate the experimental procedures. Figure 1 shows the scheme for constructing the zeaxanthin biosynthetic pathway. Briefly, for each individual gene in the zeaxanthin pathway, an expression cassette including a promoter, a structural gene, and a terminator is PCR-amplified and assembled using overlap extension PCR (OE-PCR) [20]. The 5′ end of the first gene expression cassette is designed to overlap with a vector, while the 3′ end is designed to overlap with the second cassette. Each successive cassette is designed to overlap with the two flanking ones and the 3′ end of the last cassette overlaps with the vector. All overlaps are designed to be at least 50 bp for efficient in vivo homologous recombination (*see* **Note 1**). The resulting multiple expression cassettes are co-transformed into *S. cerevisiae* with the linearized vector through electroporation, which allows the entire pathway to be assembled into a vector. Restriction digestion or DNA sequencing is subsequently used to verify the correctly assembled pathway, after which the cells carrying the correct construct are checked for zeaxanthin production.

For engineering the aureothin gene cluster, as shown in Fig. 2, pathway fragments encoding the aureothin biosynthetic pathway

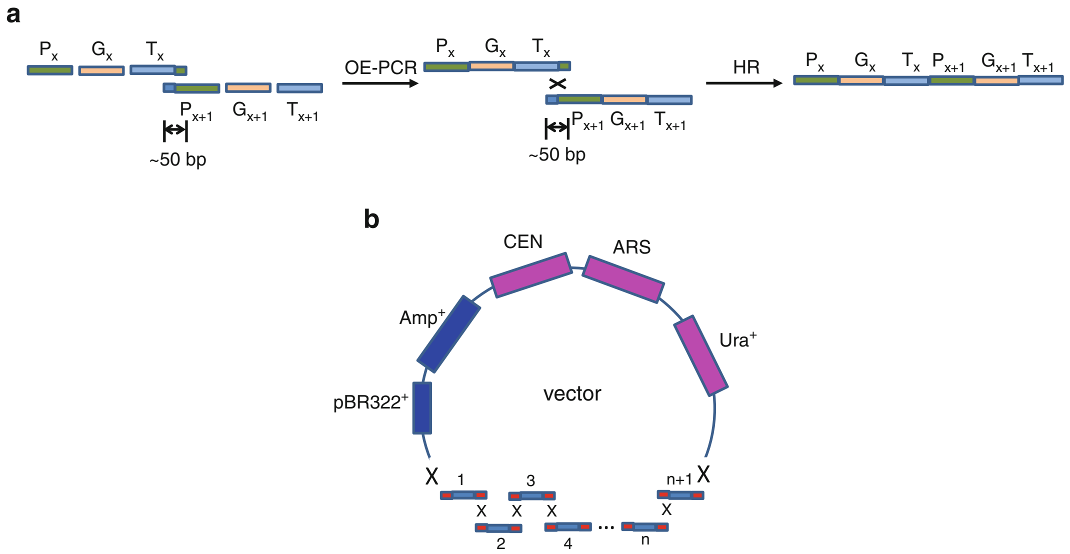


Fig. 1 (**a**) Preparation of each gene expression cassette using OE-PCR. Promoters (P_x, P_{x+1}), genes (G_x, G_{x+1}), and terminators (T_x, T_{x+1}) are individually PCR-amplified and joined together by OE-PCR. The resulting two cassettes are fused through the in vivo homologous recombination (HR) process. To generate an overlap of approximately 50 bp, the reverse primer used to amplify T_x contains a sequence of the first 20–25 nucleotides of P_{x+1}, and the forward primer used to amplify P_{x+1} contains a sequence of the last 20–25 nucleotides of T_x. (**b**) One-step method for assembly of a biochemical pathway using in vivo homologous recombination (HR) in *S. cerevisiae*

are amplified from the isolated genomic DNA of the native aureothin producer *Streptomyces thioluteus*. The helper fragments carrying the genetic elements needed for DNA maintenance and replication in *S. cerevisiae*, *E. coli*, and the target heterologous expression host *Streptomyces lividans* (*see* **Note 2**) are amplified from the corresponding vectors. Since PCR primers are designed to generate an overlap region between two adjacent fragments (*see* **Note 3**), these fragments will be assembled into a single DNA molecule in *S. cerevisiae* through homologous recombination after co-transformation. The isolated plasmids are transformed into *E. coli* for plasmid enrichment and verification, and the correct construct is transformed into *S. lividans* for heterologous expression of the aureothin biosynthetic pathway. Of special note, because the pathway fragments can be readily modified by PCR, various sophisticated genetic manipulations such as point mutagenesis and scarless gene substitution and deletion can be easily performed to confirm gene function, locate key amino acid residues, study biosynthetic mechanisms, express biosynthetic pathways heterologously, and generate new derivatives (*see* **Note 4**).

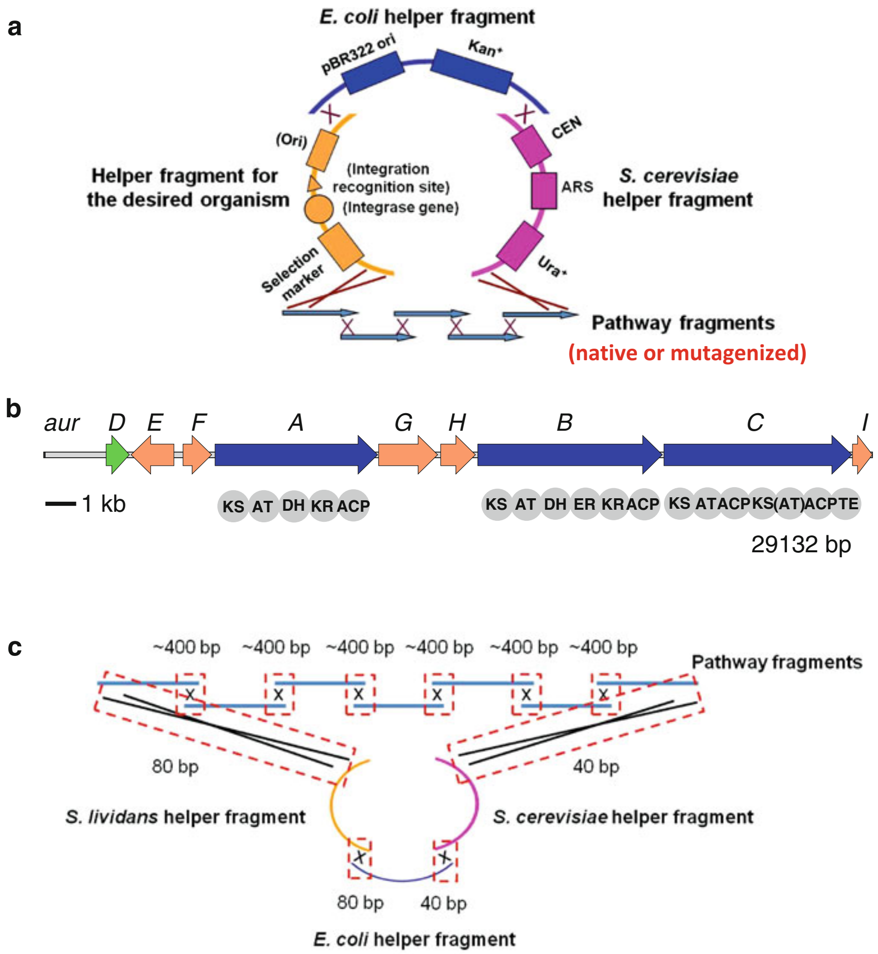


Fig. 2 (**a**) The DNA assembler-based strategy for efficient manipulation of natural product biosynthetic pathways. Various genetic modifications are introduced in the pathway fragments to be assembled. (**b**) The aureothin biosynthetic gene cluster from *S. thioluteus*. (**c**) The overlap lengths between adjacent fragments in assembly

2 Materials

Prepare all solutions using ultrapure water, prepared by purifying deionized water to attain a resistivity of 18.2 mΩ cm at 25 °C. Prepare and store all reagents at room temperature unless indicated otherwise.

2.1 DNA Preparation

1. pRS416: Obtained from New England Biolabs (Beverly, MA, USA) and serves as the template for amplifying the *S. cerevisiae* helper fragment for assembling the aureothin gene cluster variants.

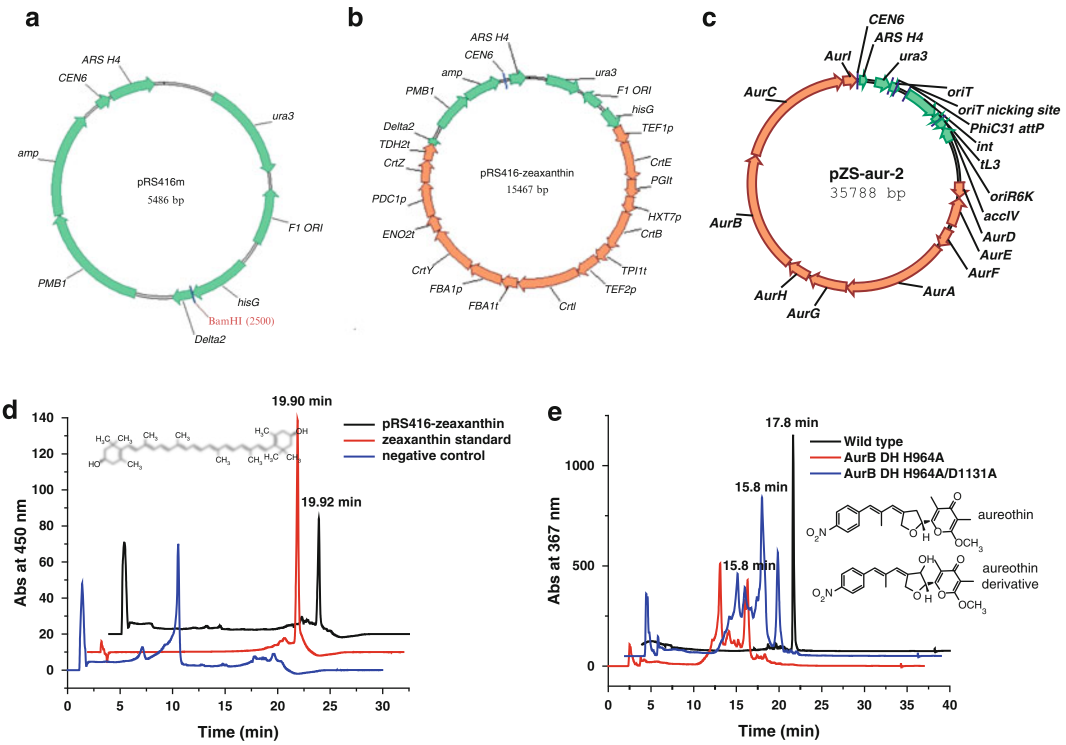


Fig. 3 (**a**) The vector map of pRS416m. (**b**) The vector map of the construct pRS416m-zeaxanthin. *CEN6*: centromere; *ARS H4*: automatic replication sequence; *ura3*: selection marker in *S. cerevisiae*; *amp*: selection marker in *E. coli*; *hisG* and *delta2*: two regions flanking the zeaxanthin biosynthetic pathway; *PMB1*: *E. coli* origin of replication. *F1 ORI*: this region is contained in the original pRS416 vector, but is not required for the construction of the plasmid containing the zeaxanthin biosynthetic pathway. (**c**) The vector map of the construct pZS-aur-2. *oriR6K*: *E. coli* origin of replication; *accIV*: apramycin resistance gene (it can be used as the selection marker in both *Streptomyces* and *E. coli*); *oriT*: conjugal transfer ori; *PhiC31 attP*: the ϕC31 recognition site; *int*: the ϕC31 integrase; *tL3*: a terminator. (**d**) HPLC analysis of the cell extracts from *S. cerevisiae* carrying the zeaxanthin biosynthesis pathway. Cells carrying the empty vector pRS416 were used as the negative control. Zeaxanthin structure was shown. (**e**) LC-MS analysis of the *S. lividans* clones carrying the wild type aureothin biosynthetic pathway and the mutant aureothin biosynthetic pathways. The structures of aureothin and its derivative were shown

2. pRS416m: pRS416 with a *hisG* sequence and a *delta2* [21] sequence that flank the multiple cloning site and serves as the vector for assembly of the zeaxanthin pathway (Fig. 3a) (*see* **Note 5**).

3. pCAR-ΔCrtX: It contains the genes *crtE*, *crtB*, *crtI*, *crtY*, and *crtZ* from *Erwinia uredovora* for zeaxanthin biosynthesis (Prof. E.T. Wurtzel, City University of New York, NY, USA) [22–24].

4. *S. thioluteus* strain: Obtained from the American Tissue Culture Collection (ATCC 12310, Manassas, VA, USA).

5. The genomic DNA of *S. thioluteus*: Isolated from *S. thioluteus* using Wizard Genomic DNA Isolation Kit (Promega, Madison, WI, USA).
6. pAE4: A *Streptomyces–E. coli* shuttle vector obtained from Professor William Metcalf (University of Illinois, Urbana, IL, USA) and serves as the template for amplifying the *S. lividans* helper fragment and the *E. coli* helper fragment for assembly of the aureothin gene cluster (the complete sequence of the plasmid can be obtained by request from the authors).
7. *E. coli* strain WM6026: Obtained from Professor William Metcalf and serves as the donor strain for conjugating plasmids to *S. lividans* (the details of strain construction can be obtained by request from the authors).
8. 0.5 M ethylenediamine tetraacetic acid (EDTA) solution (pH 8.0): For a 500 mL of stock solution of 0.5 M EDTA, weigh out 93.05 g of EDTA disodium salt (MW = 372.2) and dissolve it in 400 mL of deionized water. Adjust to pH 8.0 with NaOH and correct the final volume to 500 mL. EDTA will not be dissolved completely in water unless the pH is adjusted to about 8.0.
9. Concentrated stock solution of TAE (50×): Weigh 242 g of Tris base (MW = 121.14) and dissolve it in approximately 750 mL of deionized water. Carefully add 57.1 mL of glacial acid and 100 mL of 0.5 M EDTA, and adjust the solution to a final volume of 1 L. This stock solution can be stored at room temperature. The pH of this buffer is not adjusted and should be about 8.5.
10. Working solution of TAE buffer (1×): Dilute the stock solution by 50-fold with deionized water. Final solute concentrations are 40 mM Tris acetate and 1 mM EDTA.
11. 0.7 % agarose gel in 1× TAE buffer: Add 0.7 g of agarose into 100 mL of 1× TAE buffer and microwave until agarose is completely melted. Cool the solution to approximately 70–80 °C. Add 5 μL of ethidium bromide into the solution and mix well. Pour 25–30 mL of solution onto an agarose gel rack with appropriate 2-well or 8-well combs.
12. Wizard Genomic DNA Isolation Kit (Promega, Madison, WI, USA).
13. QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA, USA).
14. QIAprep Miniprep Kit (QIAGEN, Valencia, CA, USA).
15. DNA polymerase: Any polymerase with high fidelity can be used.
16. Failsafe PCR 2× Premix G: Containing dNTPs and PCR buffer (EPICENTRE Biotechnologies, Madison, WI, USA).

17. 5× Phusion GC Reaction Buffer (New England Biolabs, Beverly, MA, USA).
18. 40× dNTPs premix: 10 mM each nucleotide.
19. Dimethyl sulfoxide (DMSO).
20. *BamH*I restriction enzyme (New England Biolabs, Ipswich, MA, USA).
21. 3 M sodium acetate pH 5.0: Weigh 12.3 g of sodium acetate (MW = 82.03) and dissolve it into 50 mL of deionized water. Adjust to pH 5.0 by HCl.
22. 10 mg/mL glycogen: Dissolve 10 mg of glycogen in 1 mL of deionized water.
23. NanoDrop2000c: Used to measure the concentration of DNA and check the OD_{600} of the cells (Thermo Scientific, Wilmington, DE, USA).
24. Benchtop centrifuges to separate cells and supernatant.
25. Molecular imager gel doc: Used to check DNA on the agarose gel (Bio-Rad, Hercules, CA, USA).

2.2 Transformation

1. *S. cerevisiae* HZ848 (*MATα*, *ade2-1*, *Δura3*, *his3-11*, *15*, *trp1-1*, *leu2-3*, *112*, *can1-100*): Used as the host for DNA assembly (*see* **Note 6**).
2. YPAD medium: Dissolve 6 g of yeast extract, 12 g of peptone, 12 g of dextrose, and 60 mg of adenine hemisulfate in 600 mL of deionized water. Autoclave at 121 °C for 15 min.
3. Synthetic complete drop-out medium lacking uracil (SC-Ura): Dissolve 3 g of ammonium sulfate, 1 g of yeast nitrogen source without ammonium sulfate and amino acids, 0.5 g of complete synthetic medium minus uracil (CSM-Ura; MP Biomedicals, Solon, OH, USA), 26 mg of adenine hemisulfate, and 12 g of dextrose in 600 mL of deionized water, and adjust the pH to 5.6 by NaOH. Autoclave at 121 °C for 15 min.
4. SC-Ura-agar: SC-Ura and 20 g/L of agar.
5. 1 M sorbitol solution: Dissolve 91.1 g of sorbitol (MW = 182.17) in 400 mL of deionized water and adjust to a final volume of 500 mL. Sterilize the solution by filtering it through a filter with a pore size of 0.22 μm.
6. Gene Pulser II and Pulse controller plus: Used to transform plasmids into *S. cerevisiae* and *E. coli* through electroporation (Bio-Rad, Hercules, CA, USA).

2.3 Verification of the Clones

1. Zymoprep II yeast plasmid miniprep (Zymo Research, Orange, CA, USA).
2. *E. coli* strain BW25141 (*F-*, *Δ*(*araD-araB*)*567*, *ΔlacZ4787*(:: *rrnB-3*), *Δ*(*phoB-phoR*)*580*, *λ*$^-$, *galU95*, *ΔuidA3*::*pir*$^+$, *recA1*,

endA9(del-ins)::FRT, rph-1, Δ(rhaD-rhaB)568, hsdR514): Work as a standard cloning strain (obtained from Professor William Metcalf at the University of Illinois, Urbana, IL).

3. *Sca*I, *Psi*I, *Sac*I, and *Asc*I restriction enzymes (New England Biolabs, Ipswich, MA, USA) (*see* **Note** 7).
4. 1 M glucose solution: Dissolve 90 g of D-glucose in 400 mL of deionized water and adjust to a final volume of 500 mL. Filter-sterilize it.
5. SOC medium: Add 20 g of bacto-tryptone, 5 g of yeast extract, 0.5 g of NaCl, and 186.4 mg of KCl into 980 mL of deionized water. Adjust the pH to 7.0 with NaOH. Autoclave at 121 °C for 15 min. After the solution cools down to 70–80 °C, add 20 mL of sterile 1 M glucose.
6. 100 mg/mL ampicillin stock solution: Dissolve 1 g of ampicillin powder in 10 mL of deionized water and filter-sterilize it.
7. 50 mg/mL apramycin stock solution: Dissolve 0.5 g of apramycin powder in 10 mL of deionized water and filter-sterilize it.
8. LB broth: Add 10 g of bacto-tryptone, 5 g of yeast extract, and 10 g of NaCl into 1 L of deionized water. Autoclave at 121 °C for 15 min.
9. LB-agar: LB broth and 20 g/L agar.
10. LB-Amp$^+$ agar plates: Autoclave LB-agar and when the solution cools down to 70–80 °C, add 1 mL of 100 mg/mL ampicillin to 1 L of LB-agar. Pour 20–25 mL into each Petri dish.
11. LB-Apr$^+$ agar plates: Autoclave LB-agar and when the solution cools down to 70–80 °C, add 1 mL of 50 mg/mL apramycin to 1 L of LB-agar. Pour 20–25 mL into each Petri dish.

2.4 Detection of Zeaxanthin

1. 0.1 % trifluoroacetic acid (TFA) buffer: Add 1 mL of TFA into 999 mL of deionized water.
2. 10 μg/mL zeaxanthin standard solution: Dissolve 1 mg of zeaxanthin (Sigma, St Louis, MO, USA) in 100 mL of methanol.
3. French pressure cell press: Used to lyse the yeast cells.
4. Rotary evaporator R-205: Used to evaporate the solvent.
5. High performance liquid chromatography (HPLC) equipment.
6. ZORBAX SB-C18 column (Agilent Technologies, Palo Alto, CA, USA).

2.5 Conjugation of Plasmids into *S. lividans*

1. 3.8 mg/mL 2,6-diaminopimelic acid (DAP) stock solution: Dissolve 38 mg of DAP powder in 10 mL of deionized water and filter-sterilize it.

2. LB-Apr$^+$-DAP plates: LB-Apr$^+$ plates and 19 μg/mL DAP.
3. A mixture of nalidixic acid and apramycin each at the concentration of 1 mg/mL: Dissolve 20 mg of nalidixic acid powder in 15 mL of deionized water. Nalidixic acid will not be dissolved completely in water unless the pH is adjusted to approximately 12.0. Slowly adjust pH to 9.0–9.5 with 1 M HCl and add 400 μL of 50 mg/mL apramycin. The final pH after supplementing apramycin should drop to approximately 7.8–8.2. Correct the final volume to 20 mL and filter-sterilize it.
4. *S. lividans* spores: Prepared according to the protocol in the book "Practical Streptomyces Genetics," second edition, pp. 44–47 [25].
5. R2 no-sucrose agar plates: Add 0.25 g of K_2SO_4, 10.12 g of $MgCl_2{\cdot}6H_2O$, 10 g of glucose, 0.1 g of Difco casamino acids, and 5.73 g of TES into 990 mL of deionized water. Autoclave at 121 °C for 15 min. After the solution cools down to 70–80 °C, add 1 mL of 5 mg/mL KH_2PO_4, 8 mL of 36.8 mg/mL $CaCl_2{\cdot}2H_2O$, 1.5 mL of 0.2 g/mL L-proline, and 0.5 mL of 1 M NaOH (all these solutions need to be filter-sterilized in advance). Pour 20–25 mL into each Petri dish. Please refer to the recipe in the book "Practical Streptomyces Genetics," p. 408 [25].
6. ISP2-Apr$^+$ agar plates: Add 10 g of malt extract, 4 g of yeast extract, 4 g of glucose, and 20 g of agar into 1 L of deionized water. Adjust pH to 7.2 and autoclave at 121 °C for 15 min. When the solution cools down to 70–80 °C, add 1 mL of 50 mg/mL apramycin. Mix well and pour 20–25 mL into each Petri dish.

2.6 Culturing of the Aureothin Pathway Variants and Detection of Aureothin and Its Derivatives

1. MYG medium: Add 10 g of malt extract, 4 g of yeast extract, and 4 g of glucose into 1 L of deionized water. Adjust pH to 7.2 and autoclave at 121 °C for 15 min.
2. Ethyl acetate.
3. 100 μg/mL aureothin standard solution: Dissolve 1 mg of aureothin (BioAustralis, Smithfield, Australia) in 10 mL of methanol.
4. Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA).

3 Methods

3.1 DNA Preparation for Assembling the Zeaxanthin Biosynthetic Pathway

1. Amplify the genes *crtE*, *crtB*, *crtI*, *crtY*, and *crtZ* from the plasmid pCAR-ΔCrtX and amplify the corresponding promoter and terminator from the genomic DNA of *S. cerevisiae* using the primers listed in Table 1. Set up the reaction mixtures as follows: 50 μL of FailSafe PCR 2xPreMix G, 2.5 μL of

Table 1
The primers used in assembling the zeaxanthin pathway in pRS416m

Name	Sequence
hisG-f	GGCCAGTGAGCGCGCGTAATACGACTCACTATAGGCGCGCCTGCGTGAAGTCGAAG
hisG-r	GGAGTAGAAACATTTTGAAGCTATTTCCAGTCAATCAGGGTATTG
TEF1p-f	CTTCAATACCCTGATTGACTGGAAATAGCTTCAAAATGTTTCTACTC
TEF1p-r	GAACGTGTTTTTTTGCGCAGACCGTCATTTTGTAATTAAAACTTAGATTAG
CrtE-f	CTAATCTAAGTTTTAATTACAAAATGACGGTCTGCGCAAAAAAACACGTTC
CrtE-r	GGTATATATTTAAGAGCGATTTGTTTAACTGACGGCAGCG
PGIt-f	CGCTGCCGTCAGTTAAACAAATCGCTCTTAAATATATACC
PGIt-r	CCGAAATTGTTCCTACGAGAAGTGGTATACTGGAGGCTTCATGAGTTATG
HXT7p-f	CATAACTCATGAAGCCTCCAGTATACCACTTCTCGTAGGAACAATTTCGG
HXT7p-r	GAGTAACGACGGATTATTCATTTTTTGATTAAAATTAAAAAAAC
CrtB-f	GTTTTTTTAATTTTAATCAAAAAATGAATAATCCGTCGTTACTC
CrtB-r	GATAATATTTTTATATAATTATATTAATCCTAGAGCGGGCGCTGCCAGAG
TPI1t-f	CTCTGGCAGCGCCCGCTCTAGGATTAATATAATTATATAAAAATATTATC
TPI1t-r	CTATATGTAAGTATACGGCCCTATATAACAGTTGAAATTTGG
TEF2p-f	CCAAATTTCAACTGTTATATAGGGCCGTATACTTACATATAG
TEF2p-r	CACCAATTACCGTAGTTGGTTTCATGTTTAGTTAATTATAGTTCGTTG
CrtI-f	CAACGAACTATAATTAACTAAACATGAAACCAACTACGGTAATTGGTG
CrtI-r	CTCATTAAAAAACTATATCAATTAATTTGAATTAACTCATATCAGATCCTCCAGCATC
FBA1t-f	GATGCTGGAGGATCTGATATGAGTTAATTCAAATTAATTGATATAGTTTTTTAATGAG
FBA1t-r	GTTCAAGCCAGCGGTGCCAGTTGGAGTAAGCTACTATGAAAGACTTTAC

(continued)

Table 1
(continued)

Name	Sequence
FBA1p-f	GTAAAGTCTTTCATAGTAGCTTACTCCAACTGGCACCGCTGGCTTGAAC
FBA1p-r	CAGATCATAATGCGGTTGCATTTTGAATATGTATTACTTGGTTATGG
CrtY-f	CCATAACCAAGTAATACATATTCAAAATGCAACCGCATTATGATCTG
CrtY-r	CTAATAATTCTTAGTTAAAAGCACTTTAACGATGAGTCGTCATAATGG
ENO2t-f	CCATTATGACGACTCATCGTTAAAGTGCTTTTAACTAAGAATTATTAG
ENO2t-r	GGAACATATGCTCACCCAGTCGCATGAGGTATCATCTCCATCTCCCATATG
PDC1p-f	CATATGGGAGATGGAGATGATACCTCATGCGACTGGGTGAGCATATGTTCC
PDC1p-r	GGCATTCCAAATCCACAACATTTTGATTGATTTGACTGTG
CrtZ-f	CACAGTCAAATCAATCAAAATGTTGTGGATTTGGAATGCC
CrtZ-r	CATTAAAGTAACTTAAGGAGTTAAATTTACTTCCCGGATGCGGGCTC
TDH2t-f	GAGCCCGCATCCGGGAAGTAAATTTAACTCCTTAAGTTACTTTAATG
TDH2t-r	GATCCGTTAGACGTTTCAGCTTCCAGCGAAAAGCCAATTAGTGTGATAC
Delta2-f	GTATCACACTAATTGGCTTTTCGCTGGAAGCTGAAACGTCTAACGGATC
Delta2-r	TTACGCCAAGCGCGCAATTAACCCTCACTAAAGGCGCGCCGAGAACTTCTAGTATATTC

forward primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), 1 μL of template (10–50 ng of *S. cerevisiae* genomic DNA or the plasmid pCAR-ΔCrtX), 1 μL of DNA polymerase, and 43 μL of ddH$_2$O in a total volume of 100 μL.

2. PCR condition: Fully denature at 98 °C for 30 s, followed by 25 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.
3. Load the 100 μL of PCR products onto 0.7 % agarose gels and perform electrophoresis at 120 V for 20 min.
4. Gel-purify PCR products using the QIAquick Gel Extraction Kit.

5. Check the concentrations of the purified products using NanoDrop.
6. Perform OE-PCR to generate each gene expression cassette (*see* **Note 8**). Set up the first-step reaction mixture as follows: 10 μL of FailSafe PCR 2xPreMix G, 100 ng of promoter fragment, 100 ng of *crt* gene fragment, 100 ng of terminator fragment, and 0.2 μL of DNA polymerase. Add ddH_2O to a final volume of 20 μL.
7. Reaction condition: Fully denature at 98 °C for 30 s, followed by 10 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.
8. Set up the second-step reaction mixture as follows: 50 μL of FailSafe PCR 2× PreMix G, 10 μL of first-step reaction mixture, 1 μL of DNA polymerase, 2.5 μL of forward primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), and 34 μL ddH_2O in a total volume of 100 μL.
9. Reaction condition: Fully denature at 98 °C for 30 s, followed by 25 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.
10. Digest pRS416m by *BamH*I at 37 °C for 3 h. Digestion condition: 5 μL of 10× buffer, 0.5 μL of 100× BSA, 3 μg of pRS416m, and 30 units of *BamH*I. Add ddH_2O to a final volume of 50 μL.
11. Load the PCR and digestion products onto 0.7 % agarose gels and perform electrophoresis at 120 V for 20–30 min.
12. Gel-purify the PCR and digestion products using the QIAquick Gel Extraction Kit.
13. Check the concentrations of the purified products using NanoDrop.
14. Take 200–300 ng of each fragment, mix in a tube, and calculate the final volume.
15. Add 10 % v/v 3 M sodium acetate and 2 % v/v 10 mg/mL glycogen (e.g., if there is 100 μL of mixture, add 10 μL of sodium acetate and 2 μL of glycogen) and mix well.
16. Add 2× v/v 100 % ethanol (e.g., if the final volume is about 110 μL, add 220 μL ethanol) and mix well.
17. Store the DNA mixture at −80 °C for at least an hour.
18. Centrifuge at 4 °C, 16,100 × *g* for 20 min. Usually the precipitated DNA can be seen at the bottom of the tube.
19. Remove the supernatant completely (do not touch the DNA).
20. Add 500 μL of 70 % ethanol to wash the DNA pellet, and centrifuge at room temperature, 16,100 × *g* for 3 min.

21. Remove the ethanol completely and air dry the pellet for 1–2 min (do not overdry it).
22. Resuspend the DNA pellet in 4 μL of ddH_2O. Now the DNA is ready for transformation (*see* **Note 9**).

3.2 DNA Preparation for Assembling the Aureothin Gene Cluster Variants

1. Amplify the aureothin biosynthetic pathway fragments from the genomic DNA of *S. thioluteus* and amplify the *S. cerevisiae*, *E. coli*, and *S. lividans* helper fragments from the corresponding vectors using the primers listed in Table 2. For constructing the mutant pathways, mutations are added into the corresponding primers. Set up the reaction for amplifying the *S. cerevisiae* helper fragment as follows: 20 μL of Buffer GC, 2.5 μL of dNTP premix, 2.5 μL of forward primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), 1 μL of template pRS416 at the concentration of 50–100 ng/μL, 1 μL of DNA polymerase, and 70.5 μL of ddH_2O in a total volume of 100 μL (*see* **Note 10**). Set up all the other reaction mixtures as follows: 50 μL of FailSafe PCR 2× PreMix G, 2.5 μL of forward primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), 1 μL of template (10–50 ng of *S. thioluteus* genomic DNA or the plasmid pAE4), 1 μL of DNA polymerase, 5 μL of DMSO (*see* **Note 11**), and 38 μL of ddH_2O in a total volume of 100 μL.
2. PCR condition for amplifying the *S. cerevisiae* helper fragment: Fully denature at 98 °C for 30 s, followed by 25 cycles of 98 °C for 10 s, 45 °C for 30 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min (*see* **Note 10**). PCR condition for all the other fragments: Fully denature at 98 °C for 30 s, followed by 25 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 1–3 min, with a final extension at 72 °C for 10 min (*see* **Note 11**).
3. Load the 100 μL of PCR products onto 0.7 % agarose gels and perform electrophoresis at 120 V for 20 min.
4. Gel-purify PCR products using the QIAquick Gel Extraction Kit.
5. Check the concentrations of the purified products using NanoDrop.
6. Take 200–300 ng of each fragment, mix in a tube, and calculate the final volume.
7. Add 10 % v/v 3 M sodium acetate and 2 % v/v 10 mg/mL glycogen (e.g., if there is 100 μL of mixture, add 10 μL of sodium acetate and 2 μL of glycogen) and mix well.
8. Add 2× v/v 100 % ethanol (e.g., if the final volume is about 110 μL, add 220 μL ethanol) and mix well.
9. Store the DNA mixture at −80 °C for at least an hour.
10. Centrifuge at 4 °C, 16,100 × *g* for 20 min. Usually the precipitated DNA can be seen at the bottom of the tube.

Table 2
The primers used in assembling the aureothin gene cluster variants

Name	Sequence
Aur-1-for	TGGTACTGCAAATACGGCATCAGTTACCGTGAGCAGATCGGATCAGCTCGTCCCGTTCGG
Aur-1-rev	GCTGCTCTTCTCGCATCGTC
Aur-2-for	CGTAGAGGAGCTCCAGCAGC
Aur-2-rev	CTCCTCCAGCACCTCGCAGC
Aur-3-for	TCCTGACCTTCGACTCGCTG
Aur-3-rev	CATGTTCGATCCTTCCGTTG
Aur-4-for	GACGGTGCACCAGCTGGTCA
Aur-4-rev	GTTGCCGGTCATGTGGTAGC
Aur-5-for	ATGACCAATGACGCCAAGAC
Aur-5-rev	CCGTCCATCAGGTCGAACGC
Aur-6-for	CCTACTACGGCCTGGTGGAC
Aur-6-rev	CATCGCCGTCATCGAGACGA
Aur-7-for	GCAACGAAGGACATGTCCAG
Aur-7-rev	TTATAGCACGTGATGAAAAGGACCCAGGTGGCACTTTTCGTCAGTCAGTCGTCCAGGCGC
Yeast-for	CGAAAAGTGCCACCTGGGTC
Yeast-rev	AATATTGTGAGTTTAGTATACATGCA
Strep-for	GTATTATAAGTAAATGCATGTATACTAAACTCACAATATTATGGCGCGCCGACGTGCTCA
Strep-rev	ATTAGCCATGGCATCACAGTATCGTGATGACATTAATTAACGCAATCCAGTGCAAAGCTA
E. coli-for	AGACGAAGAAGCTAGCTTTGCACTGGATTGCGTTAATTAATGTCATCACGATACTGTGAT
E. coli-rev	GCCCATGACCACCGTCGTCTCCGAACGGGACGAGCTGATCCGATCTGCTCACGGTAACTG
AurB DH H964A-for	TGTCGGCGCGGACCGAGTCCTGGCTGGCCGACGCCGTCGTGCTCGGCTCCACGCTCGTCC
AurB DH H964A-rev	GGACGAGCGTGGAGCCGAGCACGACGGCGTCGGCCAGCCAGGACTCGGTCCGCGCCGACA
AurB DH D1131A-for	GCTACGGCGTCCACCCCGCGCTCCTCGCCGCCGCACTGCACACCGCCCTCCTGAAGGAGG
AurB DH D1131A-rev	CCTCCTTCAGGAGGGCGGTGTGCAGTGCGGCGGCGAGGAGCGCGGGGTGGACGCCGTAGC

The mutated codons are underlined

11. Remove the supernatant completely (do not touch the DNA).
12. Add 500 μL of 70 % ethanol to wash the DNA pellet, and centrifuge at room temperature, 16,100 × *g* for 3 min.
13. Remove the ethanol completely and air dry the pellet for 1–2 min (do not overdry it).
14. Resuspend the DNA pellet by 4 μL of ddH$_2$O. Now the DNA is ready for transformation (*see* **Note 9**).

3.3 Transformation

1. Inoculate a single colony of HZ848 into 3 mL of YPAD medium and grow overnight in a shaker at 30 °C and 250 rpm.
2. Measure the OD$_{600}$ of the seed culture and inoculate the appropriate amount to 50 mL of fresh YPAD medium to obtain an OD$_{600}$ of 0.2 (e.g., if the overnight culture has an OD$_{600}$ of 10, then add 1 mL into 50 mL of fresh YPAD medium).
3. Continue growing the 50 mL of culture for approximately 4 h to obtain an OD$_{600}$ of 0.8 (*see* **Note 12**).
4. Spin down the yeast cells at 4 °C, 3,220 × *g* for 10 min and remove the spent medium.
5. Use 50 mL of ice-cold ddH$_2$O to wash the cells once and centrifuge again.
6. Discard water, add 1 mL of ice-cold ddH$_2$O to resuspend the cells, and move them to a sterile Eppendorf tube.
7. Spin down the cells using a bench top centrifuge for 30 s at 4 °C, 4,500 × *g*.
8. Remove water and use 1 mL of 1 M ice-cold sorbitol to wash the cells once (now the cells look slightly yellow). Centrifuge again and remove the sorbitol.
9. Resuspend the cells in 250–300 μL of chilled 1 M sorbitol and distribute them into 50 μL aliquots.
10. Now each 50 μL of cells is ready for electroporation (*see* **Note 13**). Mix the 4 μL of DNA with 50 μL of yeast cells and put the mixture into a chilled electroporation cuvette.
11. Electroporate the cells at 1.5 kV and quickly add 1 mL of pre-warmed (30 °C) YPAD medium to resuspend cells (*see* **Note 14**).
12. Grow in a shaker at 30 °C, 250 rpm for 1 h.
13. Spin down the cells in a sterile tube at 16,100 × *g* for 30 s and remove the YPAD medium.
14. Use 1 mL of room temperature sorbitol solution to wash the cells 2–3 times and finally resuspend the cells in 1 mL sorbitol.
15. Spread 100 μL of resuspended cells onto SC-Ura plates.
16. Incubate the plates at 30 °C for 2–3 days until colonies appear.

3.4 Verification of the Correctly Assembled Pathways

1. Randomly pick ten colonies from the SC-Ura plate and inoculate each colony into 1.5 mL of SC-Ura liquid medium. Grow at 30 °C for 1.5 days (*see* **Note 15**).
2. Purify yeast plasmid DNA from each 1.5 mL of culture using the Zymoprep II kit.
3. Mix 2 μL of isolated plasmid DNA with 50 μL of *E. coli* BW25141 cells and put the mixture into a chilled electroporation cuvette (*see* **Note 16**).
4. Electroporate the cells at 2.5 kV and quickly add 1 mL of SOC medium to resuspend the cells (*see* **Note 14**).
5. Grow in a shaker at 37 °C, 250 rpm for 1 h.
6. Spin down the cells, remove 800 μL of SOC medium, resuspend the pellet with the remaining 200 μL of SOC medium, and spread the cells on an LB-Amp$^+$ plate for the zeaxanthin construct or on an LB-Apr$^+$ plate for the aureothin constructs.
7. Incubate the plates at 37 °C for 16–18 h until colonies appear (*see* **Note 17**).
8. Inoculate a single colony from each plate to 5 mL of LB supplemented with 100 μg/mL ampicillin (the zeaxanthin construct) or 50 μg/mL apramycin (the aureothin constructs), and grow at 37 °C for 12–16 h.
9. Purify *E. coli* plasmids from each 5 mL of culture using the QIAgen Miniprep kit.
10. Check the plasmid concentrations by NanoDrop.
11. Verify the correctly assembled pathway through two separate restriction digestion reactions for each construct (*see* **Note 7**).
12. Digestion condition by *Sca*I at 37 °C for 3 h: 1.5 μL of 10× buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *Sca*I. Add ddH$_2$O to a final volume of 15 μL. Expected bands for the assembled zeaxanthin pathway (Fig. 3b): 1,750, 2,131, 2,628, 3,223, 5,735 bp.
13. Digestion condition by *Psi*I at 37 °C for 3 h: 1.5 μL of 10× buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *Psi*I. Add ddH$_2$O to a final volume of 15 μL. Expected bands for the assembled zeaxanthin pathway (Fig. 3b): 215, 1,389, 1,689, 1,782, 2,425, 2,752, 5,215 bp.
14. Digestion condition by *Sac*I at 37 °C for 3 h: 1.5 μL of 10× buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *Sac*I. Add ddH$_2$O to a final volume of 15 μL. Expected bands for the assembled aureothin gene cluster variants (Fig. 3c): 163, 696, 752, 839, 1,207, 1,965, 2,370, 2,956, 3,074, 3,612, 3,693, 5,250, 9,211 bp.
15. Digestion condition by *Asc*I at 37 °C for 3 h: 1.5 μL of 10× buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of

*Asc*I. Add ddH_2O to a final volume of 15 μL. Expected bands for the assembled aureothin gene cluster variants (Fig. 3c): 487, 1,890, 2,927, 3,154, 3,716, 5,419, 5,433, 5,810, and 6,952 bp.

3.5 Detection of Zeaxanthin

1. Inoculate a single colony carrying the zeaxanthin biosynthetic pathway into 3 mL of SC-Ura liquid medium and grow at 30 °C, 250 rpm for 1.5 days.
2. Inoculate 2.5 mL of seed culture into 250 mL of fresh SC-Ura medium and continue to grow at 30 °C, 250 rpm for 4 days.
3. Cells are collected by centrifugation at 3,220 × *g*, resuspended with 5 mL of acetone, and lysed by French pressure cell at 10,000 psi.
4. Supernatants are collected after centrifugation at 16,100 × *g* for 3 min and evaporated to dryness using rotary evaporator.
5. After resuspension in 0.5–1 mL of methanol, 100 μL of sample is loaded onto the Agilent ZORBAX SB-C18 column and analyzed at 450 nm by HPLC with a 0.5 mL/min flow rate as follows: buffer A: H_2O with 0.1 % TFA, buffer B: 100 % CH_3OH; 0–3 min, 60 % CH_3OH; 3–15 min, linear gradient from 60 % CH_3OH to 100 % CH_3OH; 15–17 min, 100 % CH_3OH; 17–20 min, linear gradient from 100 % CH_3OH to 60 % CH_3OH. Authentic zeaxanthin is used as standard, which was eluted at 19.9 min.

3.6 Conjugation and Heterologous Expression of the Aureothin Pathway Variants in S. lividans

1. Mix 50–100 ng of each verified plasmid with 50 μL of *E. coli* WM6026 cells and put the mixture into a chilled electroporation cuvette.
2. Electroporate the cells at 2.5 kV, and quickly add 1 mL of SOC medium and 5 μL of 38 mg/mL DAP to resuspend cells (*see* **Notes 14** and **18**).
3. Grow in a shaker at 37 °C, 250 rpm for 1 h.
4. Spread 100 μL of each culture on an LB-Apr$^+$-DAP plate.
5. Incubate the plates at 37 °C for 16 h until colonies appear.
6. Inoculate a single colony from each plate to 2 mL of LB supplemented with 50 μg/mL apramycin and 10 μL of 38 mg/mL DAP, and grow at 37 °C for approximately 2 h until OD_{600} reaches 0.6–0.8.
7. Spin down 100 μL of cell cultures in an Eppendorf tubes and wash the cell pellets each with 1 mL of fresh LB medium.
8. Spin down the cells and wash one more time.
9. Resuspend the cell pellets each with 1 mL of LB.

10. Mix 2 μL of the resuspended cells with 25 μL of *S. lividans* spores by pipetting and spot 2 μL of aliquots onto R2 no-sucrose plates. Wait until all the spotted drops are absorbed into the plates.
11. Incubate the plates at 30 °C for 16–18 h.
12. Flood the plates with 2 mL of a mixture of nalidixic acid and apramycin each at a concentration of 1 mg/mL (*see* **Note 19**).
13. Incubate the plates at 30 °C for additional 3–5 days until exconjugants appear, at which point exconjugants are picked and restreaked on ISP2-Apr$^+$ plates and allowed to grow for 2 days.
14. Inoculate single colonies from the ISP2-Apr$^+$ plates into 10 mL of MYG supplemented with 50 μg/mL apramycin and grow the cultures at 30 °C for 2 days as seed cultures, of which 2.5 mL is subsequently inoculated to 250 mL of fresh MYG supplemented with 50 μg/mL apramycin and grown for another 84 h.

3.7 Detection of Aureothin and Its Derivatives

1. Centrifuge the cultures at 3,220 × *g* for 10 min to remove the cells.
2. Extract the supernatants with an equal volume of ethyl acetate and evaporate to dryness using rotary evaporator.
3. Perform LC–MS on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer with an Agilent SB-C18 reverse-phase column. HPLC parameters for detection of aureothin and its derivatives are as follows: solvent A, 1 % acetic acid in water; solvent B, acetonitrile; gradient, 10 % B for 5 min, to 100 % B in 10 min, maintain at 100 % B for 5 min, return to 10 % B in 10 min, and finally maintain at 10 % B for 7 min; flow rate 0.3 mL/min; detection by UV spectroscopy at 367 nm. Under such conditions, aureothin and its derivative are eluted at 17.8 min and 15.8 min, respectively. Mass spectra are acquired in ultra-scan mode using electrospray ionization (ESI) with positive polarity. The MS system is operated using a drying temperature of 350 °C, a nebulizer pressure of 35 psi, a drying gas flow of 8.5 L/min, and a capillary voltage of 4,500 V (*see* **Note 20**).

4 Notes

1. If a larger biochemical pathway needs to be assembled, increasing the length of the overlaps between the adjacent fragments is necessary. For example, to assemble a pathway with a size of ~25 kb, a longer overlap (e.g., 125 bp) could ensure high

assembly efficiency (>50 %), while low efficiency (10–20 %) is obtained if the length of the overlaps is only 50 bp [13].

2. *S. cerevisiae* is used as the assembly host, *E. coli* serves as the DNA enrichment host, and *S. lividans* is a widely used heterologous host for studying gene clusters from other *Streptomyces* species.

3. As shown in Fig. 2c, to ensure high assembly efficiency, overlaps of 400 bp between internal adjacent pathway fragments are generated. For example, the forward primer for amplifying the second pathway fragment can be located at ~400 bp upstream of the annealing position of the reverse primer for amplifying the first pathway fragment. Overlaps between other fragments are generated by adding tails to primers, thus they could not be very long. We encountered difficulties in amplifying the *S. cerevisiae* helper fragment, and in the end, the correct product was only obtained by using the pair of primers without any tails. As a result, the *S. cerevisiae* helper fragment only overlaps with the last pathway fragment and the adjacent helper fragment with 40 bp. Overlaps of 80 bp are generated between the *S. lividans* helper fragment and its neighbors.

4. Site-directed mutagenesis can be easily performed using DNA assembler. Unlike the labor-intensive and time-consuming procedures used in the conventional methods [26–28], DNA assembler only requires adding site-specific mutation(s) into the PCR primers used to generate pathway fragments. As an example, the active sites of the dehydratase domains (DH) of AurB are targeted (Fig. 2b). The motifs HXXXGXXXXP and DXXX(Q/H) were found to be conserved among the DH domains of polyketide synthases and the histidine and the aspartic acid were identified as the catalytic residues [29–31]. Based on this information, AurB DH H964A mutant and AurB DH H964A/D1131A double mutant will be generated. To confirm the necessity of gene(s), gene disruption can be carried out by two strategies: (1) a stop codon can be introduced into the early portion of the gene, resulting in in situ gene inactivation; (2) the complete gene can be omitted in pathway assembly by redesigning the overlaps. Refer to ref. 14 for more details.

5. The *hisG* and *delta2* sequences are not essential for assembling a pathway. As long as the 5′ end of the first promoter and the 3′ end of the last terminator contain overlaps (at least 50 bp) with the vector backbone, the pathway can be assembled on a plasmid. Similarly, any linearized *S. cerevisiae*–*E. coli* shuttle vector containing an *ura3* gene as a selection marker can be used as the vector backbone.

6. *S. cerevisiae* HZ848 is used as the host for DNA assembly. However, any *S. cerevisiae* strain with a nonfunctional *ura3* gene can be used as a host.

7. In order to verify the correctly assembled constructs through restriction digestion, a set of digestions consisting of one or two enzymes that cut the expected construct multiple times are chosen. Usually, two to three sets of digestions need to be set up in order to ensure the correct assembly. For a plasmid with a size of 15–20 kb, such as pRS416-zeaxanthin, find one or two enzymes which cut the DNA molecule 5–9 times. For a plasmid with a size of 20–30 kb, such as pZS-aur-2, find one or two enzymes which cut the DNA molecule 9–13 times. Try to avoid using enzyme digestion that will result in multiple fragments with similar sizes. Besides restriction digestion, the correctly assembled constructs can be confirmed by DNA sequencing.

8. In the construction of the first gene expression cassette and the last gene cassette, the *hisG* sequence and the *delta2* sequence are also included. Therefore, for these two reactions, four fragments are spliced together.

9. The fragment mixture can be maintained at −20 °C for several months.

10. We encountered difficulties in amplifying the *S. cerevisiae* helper fragment, and in the end, the correct product was obtained by using the designated reagents and PCR program mentioned above.

11. 58 °C is used as a standard annealing temperature. In some cases, especially for amplifying certain fragments from *Streptomyces*, fine-tuning annealing temperature is often necessary for obtaining the correct amplicons or improving PCR yields. Due to the high GC content of *Streptomyces* genome (>70 %), including 5 % DMSO in the reaction mixture will reduce the chance of forming secondary structures in general, resulting in better amplification efficiency. However, in some cases, we did encounter lower amplification efficiencies or even failures when DMSO was used. Generally, when difficulties are encountered to amplify a certain gene, a set of PCR conditions with various annealing temperatures and inclusion or exclusion of 5 % DMSO need to be tested.

12. Normally, the doubling time for a *S. cerevisiae* laboratory strain is approximately 2 h.

13. Unlike *E. coli*, yeast competent cells need to be freshly prepared each time.

14. For an efficient electroporation, a time constant of 5.0–5.2 ms should be obtained.

15. Assembly efficiency is defined as the percentage of the correct clones among the transformants appearing on the plate. Usually, ten colonies are picked and an average efficiency of 60–80 % can be obtained for assembly of the two target pathways.

16. *E. coli* strain BW25141 was used for plasmid enrichment and verification. However, any *E. coli* strain suitable for DNA cloning, such as DH5α and JM109, can be used.

17. The number of obtained *E. coli* transformants could vary from a few to several thousands. This is mainly due to the low quality of the isolated yeast plasmids. However, as long as colonies appear, experiments can proceed.

18. WM6026 is an auxotrophic *E. coli* strain whose growth relies on exogenously supplemented DAP.

19. Nalidixic acid is used to kill *E. coli* after it donates the plasmid into *S. lividans*, and apramycin is used to select the successful *S. lividans* exconjugants.

20. Aureothin has a molecular weight of 398.4 and the aureothin derivative generated by the AurB DH H964A mutant and the AurB DH H964A/D1131A double mutant has a molecular weight of 414.4. The MS2 fragmentation patterns of the produced compounds are compared with that of authentic aureothin. The products generated by the AurB DH H964A mutant and the AurB DH H964A/D1131A mutant exhibit the expected +16 patterns in the MS and MS2 profiles.

References

1. Hjersted JL, Henson MA, Mahadevan R (2007) Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97:1190–1204
2. Keasling JD (2008) Synthetic biology for synthetic chemistry. ACS Chem Biol 3:64–76
3. Menzella HG, Reid R et al (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat Biotechnol 23:1171–1176
4. Pitera DJ, Paddon CJ et al (2007) Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. Metab Eng 9:193–207
5. Ro DK, Paradise EM et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440:940–943
6. Szczebara FM, Chandelier C et al (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. Nat Biotechnol 21:143–149
7. Dejong JM, Liu Y et al (2006) Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. Biotechnol Bioeng 93:212–224
8. Yan Y, Kohli A, Koffas MA (2005) Biosynthesis of natural flavanones in *Saccharomyces cerevisiae*. Appl Environ Microbiol 71:5610–5613
9. Gunyuzlu PL, Hollis GF, Toyn JH (2001) Plasmid construction by linker-assisted homologous recombination in yeast. Biotechniques 31:1250–1252
10. Ma H, Kunes S et al (1987) Plasmid construction by homologous recombination in yeast. Gene 58:201–216
11. Oldenburg KR, Vo KT et al (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. Nucleic Acids Res 25:451–452
12. Raymond CK, Pownder TA, Sexson SL (1999) General method for plasmid construction using homologous recombination. Biotechniques 26(134–138):140–131
13. Shao Z, Zhao H, Zhao H (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:e16
14. Shao Z, Luo Y, Zhao H (2011) Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol Biosyst 7:1056–1059
15. Dewick PM (2002) Medical natural products. A biosynthetic approach, 2nd edn. Wiley, Chichester, UK
16. Herbert RB (1989) The biosynthesis of secondary metabolites, 2nd edn. Chapman and Hall, London, UK

17. Li JW, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325:161–165
18. Challis GL (2008) Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology 154:1555–1569
19. Zerikly M, Challis GL (2009) Strategies for the discovery of new natural products by genome mining. ChemBioChem 10:625–633
20. Horton RM, Hunt HD et al (1989) Engineering hybrid genes without the use of restriction enzymes - gene-splicing by overlap extension. Gene 77:61–68
21. Lee FW, Da Silva NA (1997) Sequential delta-integration for the regulated insertion of cloned genes in Saccharomyces cerevisiae. Biotechnol Prog 13:368–373
22. Chemler JA, Yan Y, Koffas MA (2006) Biosynthesis of isoprenoids, polyunsaturated fatty acids and flavonoids in *Saccharomyces cerevisiae.* Microb Cell Fact 5:20
23. Misawa N, Nakagawa M et al (1990) Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli.* J Bacteriol 172:6704–6712
24. Misawa N, Shimada H (1997) Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts. J Biotechnol 59:169–181
25. Kieser T, Bibb JM et al (2000) Practical streptomyces genetics. The John Innes Foundation, Norwich
26. Blodgett JA, Thomas PM et al (2007) Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. Nat Chem Biol 3:480–485
27. Ito T, Roongsawang N et al (2009) Deciphering pactamycin biosynthesis and engineered production of new pactamycin analogues. ChemBioChem 10:2253–2265
28. Karray F, Darbon E et al (2010) Regulation of the biosynthesis of the macrolide antibiotic spiramycin in *Streptomyces ambofaciens.* J Bacteriol 192:5813–5821
29. Keatinge-Clay A (2008) Crystal structure of the erythromycin polyketide synthase dehydratase. J Mol Biol 384:941–953
30. Moriguchi T, Kezuka Y et al (2010) Hidden function of catalytic domain in 6-methylsalicylic acid synthase for product release. J Biol Chem 285:15637–15643
31. Pawlik K, Kotowska M et al (2007) A cryptic type I polyketide synthase (cpk) gene cluster in *Streptomyces coelicolor A3(2).* Arch Microbiol 187:87–99

Chapter 10

Assembly of Multi-gene Pathways and Combinatorial Pathway Libraries Through ePathBrick Vectors

Peng Xu and Mattheos A.G. Koffas

Abstract

As an emerging discipline, synthetic biology is becoming increasingly important to design, construct, and optimize metabolic pathways leading to desired phenotypes such as overproduction of biofuels and pharmaceuticals in genetically tractable organisms. We have recently developed a versatile gene assembly platform ePathBricks supporting the modular assembly of multi-gene pathway components and combinatorial generation of pathway diversities. In this protocol, we will detail the process to assemble a seven gene flavonoid pathway (~9 kb) on one single ePathBrick vector. We will also demonstrate that a three-gene flavonoid pathway can be easily diversified to 54 pathway equivalents differing in pathway configuration and gene order; coupled with high-throughput screening techniques, we envision that this combinatorial strategy would greatly improve our ability to exploit the full potential of microbial cell factories for recombinant metabolite production.

Key words ePathBrick, Gene assembly, Synthetic biology, Combinatorial pathway library, Metabolic engineering

1 Introduction

Synthetic biology is characterized by a *constructive* approach to understand and manipulate biological systems [1]. As the cost of commercial synthesis of genes is declining, our ability to physically construct complex biological devices/pathways from basic DNA parts is becoming a critical hurdle in implementing the vision of synthetic biology [2]. These limitations can become prohibitive when constructing multi-gene metabolic pathways and complex regulatory circuits. For example, strain development through metabolic engineering leading to valuable pharmaceutical and fuel molecules [3] typically involves the manipulation of a dozen of precursor or rate-limiting pathways [4–6]. Conventional pathway construction approaches, which largely rely on smart design to assemble multi-gene fragments in operon form, are limited in terms of automation and context-independent pathway output.

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_10, © Springer Science+Business Media New York 2013

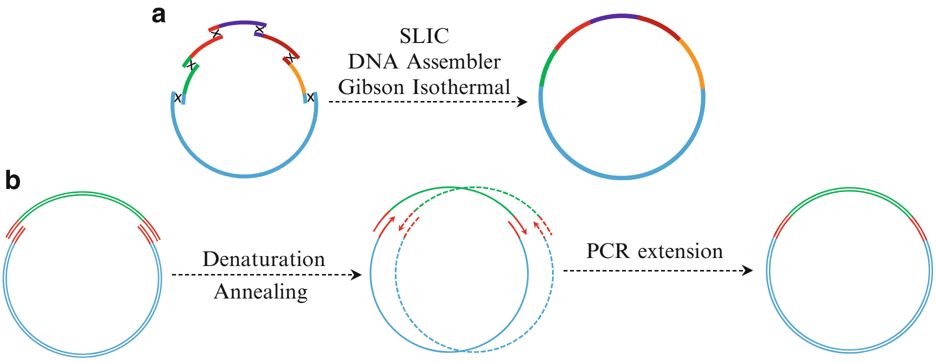


Fig. 1 DNA assembly tools for synthetic pathway construction. (**a**) Parallel gene assembly tools provide an efficient approach to construct directional multi-gene pathways from synthesized or PCR-based gene fragments with 30+ bp overlap region. SLIC and Gibson isothermal assembly are based on in vitro homologous recombination and single-strand annealing. DNA assembler is based on yeast in vivo homologous recombination. (**b**) CPEC (circular polymerase extension cloning) technique for gene assembly. Complementary overhangs attached to insertion gene fragment anneal to linearized vector and can be circularized into an intact vector by PCR extension

Therefore, synthetic metabolic engineering requires efficient tools that allow precise and concerted assembly of multiple gene fragments, leading to programmable or predictable pathway output customized for strain optimization.

Last decades' advancement in synthetic biology has led to the emergence of several robust gene assembly platforms that are suited for metabolic pathway construction and optimization. These gene assembly tools allow us to efficiently construct directional multi-gene pathways from synthesized or PCR-based gene fragments with 30+ bp overlap region (Fig. 1). For example, sequence and ligation-independent cloning (SLIC) [7, 8] and Gibson isothermal assembly [9] rely on in vitro homologous recombination and single-strand annealing. SLIC has been recently modified to allow one-step assembly of multiple medium-size DNA fragments [10] and large genomic DNA fragment (up to 28 kb) [11] using in vitro single-strand overlapping annealing. In vivo assembly tools based on yeast homologous recombination have also been developed for assembling large DNA fragments encoding an entire biochemical pathway [12] or iteratively integrating multi-gene pathways into yeast chromosome [13]. Coupled with yeast in vivo recombination, Gibson isothermal assembly has been used to assemble a self-replicating synthetic *M. genitalium* genome (1.08 Mb) [14] and high G + C genomic DNA fragment (454 kb) containing *Synechococcus elongatus* PCC 7942 photosynthetic gene clusters [15]. Circular polymerase extension cloning (CPEC) has been developed for one-step assembly of multi-gene pathways and combinatorial DNA libraries using complementary fragment annealing and PCR overlap extension [16].

Compared with conventional DNA cloning protocols, these advanced DNA assembly tools offer an efficient approach to construct multi-gene pathways in a one-step, scar-less, and sequence-independent manner. The downside of these assembly tools is the terminal flanking homology sequence of each DNA fragment must be specifically designed for each assembly junction, which tends to be laborious and error-prone [17]. Assembly of gene fragments with identical homology terminus (such as repeated promoters, repeated ribosome binding site, and repeated terminators) can be problematic, as this can lead to constructs with missing, extra or rearranged gene fragments in the wrong order. As such, these advanced cloning tools are not amenable to refine the regulatory control elements of each of the expression cassettes if identical regulatory sequences (such as promoters, operators, ribosome binding sites, and terminators) are used [18]. In addition, due to the predetermined fragment context specified by the overlap sequence, these gene assembly tools are not scalable for diversifying pathway configurations and shuffling gene orders in the final construct [19]. As a result, synthetic biology calls for new gene assembly tools adept at hierarchical pathway construction and generation of pathway diversities in a combinatorial manner.

Assembly of DNA fragment based on BioBrick paradigm has become a standard practice in metabolic pathway construction [20, 21] and genetic circuit design [22]. The BioBrick™ standards take advantage of the isocaudamer pairs *Xba*I and *Spe*I, which generate compatible cohesive ends and upon ligation result in a scar sequence which cannot be cleaved by either of the original restriction enzyme pairs. Basic molecular components flanked with these isocaudamer pairs can be assembled together by iterative restriction enzyme digestion and ligation. Recently, by expanding the BioBrick paradigm, we have developed a versatile gene assembly platform ePathBrick that can be used to streamline the process of pathway construction and optimization [19]. With four compatible restriction enzyme sites (*Avr*II, *Xba*I, *Spe*I, and *Nhe*I) allocated on strategic positions, the engineered ePathBrick vectors (Fig. 2) support the modular assembly of a number of molecular components including regulatory control elements (promoters, operators, ribosome binding sites, and terminators) and multi-gene pathways; in addition, ePathBricks provide a platform for combinatorial generation of pathway diversities with three distinct configurations. For example, for a multi-gene pathway, each of the pathway components can be organized either in operon, pseudo-operon, or monocistronic form (Fig. 3) when using different isocaudamer pairs.

In this protocol, we will demonstrate the assembly of a seven gene pathway (~9 kb) on one single ePathBrick vector. A total size of 36 kb pathways could be assembled using four compatible ePathBrick vectors, which is sufficient for expressing a typical plant secondary metabolic pathway. We will also demonstrate that

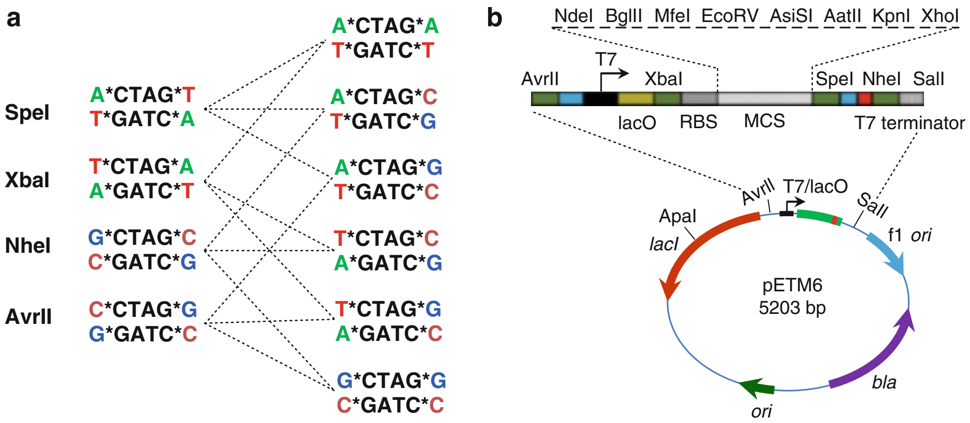


Fig. 2 (**a**) Compatible sticky ends can be generated by four isocaudamer pairs (*Avr*II, *Nhe*I, *Spe*I, and *Xba*I) with unique DNA recognition sequence. The resulting sticky ends can be rejoined together upon simple ligation thereby nullifying the old restriction sites. (**b**) The basic configuration of engineered ePathBrick vectors. Four compatible restriction enzymes are allocated on strategic positions so that regulatory control elements and expression cassettes can be reused

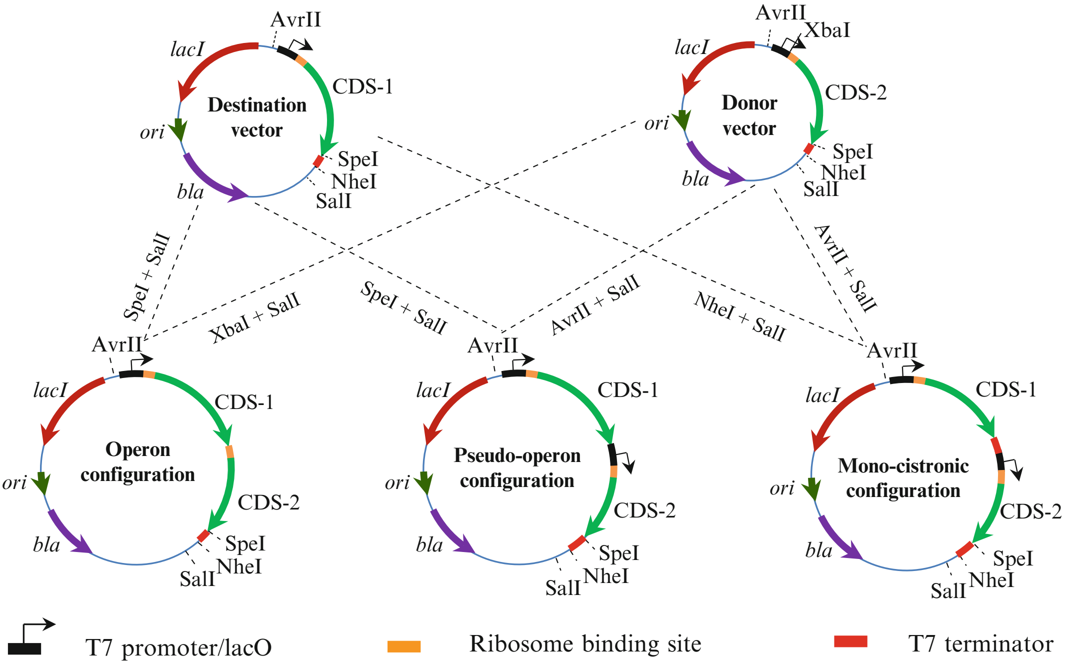


Fig. 3 ePathBrick vectors provide a versatile platform to combinatorially assemble multi-gene pathways with different configuration. Operon configuration can be achieved by digesting the donor vector with restriction enzyme pairs *Xba*I/*Sal*I and ligating it to the *Spe*I/*Sal*I digested destination vector. Pseudo-operon configuration can be achieved by digesting the donor vector with restriction enzyme pairs *Avr*II/*Sal*I and ligating it to the *Spe*I/*Sal*I digested destination vector. Monocistronic configuration can be achieved by digesting the donor vector with restriction enzyme pairs *Avr*II/*Sal*I and ligating it to the *Nhe*I/*Sal*I digested destination vector

a three-gene flavonoid pathway can be easily diversified to 54 pathway equivalents differing in pathway configuration and gene order; coupled with high-throughput screening techniques, we envision that this combinatorial strategy would greatly improve our ability to exploit the full potential of microbial cell factories for recombinant metabolite production. The ePathBrick assembly, though limited by one gene at a time, provides a versatile platform to combinatorially construct and optimize multi-gene pathways—a feature that will be proven extremely useful to pathway engineering.

2 Materials

2.1 Reagents and Molecular Kits

1. Certified™ Molecular Biology Agarose, Biorad, Cat. No. 161-3102.
2. Ampicillin Sodium Salt, Sigma-Aldrich, Cat. No. A9518-25G.
3. S.O.C. media, Invitrogen, Cat. No. 15544-034.
4. 10 mg/mL ethidium bromide and 50× TAE buffer, Biorad.
5. LB agar and LB broth, Sigma-Aldrich.
6. Phusion® DNA polymerase and buffer, New England Biolabs, Cat. No. 0530L.
7. T4 DNA ligase and buffer, New England Biolabs, Cat. No. M0202L.
8. Fermentas Fast Digest Enzymes *DpnI*, *Nde*I, *Kpn*I, *Xho*I, *Avr*II, *Xba*I, *Spe*I, *Nhe*I, *Sal*I, and *Apa*I.
9. QIAprep® Spin Miniprep kit, QIAGEN, Cat. No. 27106.
10. QIAquick® Gel Extraction kit, QIAGEN, Cat. No. 28706.
11. PureLink™ Genomic DNA Mini Kit, Invitrogen, Cat. No. K1820-02.
12. QuikChange® II Site-Directed Mutagenesis Kit, Agilent Technologies, Cat. No. 200521.
13. CCB1 buffer: Mix 5.15 g of $CaCl_2$ in 200 mL deionized (DI) H_2O. Autoclave for 20 min at 121 °C. Cool and add 2.5 mL of 1 M $MgSO_4$ solution. Store at 4 °C.
14. CCB2 buffer: Mix 2.06 g of $CaCl_2$ in 128 mL of DI H_2O and 72 mL of glycerol. Autoclave for 20 min at 121 °C. Cool and add 1 mL of 1 M $MgSO_4$. Store in 4 °C.

2.2 Strains, Plasmids, and Primers

Please refer to Tables 1 and 2 for strains, plasmids, and primers used in this protocol.

Table 1
Strains and plasmids in this protocol

Plasmid or strain	Relevant properties or genotype	Source or reference
Plasmid		
pETM5	ColE1(pBR322), Amp^r, with ePathBrick feature (without SpeI)	[19]
pETM6	ColE1(pBR322), Amp^r, with ePathBrick feature (with SpeI)	[19]
pET-PhCHS-MsCHI	pETDuet-1 carrying *chs* from *P. hybrida*, *chi* from *Medicago sativa*	[23]
pCDF-4CL2	pCDFDuet-1 carrying *4cl-2* from *Petroselinum crispum*	[23]
pETM6-Pc4CL-oPhCHS-oMsCHI	pETM6 carrying *4cl-2* from *P. crispum*, *chs* from *P. hybrida*, *chi* from *Medicago sativa*, CHS, and CHI were organized in operon form	This study
pETM6-Pc4CL-pPhCHS-pMsCHI	pETM6 carrying *4cl-2* from *P. crispum*, *chs* from *P. hybrida*, *chi* from *Medicago sativa*, CHS, and CHI were organized in pseudo-operon form	This study
pETM6-Pc4CL-mPhCHS-mMsCHI	pETM6 carrying *4cl-2* from *P. crispum*, *chs* from *P. hybrida*, *chi* from *Medicago sativa*, CHS, and CHI were organized in operon form	This study
pETM6-Flavonoid pathway library	pETM6 carrying plant-derived three-gene (*4cl-2*, *chs,* and *chi*) flavonoid pathway libraries, differing in gene configuration and gene order	This study
pETM6-Pc4CL-PhCHS-MsCHI-accDCBA	pETM6 carrying *4cl-2* from *P. crispum*, *chs* from *P. hybrida*, *chi* from *Medicago sativa*, and four subunits accDCBA from *E. coli*	This study
Strains		
E. coli DH5α	F$^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ^-	Koffas stock
E. coli K-12	Type strain	Koffas stock
E. coli BW 27784	F-Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ^-, Δ(araH-araF)570(::FRT), ΔaraEp-532::FRT, φP_{cp18}araE533, Δ(rhaD-rhaB)568, hsdR514	Yale, CGSC

Table 2
Primers used in this protocol

No.	Primer name	Nucleotide sequence (5′ > 3′)
1	4Cl_NdeF	GAACATATGGGAGACTGTGTAGCACCC
2	4Cl_KpnR	GACGGTACCCTATTATTTGGGAAGATCACCG
3	CHS_NdeF	GAGCATATGGTGACAGTCGAGGAGTATC
4	CHS_KpnR	GGAGGTACCCTATTAAGTAGCAACACTGTGG
5	CHI_NdeF	GGACATATGGCTGCATCAATCACCGCAATC
6	CHI_KpnR	GGCGGTACCCTATCAGTTTCCAATCTTGAAAG
7	CHS_SpeF	GTCTTTTGCACAACCAGTGGTGTGGACATG
8	CHS_SpeR	CATGTCCACACCACTGGTTGTGCAAAAGAC
9	AccA_NdeF	GGGCGCATATGAGTCTGAATTTCCTTG
10	AccA_XhoR	ATGGGCTCGAGTTACGCGTAACCGTAGCTC
11	AccB_NdeF	CGGGGCATATGGATATTCGTAAG
12	AccB_XhoR	GGGCTCTCGAGTTACTCGATGACGACC
13	AccC_NdeF	GCGCCCATATGCTGGATAAAATTG
14	AccC_XhoR	GGCGGCTCGAGTTATTTTTCCTGAAGACC
15	AccD_NdeF	GGCTCCATATGAGCTGGATTGAACG
16	AccD_XhoR	GGATTCTCGAGTCAGGCCTCAGGTTCCTG

3 Methods

3.1 Preparation of DH5α and BW27784 Chemically Competent Cell

1. Inoculate *E. coli* DH5α and BW27784 strain in 3 mL of LB in 15 mL Corning tubes and incubate at 37 °C and 250 rpm overnight.
2. Measure overnight culture OD and make a dilution in 40 mL of LB in sterile flask so that initial OD is about 0.05.
3. Incubate at 37 °C until OD is about 0.6 (about 1.5–2.5 h depending on cell strain).
4. Transfer the cell culture to 50 mL Corning tubes and cool in ice for 10 min. Meanwhile cool centrifuge to 4 °C.
5. Centrifuge for 10 min at 3,400 × g and 4 °C.
6. Drain supernatant and gently resuspend cell pellet with 20 mL of ice cold CCB1 and incubate on ice for another 15 min.
7. Repeat **steps** 7 and **8** but resuspend cell pellet with 800 μL of ice cold CCB1.

8. Incubate cells in CCB1 for 2 h (overnight preferred) on ice in 4 °C fridge.
9. Gently mix with 1:1 (800 μL) of CCB2 and aliquot into the micro-centrifuge tubes (100 μL competent cell each vial). Instantaneously freeze the cell using dry ice and store vials at −80 °C freezer.

3.2 Heat Shock Transformation

1. Thaw chemically competent cells on ice.
2. Add up to 0.5 μg of DNA (1–5 μL ligation reaction) to 50–100 μL chemically competent cells; tap tube gently to mix.
3. Incubate on ice for at least 15 min. Meanwhile prepare a 42 °C heating block with water bath.
4. Heat shock cells for 30–45 s in 42 °C water bath.
5. After heat shock, return vial to ice bath immediately and let cell sit on ice for 2 min.
6. Add 3× volume of room temperature SOC media (~200 μL).
7. Cap tightly and shake horizontally for 1 h at 37 °C and 250 rpm.
8. Plate ~200 μL cell culture onto selective plates with appropriate antibiotics.
9. Incubate plates for 18 h at 37 °C.
10. Pick up colony and inoculate into 4 mL LB supplemented with appropriate antibiotics.

3.3 Preparation of PCR Fragments

1. Inoculate *E. coli* DH5α harboring plasmid pET-PhCHS-MsCHI into 4 mL LB supplemented with 100 μg/mL of ampicillin, grow in 15 mL Corning tubes at 37 °C and 250 rpm for 16 h in shaker incubator.
2. Inoculate *E. coli* DH5α harboring plasmid pCDF-4CL2 into 4 mL LB supplemented with 50 μg/mL of streptomycin, grow in 15 mL Corning tubes at 37 °C and 250 rpm for 16 h in shaker incubator.
3. Inoculate *E. coli* K-12 into 4 mL LB, grow in 15 mL Corning tubes at 37 °C and 250 rpm for 16 h in shaker incubator.
4. Centrifuge the above cell cultures at 3,400 × g for 15 min to collect cell pellets.
5. Mini-preparation of plasmids pET-PhCHS-MsCHI and pCDF-4CL2 using QIAprep® Spin Miniprep kit. Elute each plasmid with 25 μL sterilized MilliQ water. Repeat the elution process two times. Store plasmids at −20 °C freezer for future use.
6. Mini-preparation of *E. coli* K-12 genomic DNA using Pure-Link™ Genomic DNA Mini Kit. Elute K-12 genomic DNA

with 25 μL sterilized MilliQ water. Repeat the elution process two times. Store genomic DNA at −20 °C freezer for future use.

7. Set up PCR reactions to amplify pc4CL2 gene from pCDF-4CL2. Add the following components to a PCR tube and mix:

Components	Volume (for 40 μL reaction)
10 μM forward primer accA_NdeF	1 μL
10 μM reverse primer accA_XhoR	1 μL
10 mM dNTPs	0.8 μL
5× Phusion® HF buffer	8 μL
Template (K-12 genomic DNA)	0.8 μL
Phusion® DNA polymerase	0.4 μL
MilliQ H_2O	28 μL

8. Set up PCR reactions to amplify PhCHS and MsCHI genes from pET-PhCHS-MsCHI. Repeat **step 7** but use pET-PhCHS-MsCHI as template and CHS_NdeF/CHS_KpnR and CHI_NdeF/CHI_KpnR as primers, respectively (*see* **Note 1**).

9. Set up PCR reactions to amplify accA gene from *E. coli* K-12 genomic DNA. Add the following components to a PCR tube and mix:

Components	Volume (for 40 μL reaction)
10 μM forward primer accA_NdeF	1 μL
10 μM reverse primer accA_XhoR	1 μL
10 mM dNTPs	0.8 μL
5× Phusion® HF buffer	8 μL
Template (K-12 genomic DNA)	0.8 μL
Phusion® DNA polymerase	0.4 μL
MilliQ H_2O	28 μL

10. Set up PCR reactions to amplify *accB*, *accC*, and *accD* genes from K-12 genomic DNA. Repeat **step 9** but use accB_NdeF/accB_XhoR, accC_NdeF/accC_XhoR, and accD_NdeF/accD_XhoR as primers, respectively.

11. Perform PCR reactions using the following parameters:

Steps	Cycles	Temperature (°C)	Time (mm:ss)
Initial denaturation	1	98	00:45 (plasmid) or 01:30 (genomic DNA)
Denaturation	25–30	98	00:10
Annealing		62–64	00:25
Extension		72	00:30 per kb
Additional extension	1	72	5–8 min
Storage	1	16	As needed

3.4 Assembly of Three Gene Flavonoid Pathway

1. Digestion of PCR products PhCHS, MsCHI, Pc4CL2, and plasmid pETM6 with Fast digestion restriction enzymes NdeI and KpnI. For 20 μL reaction system, use 4 μL PCR products or plasmid, 1 μL NdeI, 1 μL KpnI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
2. Run gel to check and purify digested PCR products and vectors. Load the entire 20 μL digestion reaction and run gel using 1× TAE buffer at 100 V for 30 min. Transilluminate digested DNA fragments under UV lights and excise the gel area that contains the corresponding gene fragment. Purify all the digested DNA fragments using QIAquick® Gel Extraction kit. Elute gene fragments with 15 μL sterilized MilliQ water. Repeat the elution process two times. Digested gene fragments can be stored at −20 °C for future use.
3. Ligate PhCHS, MsCHI, and Pc4CL2 gene fragments to digested vector pETM6. For 20 μL reaction volume, use 4 μL digested vector, 6 μL insert, 2 μL 10× T4 DNA ligase buffer, 1 μL T4 DNA ligase, and 7 μL sterilized MilliQ water. Keep the reaction at 25 °C for at least 1 h (*see* **Note 2**).
4. Heat shock transformation of 5 μL ligation reaction into 100 μL DH5α chemically competent cells following the protocol described in Subheading 3.2. Recover the cells in SOC media for 1 h in a 37 °C shaking incubator. Plate ~200 μL cell culture onto LB plates supplemented with 100 μg/mL ampicillin.
5. Incubate plates for 18 h at 37 °C.
6. Pick up individual colonies and inoculate into 4 mL LB supplemented with 100 μg/mL ampicillin. Grow at 37 °C with shaking at 250 rpm overnight (*see* **Note 3**).

7. Mini-preparation of plasmids from the overnight culture using QIAprep® Spin Miniprep kit. Elute each plasmid using 50 μL MilliQ water.
8. Screen positive transformants by double digestion. For a 20 μL reaction system, use 3 μL plasmid DNA, 1 μL ApaI, 1 μL KpnI, 2 μL 10× Fast digestion green buffer, and 13 μL MilliQ water. Keep reaction at 37 °C for 30 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
9. Run gel to check digestion results. Select plasmid samples that contain correct insert size for further gene assembly. Positive clones are named pETM6-Pc4CL, pETM6-PhCHS, and pETM6-MsCHI, respectively (*see* **Note 4**).
10. PhCHS fragment obtained by digestion of pETM6-PhCHS with AvrII and SalI was gel purified and ligated to the NheI and SalI-digested pETM6-Pc4CL to give construct pETM6-Pc4CL-PhCHS following standard cloning procedures (**step 1** through **step 9** in this section).
11. MsCHI fragment obtained by digestion of pETM6-MsCHI with AvrII and SalI was gel purified and ligated to the NheI and SalI-digested pETM6-Pc4CL-PhCHS to give construct pETM6-Pc4CL-PhCHS-MsCHI following standard cloning protocol (**step 1** through **step 9** in this section).

3.5 Assembly of Four Gene Acetyl-CoA Carboxylase

1. Digest PCR products *accA*, *accB*, *accC*, *accD*, and plasmid pETM6 with Fast digestion restriction enzymes NdeI and XhoI. For 20 μL reaction system, use 4 μL PCR products or plasmid, 1 μL NdeI, 1 μL XhoI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
2. Run gel to check and purify digested PCR products and vectors. Load all the 20 μL digestion reaction and run gel using 1× TAE buffer at 100 V for 30 min. Transilluminate digested DNA fragments under UV light and excise the gel area that contains the corresponding gene fragment. Purify all the digested DNA fragments using QIAquick® Gel Extraction kit. Elute gene fragments with 15 μL sterilized MilliQ water. Repeat the elution process two times. Digested gene fragments can be stored at −20 °C for future use.
3. Ligate *accA*, *accB*, *accC*, and *accD* gene fragments to digested vector pETM6. For 20 μL reaction system, use 4 μL digested vector, 6 μL insert, 2 μL 10× T4 DNA ligase buffer, 1 μL T4 DNA ligase, and 7 μL sterilized MilliQ water. Keep the reaction at 25 °C for at least 1 h.

4. Heat shock transformation of 5 μL ligation reaction into 100 μL DH5α chemically competent cells following the protocol described in Subheading 3.2. Recover the cells in SOC media for 1 h in a 37 °C shaking incubator. Plate ~200 μL cell culture onto LB plates supplemented with 100 μg/mL ampicillin.
5. Incubate plates for 18 h at 37 °C.
6. Pick single colonies and inoculate into 4 mL LB supplemented with 100 μg/mL ampicillin. Grow at 37 °C with shaking at 250 rpm overnight.
7. Mini-preparation of plasmids from the overnight culture using QIAprep® Spin Miniprep kit. Elute each plasmid with 50 μL of MilliQ water.
8. Screen positive transformants by double digestion. For a 20 μL reaction system, use 3 μL plasmid DNA, 1 μL ApaI, 1 μL XhoI, 2 μL 10× Fast digestion green buffer, and 13 μL MilliQ water. Keep reaction at 37 °C for 30 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
9. Run gel to check digestion results. Select plasmid samples that contain correct insert size for further gene assembly. Positive clones are named pETM6-accA, pETM6-accB, pETM6-accC, and pETM6-accD, respectively.
10. *AccC* fragment obtained by digestion of pETM6-accC with AvrII and SalI was gel purified and ligated to the NheI and SalI digested pETM6-accD to give construct pETM6-accD-accC following standard cloning protocol (**step 1** through **step 9** in this section).
11. *AccA* fragment obtained by digestion of pETM6-accA with AvrII and SalI was gel purified and ligated to the NheI and SalI digested pETM6-accB to give construct pETM6-accB-accA following standard cloning protocol (**step 1** through **step 9** in this section).
12. *AccB-accA* fragment obtained by digestion of pETM6-accB-accA with AvrII and SalI was gel purified and ligated to the NheI and SalI digested pETM6-accD-accC to give construct pETM6-accD-accC-accB-accA following standard cloning protocol (**step 1** through **step 9** in this section).

3.6 Assembly and Verification of Seven-Gene Flavonoid Biosynthetic Pathway

1. Digestion of pETM6-accD-accC-accB-accA with restriction enzymes AvrII and ApaI (Fig. 4). For a 20 μL reaction, add 4 μL plasmid, 1 μL AvrII, 1 μL ApaI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.

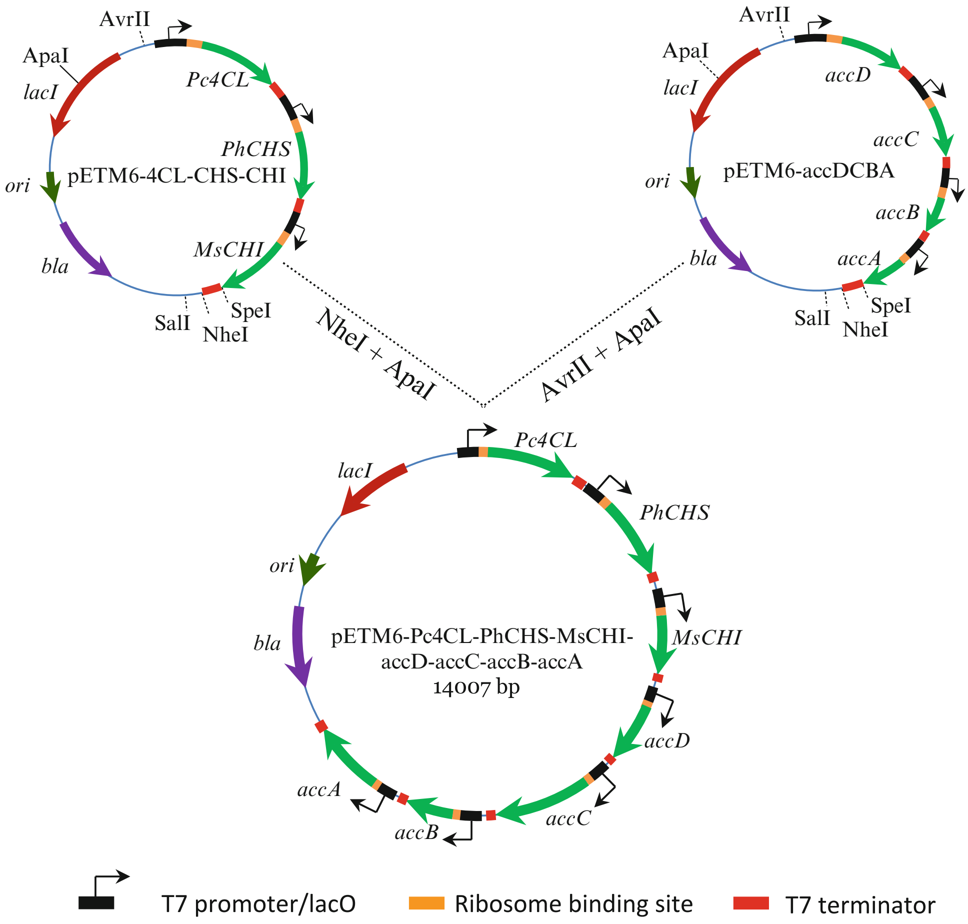


Fig. 4 Assembly of seven gene flavonoid biosynthetic pathway

2. Digestion of pETM6-Pc4CL-PhCHS-MsCHI with restriction enzymes NheI and ApaI (Fig. 4). For a 20 μL reaction, add 4 μL plasmid, 1 μL NheI, 1 μL ApaI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
3. Run gel to check digestion results and purify the digested DNA fragments. Recovered DNA fragments are Pc4CL-PhCHS-MsCHI (5,250 bp) and *accDCBA* (8,757 bp).
4. Ligate Pc4CL-PhCHS-MsCHI gene fragment to ApaI and AvrII digested pETM6-accDCBA (Fig. 4). For 20 μL reaction system, use 4 μL digested Pc4CL-PhCHS-MsCHI, 4 μL digested accDCBA, 2 μL 10× T4 DNA ligase buffer, 2 μL T4 DNA ligase, and 8 μL sterilized MilliQ water. Keep the reaction at 25 °C for at least 1 h (*see* **Note 5**).

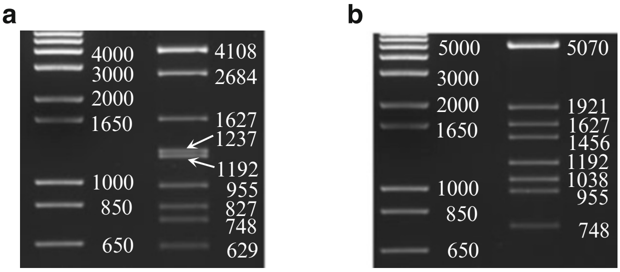


Fig. 5 (**a**) Digestion pattern for seven-gene constructs pETM6-Pc4CL-PhCHS-MsCHI-accDCBA with *Apa*I and *Xho*I. *First lane*, 1 kb plus ladder; *second lane*, digested plasmid (*note*: CHS gene contains an intergenic *Xho*I site). (**b**) Digestion pattern for seven-gene construct pETM6-Pc4CL-PhCHS-MsCHI-accDCBA with *Apa*I and *Nde*I. *First lane*, 1 kb plus ladder; *second lane*, digested plasmid

5. Heat shock transformation of 5 μL ligation reaction into 100 μL DH5α chemically competent cells following the protocol as described in Subheading 3.2. Recover the cells in SOC media for 1 h in a 37 °C shaking incubator. Plate ~200 μL cell culture onto LB plates supplemented with 100 μg/mL ampicillin.
6. Pick single colonies and inoculate to 4 mL LB (with 100 μg/mL ampicillin) and grow at 37 °C overnight.
7. Mini-preparation of plasmids from the overnight culture using QIAprep® Spin Miniprep kit. Elute each plasmid using 50 μL MilliQ water.
8. Screen positive transformants by double digestion. Appropriate digestion sites can be chosen by Vector NTI to facilitate the screening process. For a 20 μL reaction system, use 3 μL plasmid DNA, 1 μL ApaI, 1 μL XhoI or 1 μL NdeI, 2 μL 10× Fast digestion green buffer, and 13 μL MilliQ water. Keep reaction at 37 °C for 45 min. At the same time, prepare 1.0 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
9. Run gel to check digestion results. Select plasmid samples that contain all the correct insert sizes. For positive clones, nine fragments (4,108, 2,684, 1,627, 1,237, 1,192, 955, 827, 748, and 629 bp) should be obtained if the plasmids were digested with ApaI and XhoI (Fig. 5a); eight fragments (5,070, 1,921, 1,627, 1,456, 1,192, 1,038, 955, and 748 bp) should be obtained if the plasmids were digested with ApaI and NdeI (Fig. 5b). Positive clones are named pETM6-Pc4CL-PhCHS-MsCHI-accD-accC-accB-accA.

3.7 Elimination of SpeI Site on PhCHS (See Note 6)

1. Silent mutation of SpeI site inside PhCHS gene using QuikChange® II Site-Directed Mutagenesis Kit. Add the following components to a PCR tube and mix:

Components	Volume (for 30 μL reaction)
10 μM forward primer CHS_SpeF	1.2 μL
10 μM reverse primer CHS_SpeR	1.2 μL
Template plasmid pETM5-PhCHS	1.2 μL
10× Pfu high fidelity buffer	3.0 μL
10 mM dNTPs	0.6 μL
Pfu ultra DNA polymerase	0.6 μL
MilliQ water	22.2 μL

2. Perform PCR reactions using the following parameters:

Steps	Cycles	Temperature (°C)	Time (mm:ss)
Initial denaturation	1	95	01:00
Denaturation	16	95	01:30
Annealing		62	00:45
Extension		72	07:00 (1 kb per minute)
Additional extension	1	72	10 min
Storage	1	16	As needed

3. Digest PCR products with DpnI to remove template plasmid pETM5-PhCHS. For a 20 μL reaction system, use 16.5 μL PCR products, 2 μL 10× Fast digestion clear buffer, and 1.5 μL DpnI. Keep reactions at 37 °C for 40 min.
4. Heat shock transformation of 8 μL DpnI digested PCR products into 100 μL chemically competent BW27784 cells following the protocol as described in Subheading 3.2. Recover the cells in SOC media for 1 h in a 37 °C shaking incubator. Plate ~200 μL cell culture onto LB plates supplemented with 100 μg/mL ampicillin.
5. Pick single colonies and inoculate to 4 mL LB (with 100 μg/mL ampicillin) and grow at 37 °C overnight.
6. Mini-preparation of plasmids from the overnight culture using QIAprep® Spin Miniprep kit. Elute each plasmid using 50 μL MilliQ water.
7. Screen positive transformants by double digestion. For a 20 μL reaction system, add 2 μL plasmid, 1 μL ApaI, 1 μL SpeI, 2 μL 10× Fast digestion green buffer, and 14 μL MilliQ water. Keep reactions at 37 °C for 30 min. Use original pETM5-PhCHS as

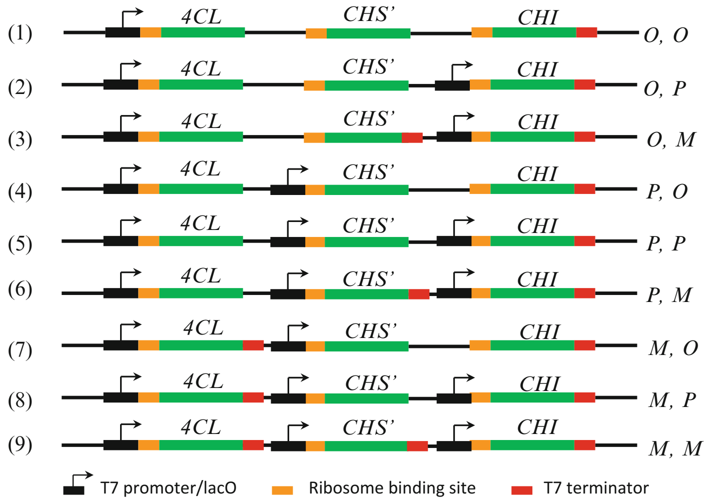


Fig. 6 A three gene pathway could take nine functional configurations. A three-gene pathway has two connection nodes and each node could take three different configurations. *O* denotes operon, *P* denotes pseudo-operon, and *M* denotes monocistronic configuration

negative control. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.

8. Run gel to check digestion results. Positive mutations are plasmid samples that are only cut by ApaI resulting in a single band at 6,327 bp; negative mutations are plasmid samples that are cut by both ApaI and SpeI resulting in two bands at 4,894 bp and 1,433 bp, respectively.
9. SpeI mutated plasmid samples was renamed pETM5-PhCHS' and was kept at −20 °C for future use.

3.8 Combinatorial Pathway Assembly: Adding the Second Gene

A three gene pathway could take nine functional configurations (Fig. 6). Using pETM6-PhCHS' as a donor vector, the second gene PhCHS' can be assembled in three distinctive configurations. In this section, we will show the details of how we diversify a three-gene-pathway by combinatorial assembly (*see* **Note 7**).

1. Double digestion of destination vector pETM6-Pc4CL with SpeI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-Pc4CL, 1 μL SpeI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
2. Double digest destination vector pETM6-Pc4CL and pETM5-Pc4CL with NheI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-Pc4CL, 1 μL NheI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.

3. Double digestion of donor vector pETM6-PhCHS’ with XbaI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-PhCHS’, 1 μL XbaI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
4. Double digestion of donor vector pETM6-PhCHS’ with AvrII and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-PhCHS’, 1 μL AvrII, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
5. Run gel to check digestion results and purify all the digested DNA fragments using QIAquick® Gel Extraction kit. Each of the digestion reaction gives two DNA fragments. Recover the bigger fragment for destination vector and the smaller fragment for donor vector. Elute gene fragments with 15 μL sterilized MilliQ water. Repeat the elution process two times. Digested gene fragments can be stored at −20 °C for future use.
6. To construct operon configuration, ligate XbaI and SalI digested donor vector (pETM6-PhCHS’) to the SpeI and SalI digested destination vector (pETM6-Pc4CL). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase, and 13 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h.
7. To construct pseudo-operon configuration, ligate AvrII and SalI digested donor vector (pETM6-PhCHS’) to the SpeI and SalI digested destination vector (pETM6-Pc4CL). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase and 13 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h.
8. To construct monocistronic configuration, ligate AvrII and SalI digested donor vector (pETM6-PhCHS’) to the NheI and SalI digested destination vector (pETM5-Pc4CL). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase and 7 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h. *Note here the destination vector is pETM5-Pc4CL instead of pETM6-Pc4CL* (*see* **Note 8**).
9. Heat shock transformation of 5 μL ligation reaction into 100 μL chemically competent DH5α cell following the protocol as described in Subheading 3.2. Recover the cells in SOC media for 1 h in a 37 °C shaking incubator. Plate ~200 μL cell culture onto LB plates supplemented with 100 μg/mL ampicillin.

10. Pick single colonies and inoculate to 4 mL LB (with 100 μg/mL ampicillin) and grow at 37 °C overnight.
11. Mini-preparation of plasmids from the overnight culture using QIAprep® Spin Miniprep kit. Elute each plasmid by 50 μL MilliQ water.
12. Screen positive transformants by double digestion. For a 20 μL reaction system, add 3 μL plasmid, 1 μL ApaI, 1 μL KpnI, 2 μL 10× Fast digestion green buffer, and 13 μL MilliQ water. Keep reactions at 37° for 30 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.
13. Run gel to check digestion results. Perform in silico digestion analysis using Vector NTI and select plasmid samples that contain the correct insert size. Positive clones are named pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM6-Pc4CL-mPhCHS', respectively. All these intermediary constructs still contain unique restriction sites AvrII, SpeI, NheI, and SalI, which can be used for next-round gene assembly.

3.9 Combinatorial Pathway Assembly: Adding the Third Gene

A three gene pathway could take nine functional configurations (Fig. 6). Using pETM6-MsCHI as donor vector, the third gene MsCHI can be assembled in three distinctive configurations. In this section, we will show the details of how we diversify a three-gene-pathway by combinatorial assembly.

1. Double digestion of destination vector pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM5-Pc4CL-mPhCHS' with SpeI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-Pc4CL, 1 μL SpeI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
2. Double digestion of destination vector pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM6-Pc4CL-mPhCHS' with NheI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-Pc4CL, 1 μL NheI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
3. Double digestion of donor vector pETM6-MsCHI with XbaI and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-MsCHI, 1 μL XbaI, 1 μL SalI, 2 μL 10× Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at 37 °C for 45 min.
4. Double digestion of donor vector pETM6-MsCHI with AvrII and SalI. For a 20 μL reaction system, use 4 μL plasmid pETM6-MsCHI, 1 μL AvrII, 1 μL SalI, 2 μL 10 × Fast digestion green buffer, and 12 μL MilliQ water. Keep reactions at

37 °C for 45 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.

5. Run gel to check digestion results and purify all the digested DNA fragments using QIAquick® Gel Extraction kit. Each of the digestion reactions gives two DNA fragments. Recover the bigger fragment for destination vector and the smaller fragment for donor vector. Elute gene fragments with 15 μL sterilized MilliQ water. Repeat the elution process two times. Digested gene fragments can be stored at −20 °C for future use.

6. To assemble the third gene in operon configuration, ligate XbaI and SalI digested donor vector (pETM6-MsCHI) to the SpeI and SalI digested destination vectors (pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM6-Pc4CL-mPhCHS'). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase, and 13 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h.

7. To assemble the third gene in pseudo-operon configuration, ligate AvrII and SalI digested donor vector (pETM6-MsCHI) to the SpeI and SalI digested destination vectors (pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM6-Pc4CL-mPhCHS'). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase, and 13 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h.

8. To assemble the third gene in monocistronic configuration, ligate AvrII and SalI digested donor vector (pETM6-MsCHI) to the NheI and SalI digested destination vectors (pETM6-Pc4CL-oPhCHS', pETM6-Pc4CL-pPhCHS', and pETM6-Pc4CL-mPhCHS'). For a 20 μL reaction system, use 5 μL digested donor vector, 5 μL digested destination vector, 2 μL 10× Fast digestion green buffer, 1 μL T4 DNA ligase, and 13 μL MilliQ water. Keep the ligation reaction at 25 °C for 1.5 h.

9. Repeat **steps 9–12** as described in Subheading 3.8 to perform heat shock transformation and double digestion screening.

10. Run gel to check digestion results. Perform in silico digestion analysis using Vector NTI and select plasmid samples that contain all the correct insert size. Positive clones are named pETM6-Pc4CL-oPhCHS'-oMsCHI, pETM6-Pc4CL-oPhCHS'-pMsCHI, pE TM6-Pc4CL-oPhCHS'-mMsCHI; pETM6-Pc4CL-pPhCHS'-oMsCHI, pETM6-Pc4CL-pPhC HS'-pMsCHI, pETM6-Pc4CL-pPhCHS'-mMsCHI; and pETM6-Pc4CL-mPhCHS'-oMsCHI, pETM6-Pc4CL-mPhC HS'-pMsCHI, pETM6-Pc4CL-mPhCHS'-mMsCHI, respectively (Fig. 6).

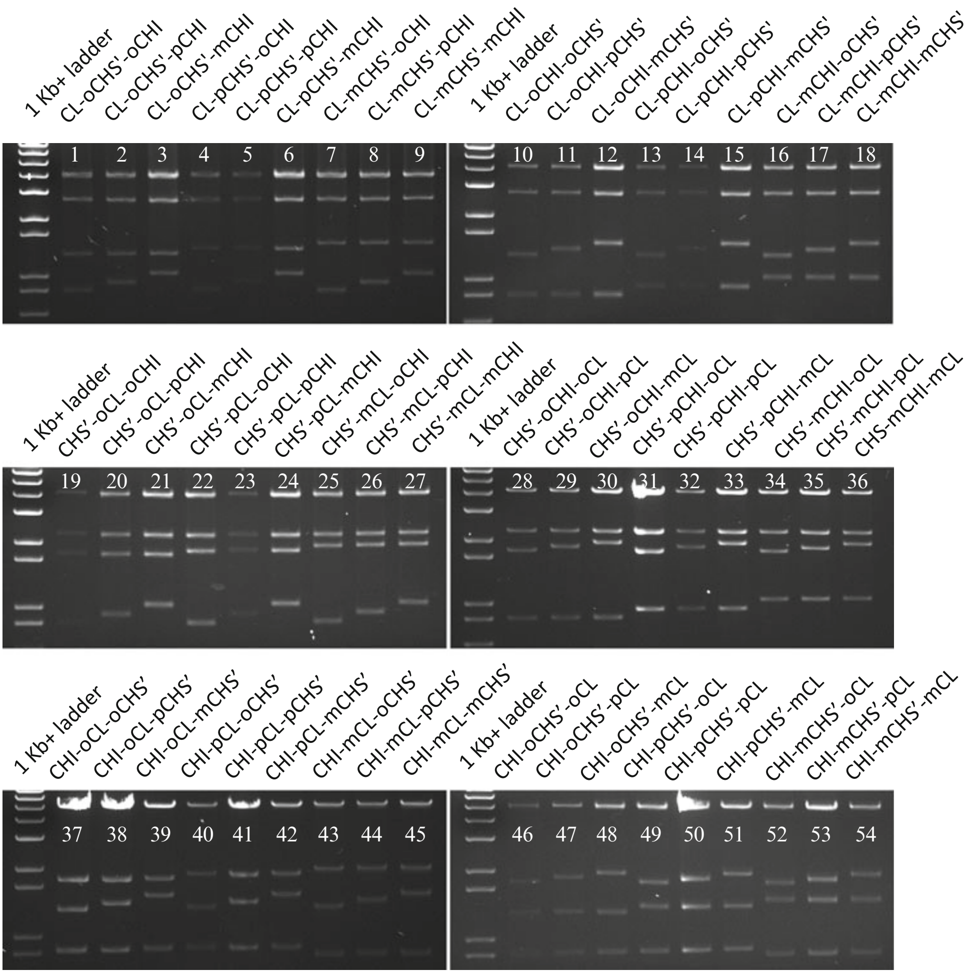


Fig. 7 Digestion pattern of combinatorial pathway libraries 1–54. Pathway 1–36 were digested with ApaI and KpnI and pathway 37–54 were digested with AvrII and KpnI

3.10 Shuffling Gene Order and Verification of Combinatorial Flavonoid Pathway Libraries

1. A three-gene pathway has six different orders (permutation of three equals six). For example, the above-mentioned three gene flavonoid pathway could be arranged to six different forms: 4CL-CHS'-CHI, 4CL-CHI-CHS', CHS'-4CL-CHI, CHS'-CHI-4CL, CHI-4CL-CHS', and CHI-CHS'-4CL. Considering gene order and configuration, a three gene flavonoid pathway could be combinatorially diversified to 54 equivalent pathway libraries (Fig. 7).
2. For each of the possible gene orders, repeat Subheading 3.9 and 3.10 to further diversify the flavonoid pathway library.
3. Screen positive transformants by double digestion. For a 20 μL reaction system, add 3 μL plasmid, 1 μL ApaI or AvrII, 1 μL KpnI, 2 μL 10× Fast digestion green buffer, and 13 μL MilliQ

water. Keep reactions at 37 °C for 45 min. At the same time, prepare 0.8 % agarose gel supplemented with 0.5 μg/mL ethidium bromide.

4. Run gel to check digestion results. Perform in silico digestion analysis using Vector NTI and select plasmid samples that contain all the correct insert size (*see* **Note 8, 9**). A ladder digestion pattern can be obtained for the operon, pseudo-operon, and monocistronic gene configuration (Fig. 7). For example, pseudo-operon configurations are 80 bases longer than operon configurations and monocistronic configurations are 96 bases longer than pseudo-operon configurations.

4 Notes

1. Studies have shown that double promoter or double terminator architecture might have a profound impact on the expression of target genes. ePathBrick vector supports the assembly of gene cassettes with multiple promoter/terminator constructs as well.
2. If no colony is formed after heat shock transformation, try overnight ligation. Alternatively, try using fresh T4 DNA ligase and fresh T4 ligase buffer.
3. For PCR-based cloning, pick 5–8 colonies for each construct when performing screening. For subcloning and gene assembly, pick 1–3 colonies for each construct when screening.
4. Post-cloning gene sequencing of the assembled fragments is not necessary since all the gene assembly is performed at subcloning level. There is little chance to introduce mutations during subcloning. However, for PCR-based cloning, we recommend sequencing each construct before proceeding to further assembly steps since PCR tends to introduce mutations.
5. It is always preferred that the insert size and vector size are comparable for large fragment assembly when performing ligation reaction. Disproportional insert and vector sizes tend to lead to ligation failure. For example, the 14 kb pETM6-Pc4CL-PhCHS-MsCHI-accDCBA was constructed by ligating AvrII and ApaI digested pETM6-accDCBA (8,757 bp) with NheI and ApaI digested pETM6-Pc4CL-PhCHS-MsCHI (5,250 bp).
6. For ePathBrick directed gene assembly, it is required that the insertion fragment is free of AvrII, XbaI, SpeI, NheI, and SalI

(or ApaI) restriction sites. If any of these sites appears in the insertion fragment, site-directed mutagenesis must be used to silently mutate these sites. In this protocol, PhCHS gene fragment contains an intergenic SpeI site; we have eliminated this site by site-directed PCR.

7. For multi-gene assembly or construction of combinatorial pathway libraries, parallel assembly is possible if arranged properly. For example, we have finished 75 constructs in three rounds of cloning in 1 week. This high assembly efficiency demonstrated the possibility to diversify natural products biosynthetic pathways in a combinatorial manner.
8. In most of the cases, the four isocaudamers AvrII, XbaI, SpeI, and NheI are still unique for each round of gene assembly. However, when using pETM6 vector to perform combinatorial assembly, if the second gene is arranged in monocistronic configuration, the SpeI site won't be unique and this will lead to assembly failure if the third gene is arranged either in operon or pseudo-operon form. A way to resolve this issue is to use pETM5 vector, free of SpeI site, to arrange the first gene. In this protocol, the donor vector for pETM6-Pc4CL-mPhCHS' is pETM5-Pc4CL instead of pETM6-Pc4CL.
9. For screening multi-gene constructs, it is preferable to choose digestion sites that can distinguish each of the insertion fragments. In most of the cases, in silico digestion tools like Vector NTI can be used to facilitate this process.

References

1. Cheng AA, Lu TK (2012) Synthetic biology: an emerging engineering discipline. Annu Rev Biomed Eng 14:155–178
2. Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 3:109–118
3. Xu P, Koffas MAG (2010) Metabolic engineering of Escherichia coli for biofuel production. Biofuels 1:493–504
4. Westfall PJ et al (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci USA 109:E111–E118
5. Ajikumar PK et al (2010) Isoprenoid pathway optimization for taxol precursor overproduction in Escherichia coli. Science 330:70–74
6. Xu P, Ranganathan S et al (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng 13:578–587
7. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
8. Jeong J-Y et al (2012) One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl Environ Microbiol 78:5440–5443
9. Gibson DG et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
10. Schmid-Burgk JL et al (2012) Rapid hierarchical assembly of medium-size DNA cassettes. Nucleic Acids Res 40:e92
11. Wang R-Y, Shi Z-Y et al (2012) Cloning large gene clusters from E. coli using in vitro single-

strand overlapping annealing. ACS Synth Biol 1:291–295

12. Shao Z, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:e16
13. Wingler LM, Cornish VW (2011) Reiterative recombination for the in vivo assembly of libraries of multigene pathways. Proc Natl Acad Sci USA 108:15135–15140
14. Gibson DG et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56
15. Noskov VN et al (2012) Assembly of large, high G + C bacterial DNA fragments in yeast. ACS Synth Biol 1:267–273
16. Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242–251
17. Hillson NJ, Rosengarten RD, Keasling JD (2011) j5 DNA assembly design automation software. ACS Synth Biol 1:14–21
18. Xu P, Bhan N, Koffas MAG (2013) Engineering plant metabolism into microbes: from systems biology to synthetic biology. Curr Opin Biotechnol 24:291–299. doi:10.1016/j.copbio.2012.08.010
19. Xu P, Vansiri A et al (2012) ePathBrick: a synthetic biology platform for engineering metabolic pathways in E. coli. ACS Synth Biol 1:256–266
20. Vick JE et al (2011) Optimized compatible set of BioBrick™ vectors for metabolic pathway engineering. Appl Microbiol Biotechnol 92:1275–1286
21. Du L, Villarreal S, Forster AC (2012) Multigene expression in vivo: supremacy of large versus small terminators for T7 RNA polymerase. Biotechnol Bioeng 109: 1043–1050
22. Norville JE et al (2010) Introduction of customized inserts for s-treamlined assembly and optimization of BioBrick synthetic genetic circuits. J Biol Eng 4:17
23. Leonard E, Lim K et al (2007) Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli. Appl Environ Microbiol 73:3877–3886

Chapter 11

Tandem Recombineering by SLIC Cloning and Cre-LoxP Fusion to Generate Multigene Expression Constructs for Protein Complex Research

Matthias Haffke, Cristina Viola, Yan Nie, and Imre Berger

Abstract

A robust protocol to generate recombinant DNA containing multigene expression cassettes by using sequence and ligation independent cloning (SLIC) followed by multiplasmid Cre-LoxP recombination in tandem for multiprotein complex research is described. The protocol includes polymerase chain reaction (PCR) amplification of the desired genes, seamless insertion into the target vector via SLIC, and Cre-LoxP recombination of specific donor and acceptor plasmid molecules, optionally in a robotic setup. This procedure, called tandem recombineering, has been implemented for multiprotein expression in *E. coli* and mammalian cells, and also for insect cells using a recombinant baculovirus.

Key words Sequence and ligation independent cloning, Cre recombinase, Cre-LoxP fusion, Multigene delivery, Multiprotein complexes, MultiBac, ACEMBL automation, Robotics

1 Introduction

High flexibility and diversity in cloning techniques are essential aspects for the creation of multigene constructs and multiprotein assemblies in synthetic biology [1]. Common techniques used to insert polymerase chain reaction (PCR) products into vectors for gene expression are restriction enzyme-dependent cloning [2], blunt end cloning [3], and Gateway cloning [4]. However, such cloning techniques possess limitations due to the requirements for specific DNA sequences and/or restriction enzyme sites and are therefore not feasible for high-throughput applications or automation. Since sequence and ligation independent cloning (SLIC) removes the requirement for specific DNA sequences and furthermore does not require restriction enzyme sites, it is more suitable for integration in an automated setup [5, 6].

In a typical SLIC reaction, the gene of interest (GOI) is amplified using primers which provide a homology sequence to the

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_11, © Springer Science+Business Media New York 2013

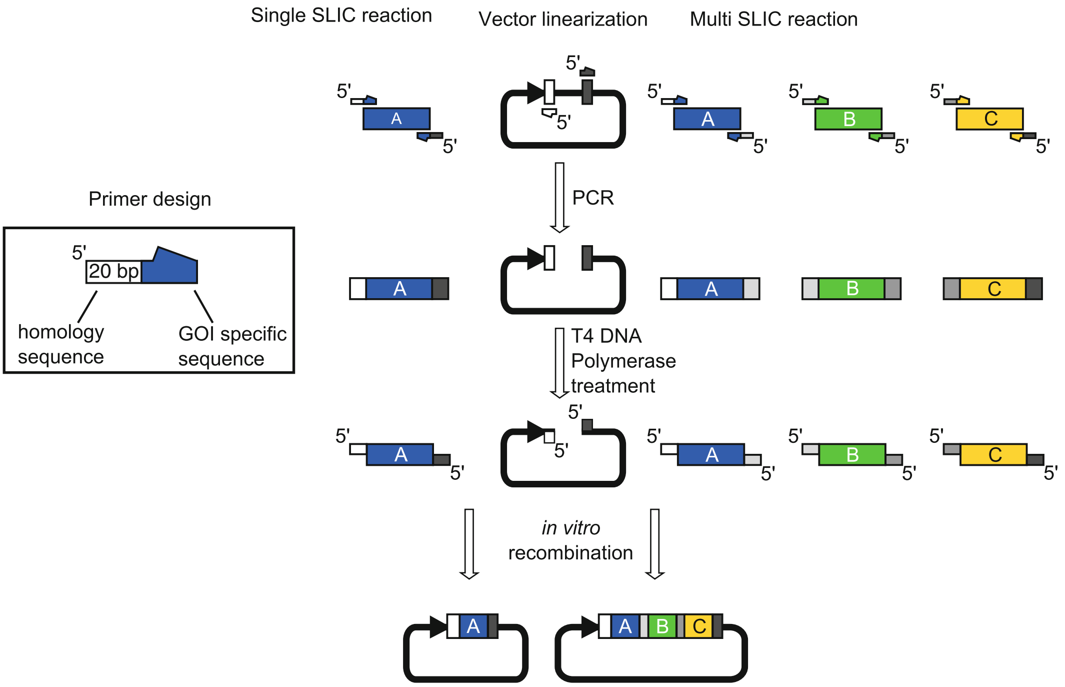


Fig. 1 Schematic overview of single- and multigene SLIC reactions. Genes of interest (A, B, C) are shown as *colored boxes*. 5′ sites in primers and T4 DNA polymerase-treated PCR products are indicated. Regions of homology are indicated by different grayscales. *Inset*: schematic representation of the primer design for SLIC reactions. The homology sequence should be 20 bp long, a similar length should be chosen for the GOI-specific sequence

vector on their 5′ sites, followed by a GOI-specific sequence (Fig. 1). Primers for the creation of multigene constructs are designed in a similar way, providing a complementary sequence to the 5′ adjacent GOI or to the homology sequence of the vector (Fig. 2). Primers for linearization of the vector are complementary to the homology sequences chosen in the GOI primers. The PCR products and the linearized vector are treated with T4 DNA polymerase, which exhibits 3′ exonuclease activity in the absence of dNTPs to generate 5′ overhangs. In vitro recombination is achieved by annealing of the T4 DNA polymerase-treated fragments and transformation of competent *E. coli* cells with the reaction mix.

The combination of SLIC with Cre-LoxP recombination of specific acceptor and donor plasmids in vitro, called tandem recombineering, further increases versatility and flexibility of the generation of multigene constructs for multiprotein expression. The ACEMBL technology is available for *E. coli* (*MultiColi*) [6, 7], mammalian cells (*MultiMam*) [8] and insect cells via a recombinant baculovirus (*MultiBac*) [9, 10] (Fig. 4). Both acceptor and donor plasmids contain LoxP sites for recombination via Cre recombinase in vitro. Acceptor plasmids can be recombined with multiple

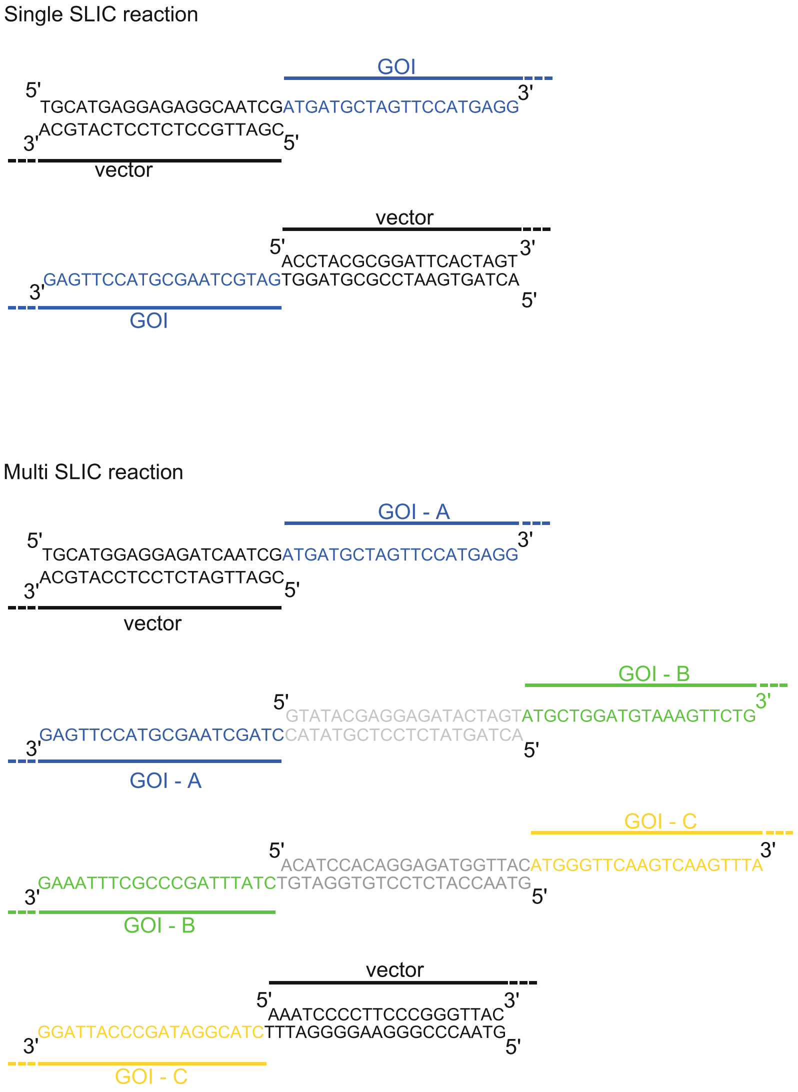

Single SLIC reaction
GOI
5'
TGCATGAGGAGAGGCAATCGATGATGCTAGTTCCATGAGG
3'
3'
ACGTACTCCTCTCCGTTAGC
5'
vector
vector
5'
ACCTACGCGGATTCACTAGT
3'
3'
GAGTTCCATGCGAATCGTAGTGGATGCGCCTAAGTGATCA
5'
GOI
Multi SLIC reaction
GOI - A
5'
TGCATGGAGGAGATCAATCGATGATGCTAGTTCCATGAGG
3'
3'
ACGTACCTCCTCTAGTTAGC
5'
vector
GOI - B
5'
GTATACGAGGAGATACTAGTATGCTGGATGTAAAGTTCTG
3'
3'
GAGTTCCATGCGAATCGATCCATATGCTCCTCTATGATCA
5'
GOI - A
GOI - C
5'
ACATCCACAGGAGATGGTTACATGGGTTCAAGTCAAGTTTA
3'
3'
GAAATTTCGCCCGATTTATCTGTAGGTGTCCTCTACCAATG
5'
GOI - B
vector
5'
AAATCCCCTTCCCGGGTTAC
3'
3'
GGATTACCCGATAGGCATCTTTAGGGGAAGGGCCCAATG
5'
GOI - C

Fig. 2 Examples for primer design for single- and multigene SLIC reactions. Complementary sequences to GOIs and vectors are indicated by lines, as well as 5′ and 3′ sites. Homology regions for multi SLIC reactions are shown in different grayscales. The sequences shown do not refer to a specific vectors or GOIs and need to be changed accordingly

donors to create fused plasmids for multiprotein expression (Fig. 3). Since donor plasmids carry a conditional origin of replication (R6Kγ), they are only propagated in *pir* positive *E. coli* strains or after fusion with one/multiple acceptor plasmids in conventional cloning (*pir* negative) strains [6, 7]. This, in combination with different antibiotic resistances, (Fig. 4) allows for specific selection of the desired Cre-LoxP recombined multiplasmid constructs. The methods described here were optimized to be integrated in an automated robotic setup with a liquid handling system [6].

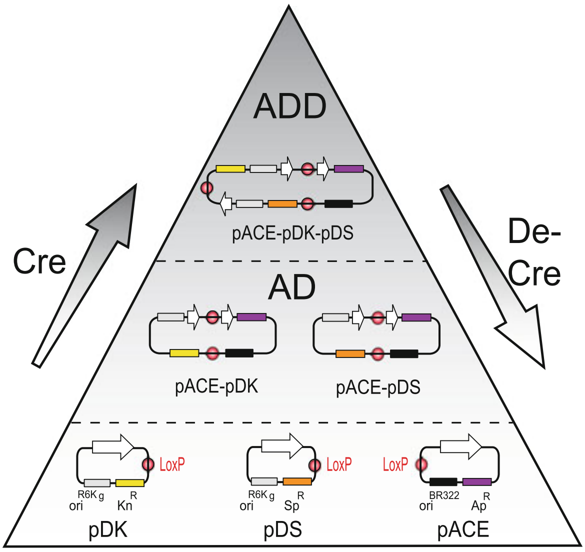


Fig. 3 Schematic representation of the Cre-LoxP recombination process. The Cre recombination process is an equilibrium reaction and gives rise to all combinations of the acceptor (A) and donor (D) fusions. One acceptor can be fused with multiple donors. Desired acceptor-donor fusions (A-D1/A-D2/A-D1-D2) are selected via specific antibiotics (*colored boxes*). The process of Cre-LoxP recombination is reversible (De-Cre reaction). LoxP sites are shown as red balls. Adapted from [11]

2 Materials

All solutions should be prepared using ultrapure water (Millipore Milli-Q system or equivalent; resistivity of 18.2 MΩ cm at 25 °C) and analytical grade reagents. Store all buffers, antibiotics, and enzymes at −20 °C.

2.1 Preparation of Vector

1. LB Broth (Miller).
2. Purified Agar Agar.
3. Sterile polystyrene Falcon tube (15 mL).
4. QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27104).
5. Antibiotics: Ampicillin, Chloramphenicol, Spectinomycin, Tetracycline, Gentamycin, Kanamycin. Concentration of stock solutions (1,000×): Ampicillin 50 mg/mL in water, Carbenicillin 50 mg/mL in 50 % ethanol, Chloramphenicol 34 mg/mL in ethanol, Spectinomycin 10 mg/mL in water, Tetracycline 12.5 mg/mL in 70 % ethanol, Gentamycin 10 mg/mL in water, Kanamycin 30 mg/mL in water (*see* **Note 1**).

2.2 PCR and Linearization of Vector

1. Phusion High-Fidelity DNA Polymerase (Thermo Scientific).
2. 5× Phusion HF Buffer (included in kit).
3. 10 mM dNTP mix (New England Biolabs Inc.).
4. Thermocycler (e.g., Biometra GmbH, Thermocycler T3000).

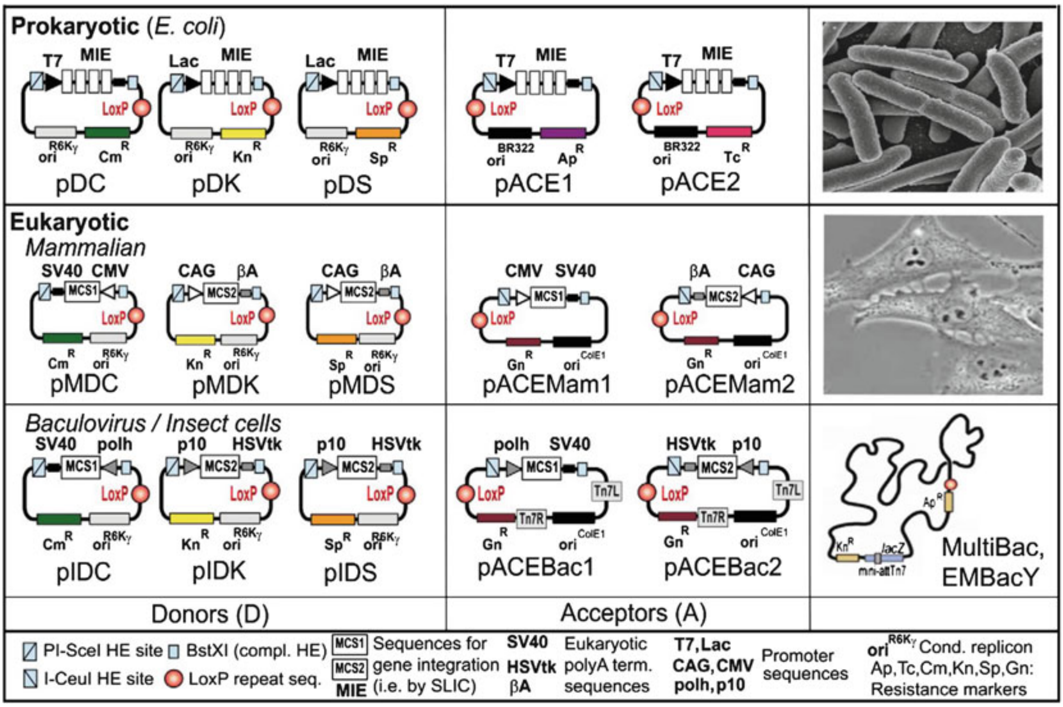

Prokaryotic (E. coli)
T7 MIE LoxP R6Kγ ori Cm^R pDC
Lac MIE LoxP R6Kγ ori Kn^R pDK
Lac MIE LoxP R6Kγ ori Sp^R pDS
T7 MIE LoxP BR322 ori Ap^R pACE1
T7 MIE LoxP BR322 ori Tc^R pACE2
Eukaryotic
Mammalian
SV40 CMV MCS1 LoxP Cm^R R6Kγ ori pMDC
CAG βA MCS2 LoxP Kn^R R6Kγ ori pMDK
CAG βA MCS2 LoxP Sp^R R6Kγ ori pMDS
CMV SV40 MCS1 LoxP Gn^R ColE1 ori pACEMam1
βA CAG MCS2 LoxP Gn^R ColE1 ori pACEMam2
Baculovirus / Insect cells
SV40 polh MCS1 LoxP Cm^R R6Kγ ori pIDC
p10 HSVtk MCS2 LoxP Kn^R R6Kγ ori pIDK
p10 HSVtk MCS2 LoxP Sp^R R6Kγ ori pIDS
polh SV40 MCS1 LoxP Tn7L Tn7R Gn^R ColE1 ori pACEBac1
HSVtk p10 MCS2 LoxP Tn7L Tn7R Gn^R ColE1 ori pACEBac2
Ap^R Kn^R lacZ mini-attTn7
MultiBac, EMBacY
Donors (D)
Acceptors (A)
▨ PI-SceI HE site
□ BstXI (compl. HE)
▨ I-CeuI HE site
● LoxP repeat seq.
MCS1 MCS2 Sequences for gene integration (i.e. by SLIC)
MIE
SV40 HSVtk βA Eukaryotic polyA term. sequences
T7,Lac CAG,CMV polh,p10 Promoter sequences
ori^{R6Kγ} Cond. replicon
Ap,Tc,Cm,Kn,Sp,Gn: Resistance markers

Fig. 4 Overview of available ACEMBL systems showing all acceptor and donor plasmids for prokaryotic (*MultiColi*), mammalian (*MultiMam*), and baculovirus expression (*MultiBac*). Reprinted from: Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells, 175(2), Vijayachandran, L.S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., Richmond, T.J., and Berger, I., pages 198–208, Copyright 2011, with permission from Elsevier. For reagents, contact: iberger@embl.fr

2.3 DpnI Digest

1. *Dpn*I (New England Biolabs Inc.).
2. 10× NEBuffer 4 (included in kit).
3. 37 °C water bath.
4. Qiagen spin column (QIAquick Gel Extraction Kit).
5. Qiagen buffers (included in kit).
6. Agarose gel electrophoresis system (e.g., BioRad, Mini-Sub Cell GT System). 5× TBE Buffer: 0.89 M Tris base, 0.89 M boric acid, 20 mM EDTA (pH 8.0). Weigh 108 g Tris base (MW: 121.10 g/mol) and 55 g boric acid (MW: 61.83 g/mol) and add 40 mL of 0.5 M EDTA, pH 8.0 in a 2-L graduated cylinder. Having water on the bottom of the cylinder (~400 mL) and stirring while adding Tris base and boric acid helps to dissolve these components. Fill up to a total volume of 2 L with water. Filter through 0.22 μm filter and autoclave to prevent precipitation during long-term storage. Store at room temperature.

7. Agarose Type D-5 DNA-grade (Euromedex, ref. D5-D).
8. 6× DNA Loading Dye: 30 % (v/v) glycerol, 0.125 % (w/v) bromophenol blue, 0.125 % (w/v) Xylene cyanol FF (*see* **Note 2**).
9. 1 kb DNA Ladder (New England Biolabs Inc., cat. no. N3232S) (*see* **Note 3**).

2.4 T4 DNA Polymerase Treatment

1. T4 DNA Polymerase (New England Biolabs Inc., cat. no. M0203S).
2. NEBuffer 2 (included in kit).
3. 2 M Urea.
4. 500 mM EDTA. Weigh 73.06 g EDTA (MW: 292.24 g/mol), add 400 mL of water, and adjust pH to 8.0. EDTA will not dissolve until the pH is adjusted to 8.0. Top up to a total volume of 500 mL. Filter through a 0.22 μm filter and store at room temperature.
5. 75 °C heat block.

2.5 SLIC Annealing

1. 65 °C heat block.

2.6 Transformation of Chemical Competent Cells

1. BW23474 chemical competent cells or equivalent.
2. 42 °C water bath.
3. LB Broth (Miller).
4. 37 °C shaking incubator.

2.7 Cre-LoxP Recombination

1. Cre Recombinase (New England Biolabs Inc.).
2. 10× Cre Recombinase Reaction Buffer (New England Biolabs Inc.).
3. 37 °C water bath.

3 Methods

3.1 Preparation of Vector

1. Inoculate 5 mL of LB Broth containing appropriate antibiotics in a 15 mL Falcon tube from a glycerol stock of *E. coli* cells containing the desired vector. Incubate at 37 °C, agitating at 150 rpm for 12 h. Concentrations for antibiotics: Ampicillin 50 μg/mL, Chloramphenicol 34 μg/mL, Spectinomycin 100 μg/mL, Tetracycline 12.5 μg/mL, Gentamycin 10 μg/mL, Kanamycin 30 μg/mL.
2. Centrifuge the Falcon tubes for 10 min at 5,000 × *g* at 4 °C. Take off the supernatant and invert the Falcon tubes to drain.
3. Perform a plasmid prep using the QIAprep Spin Miniprep Kit and follow the instructions in the product's manual.

4. Determine the concentration of the extracted DNA spectrophotometrically (e.g., Thermo Scientific NanoDrop 2000).

3.2 PCR and Linearization of Vector

1. Identical PCR reactions are set up for amplification of the desired insert and linearization of the vector.
2. Set up a 100 μL PCR reaction in a 0.5 mL PCR tube: Mix 1 μL template DNA (approximately 10 ng) with 20 μL 5× Phusion HF Buffer (*see* **Note 4)**, 2 μL 10 mM dNTP mix, 1 μL of forward primer (concentration 100 μM), 1 μL of reverse primer (concentration 100 μM), and 74.5 μL water.
3. Add 0.5 μL Phusion High-Fidelity DNA Polymerase and mix (*see* **Note 5**).
4. Choose appropriate annealing temperatures for the specific primers chosen to perform the PCR (*see* **Note 6**). Typically, templates are initially denatured at 98 °C for 60 s; followed by 30 cycles at 98 °C for 20 s, the specific annealing temperature for 30 s, 72 °C for 30 s (for 1 kb product size); and a single final step at 72 °C for 10 min.

3.3 Dpn1 Digest and Purification of PCR Product and Linearized Vector

1. Add 20 U of Dpn1 directly to the 100 μL PCR product and incubate at 37 °C for 2 h (*see* **Note 7**). This step is not required to insert PCR reactions if the resistance marker of the template plasmid differs from the destination vector.
2. Mix with 20 μL 6× DNA loading dye, load on 1 % TBE agarose gel, and run the gel at 100 V (*see* **Note 8**) for around 1.5 h until the 1 kb DNA ladder is well resolved.
3. Excise the band corresponding to the PCR product using a UV light box and transfer to a 2 mL sterile Eppendorf tube (*see* **Note 9**).
4. Extract the DNA from the gel slices using the QIAquick Gel Extraction Kit following the instructions in the product's manual.
5. Determine the concentration of the extracted DNA spectrophotometrically (e.g., Thermo Scientific NanoDrop 2000).

3.4 T4 DNA Polymerase Treatment of PCR Product and Linearized Vector

1. Set up the reaction in a 0.5 mL PCR tube: Mix 2 μL 10× NEBuffer 2, 1 μL 100 mM DTT, 2 μL 2 M Urea, 0.5 U T4 DNA Polymerase, and 1 μg of the purified PCR product (*see* **Note 10**) in a total volume of 20 μL. For a 20 bp overhang between PCR product and vector, incubate for 30 min at room temperature (*see* **Note 11**).
2. Stop the reaction by adding 1 μL of 500 mM EDTA.
3. Inactivate T4 DNA Polymerase by heating to 75 °C for 20 min.

3.5 SLIC Annealing

1. Set up the reaction in a 0.5 mL PCR tube: Mix 10 μL of the T4 DNA polymerase treated vector with 10 μL of T4 DNA polymerase-treated insert.
2. Incubate at 65 °C for 10 min and let cool down slowly in the heat block at RT.

3.6 Transformation of Chemical Competent Cells

1. Mix 5 μL of the annealing reaction with 50 μL of BW23474 chemical competent cells on ice and incubate for 30 min, heat shock at 42 °C for 60 s, incubate on ice for 2 min, add 400 μL of LB Broth, and incubate in a 37 °C shaker for 1 h.
2. Plate 100 μL of the cells on a selective LB agar plate with appropriate antibiotic(s), pellet the remaining cells by centrifugation at 4,000 × *g* for 1 min, take off 250 μL supernatant, and resuspend the pellet in the remaining 100 μL. Plate this concentrated cell suspension cells on a second LB agar plate.

3.7 Cre-LoxP Recombination of Acceptor and Donor Vectors

1. Set up a 10 μL Cre-LoxP recombination reaction in a 0.5 mL PCR tube: Mix 1 μg of Donor vector with a 1:1 M ratio of Acceptor, 1 μL 10× Cre Recombinase Reaction Buffer, and 0.5 μL Cre Recombinase in a 10 μL total reaction volume.
2. Incubate the reaction at 37 °C for 1 h (*see* **Note 12**).

3.8 Transformation of Chemical Competent Cells

1. Mix 5 μL of the Cre-LoxP recombination reaction with 50 μL of BW23474 chemical competent cells on ice and incubate for 30 min, heat shock at 42 °C for 60 s, incubate on ice for 2 min, add 400 μL of LB Broth, and incubate at 37 °C for overnight (*see* **Note 13**).
2. Plate 100 μL of the cells on a selective LB agar plate with appropriate antibiotic(s), pellet the remaining cells by centrifugation at 4,000 × *g* for 1 min, take off 250 μL supernatant, and resuspend the pellet in the remaining 100 μL. Plate the remaining cells on a second LB agar plate.

4 Notes

1. Carbenicillin can be used as a substitute for Ampicillin (at the same concentration) to reduce the presence of satellite colonies.
2. Add 0.125 % Orange G to the 6× DNA Loading Dye if working with small PCR products. Orange G migrates at about 50 bp in 1 % TBE agarose gels and helps to determine the time needed for electrophoresis.
3. For smaller PCR products, use a 100 bp DNA ladder (New England Biolabs Inc., cat. no. N3231S) to identify fragments in the range of 100 bp to 1 kbp more easily.

4. When using the Phusion High-Fidelity DNA Polymerase kit, the 5× GC buffer can help to increase the performance of Phusion High-Fidelity DNA Polymerase on long or GC rich templates. When working with GC rich templates, add 3 % DMSO as a PCR additive to aid denaturing of templates with high GC content. It is practical to run two PCR reactions with HF and GC buffer in parallel and compare yield and PCR product specificity for both reactions.
5. Mix by pipetting or flipping the tube. Centrifuge for 10 s at 4,000 × *g* to collect the mix on the bottom of the PCR tube. No bubbles should remain in the tube.
6. When using the Phusion High-Fidelity DNA Polymerase kit, calculate the annealing temperature with the manufacturer's Tm calculator tool on the website: http://www.finnzymes.fi/tm_determination.html
7. This is a critical step to reduce background colonies after transformation. The Dpn1 digest can be incubated longer than 2 h (e.g., overnight) to reduce background colonies.
8. Depending on the gel system used, the voltage might be increased up to 120 V to reduce separation time. Increasing the voltage can result in heating up and melting the agarose gel.
9. 1.5 mL Eppendorf tubes can be used in this step as well, depending on the size of the gel slice. When excising the desired band from the agarose gel, use longer wavelength (e.g., 365 nm or equivalent) and reduced intensity on the UV lightbox to avoid any modifications to your PCR product.
10. It is important to purify the desired PCR products as described before the T4 DNA Polymerase treatment as residual dNTPs from the PCR reaction inhibit the 3′ exonuclease activity of the T4 DNA Polymerase.
11. The incubation time is a critical step for T4 DNA Polymerase treatment. A too short incubation time will result in nonoverlapping overhangs between PCR product and vector and impede correct annealing.
12. Longer incubation times will likely lead to undesired higher molecular weight recombination products.
13. Long recovering times are essential to obtain positive transformants, especially when creating multiple acceptor-donor fusions due to the high selective pressure from the combination of antibiotics used.

Acknowledgments

We thank all members of the Berger laboratory for helpful discussions. M.H. is recipient of a Kekulé fellowship of the Fonds der Chemischen Industrie (FCI, Germany). Y.N. is a fellow of the Marie-Curie training network Chromatin Plasticity and the Boehringer Ingelheim Foundation (BIF, Germany). I.B. acknowledges support from the Swiss National Science Foundation (SNSF), the Agence Nationale de la Recherche (ANR), the Centre National de la Recherche Scientifique (CNRS), the EMBL and the European Commission (EC) through the joint EIPOD program, and the European Commission (EC) projects SPINE2-Complexes and 3D-Repertoire (Framework Program 6 (FP6)), as well as INSTRUCT, PCUBE, BioSTRUCT-X, and ComplexINC (EC FP7).

Competing financial interest statement: The authors declare competing financial interests. I.B. is author on patents and patent applications related to the methods here described.

References

1. Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol 3:109–118
2. Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and sequencing analysis of enzymatically amplified genomic sequences. Science 233:1076–1078
3. Costa GL, Grafsky A, Weiner MP (1994) Cloning and analysis of PCR-generated DNA fragments. PCR Methods Appl 3:338–345
4. Esposito D, Garvey LA, Chakiath CS (2009) Gateway cloning for protein expression. Methods Mol Biol 498:31–54
5. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
6. Bieniossek C, Nie Y, Frey D, Olieric N, Schaffitzel C, Collinson I, Romier C, Berger P, Richmond TJ, Steinmetz MO, Berger I (2009) Automated unrestricted multigene recombineering for multiprotein complex production. Nat Methods 6:447–450
7. Nie Y, Bieniossek C, Frey D, Olieric N, Schaffitzel C, Steinmetz MO, Berger I (2009) ACEMBLing multigene expression vectors by recombineering. Nat Protoc 4. doi:10.1038/nprot.2009.104
8. Kriz A, Schmid K, Baumgartner N, Ziegler U, Berger I, Ballmer-Hofer K, Berger P (2010) A plasmid-based multigene expression system for mammalian cells. Nat Commun 1. doi:10.1038/ncomms1120
9. Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3:1021–1032
10. Berger I, Fitzgerald DJ, Richmond TJ (2004) Baculovirus expression system for heterologous multiprotein complexes. Nat Biotechnol 22:1583–1587
11. Vijayachandran LS, Viola C, Garzoni F, Trowitzsch S, Bieniossek C, Chaillet M, Schaffitzel C, Busso D, Romier C, Poterszman A, Richmond TJ, Berger I (2011) Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J Struct Biol 175:198–208

Chapter 12

Combinatorial DNA Assembly Using Golden Gate Cloning

Carola Engler and Sylvestre Marillonnet

Abstract

A basic requirement for synthetic biology is the availability of efficient DNA assembly methods. We have previously reported the development of Golden Gate cloning, a method that allows parallel assembly of multiple DNA fragments in a one-tube reaction. Golden Gate cloning can be used for different levels of construct assembly: from gene fragments to complete gene coding sequences, from basic genetic elements to full transcription units, and finally from transcription units to multigene constructs. We provide here a protocol for DNA assembly using Golden Gate cloning, taking as an example the level of assembly of gene fragments to complete coding sequences, a level of cloning that can be used to perform DNA shuffling. Such protocol requires the following steps: (1) selecting fusion sites within parental sequences (sites at which parental sequences will be recombined), (2) amplifying all DNA fragments by PCR to add flanking restriction sites, (3) cloning the amplified fragments in intermediate constructs, and (4) assembling all or selected sets of intermediate constructs in a compatible recipient vector using a one-pot restriction-ligation.

Key words Synthetic biology, DNA assembly, DNA shuffling, Combinatorial, Hierarchical, Type IIS restriction enzymes, Seamless cloning, Modular cloning

1 Introduction

The emerging field of synthetic biology promises the production of organisms with novel phenotypes useful for medicine, agriculture, and industry. Unlike traditional biotechnology, which has so far produced organisms containing relatively low numbers of modified genes, synthetic biology aims to engineer organisms with larger numbers of modified genetic elements, potentially at up to genome scale. Such endeavor requires methods that can allow parallel assembly of multiple DNA fragments very efficiently. Fortunately, several methods that provide this capability have been developed recently [1]. Most of these methods are based on homologous recombination between sequences present at the ends of the fragments to assemble [2–4]. Methods that would allow assembly of multiple DNA fragments without a requirement for sequence homology would be particularly useful for combinatorial assembly

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology,* Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_12, © Springer Science+Business Media New York 2013

of libraries of standard basic genetic elements, especially at the level of transcription units, where fixed junctions between various genetic elements need to be as small as possible.

We have recently developed a cloning method, called Golden Gate cloning, that is suitable for this purpose [5, 6]. The principle of this method is based on the ability of type IIS enzymes to cleave outside of their recognition site, allowing two DNA fragments flanked by compatible restriction sites to be digested and ligated seamlessly [7–10] (Fig. 1a). Since the ligated product of interest does not contain the original type IIS recognition site, it will not be subject to redigestion in a restriction-ligation reaction. However, all other products that reconstitute the original site will be redigested, allowing their components to be made available for further ligation, leading to formation of an increasing amount of the desired product with increasing time of incubation. Since the sequence of the overhangs at the ends of the digested fragments can be chosen to be any 4-nucleotide sequence of choice, multiple compatible DNA fragments can be assembled in a defined linear order in a single restriction-ligation step.

Golden Gate assembly can be used for assembly of multigene constructs from libraries of standardized modules containing basic genetic elements such as promoters, gene coding sequences, and terminators using a succession of one-pot assembly reactions [11, 12]. A first Golden Gate cloning reaction is used for assembly of transcription units from standard basic genetic elements, while a second reaction is used for generation of multigene constructs from individual transcription units (levels 1 and 2, respectively, Fig. 2). At a lower level of assembly, Golden Gate cloning can also be used to generate new biological parts (such as new promoters or gene coding sequences) by DNA shuffling of several parental sequences (assembly from level −1 to level 0, Fig. 2). An example of novel coding sequences created by combinatorial assembly of variant gene fragments is the construction of DNA binding proteins with user-defined binding specificities [13–17]. We provide here a protocol for gene shuffling, but the conditions described for the assembly step can also be applied for other levels of cloning (i.e., for assembly of transcription units or of multigene constructs). Performing gene shuffling using Golden Gate cloning basically requires four steps: (1) defining a set of fusion sites that will be used to link the various DNA fragments to be assembled, (2) amplifying all DNA fragments by PCR to add flanking restriction sites, (3) cloning the amplified fragments in intermediate constructs, and (4) assembling all or selected sets of intermediate constructs in a compatible recipient vector using a one-pot restriction-ligation.

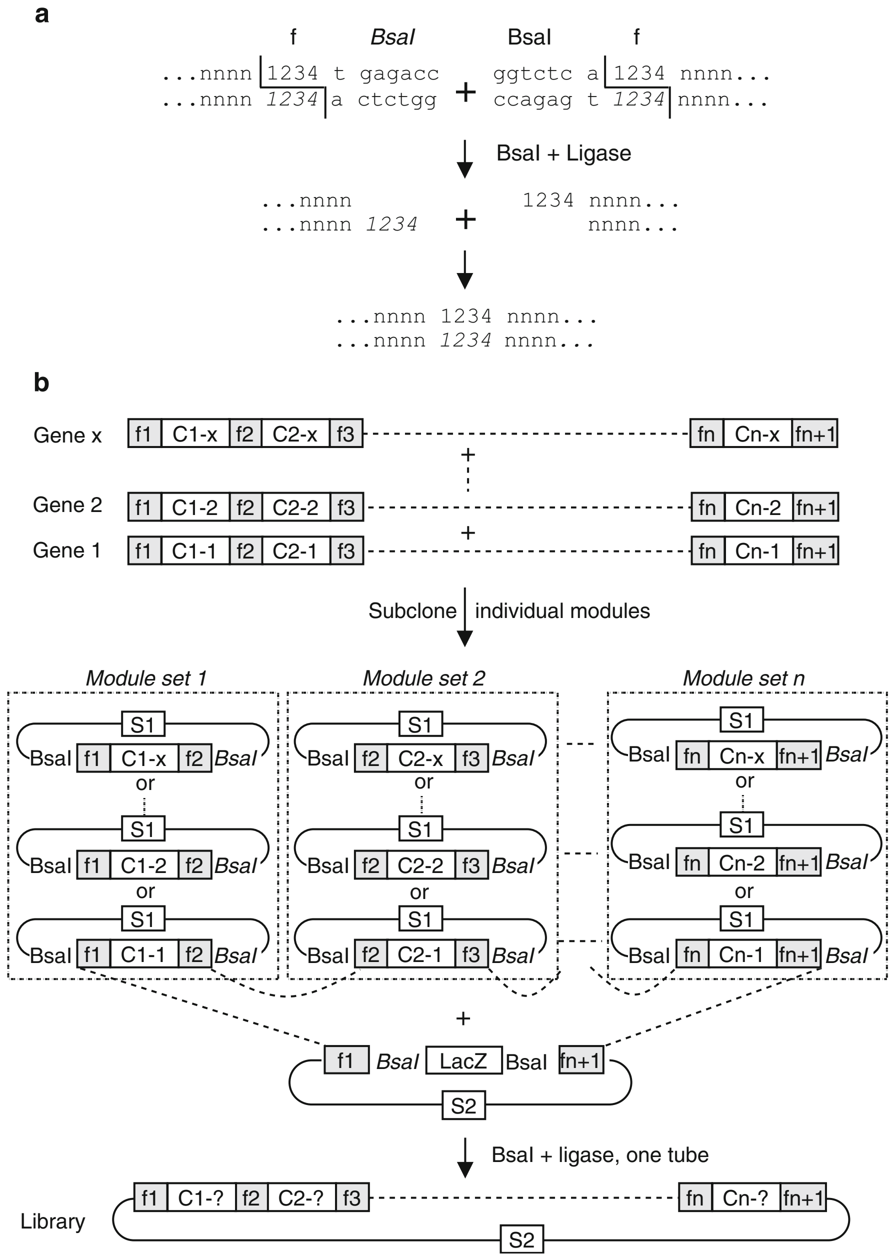


Fig. 1 DNA shuffling strategy. (**a**) Two DNA ends terminated by the same 4 nucleotides (sequence f, composed of nucleotides 1234, complementary nucleotides noted in *italics*) flanked by a BsaI recognition sequence, ggtctc, form two complementary DNA overhangs after digestion with BsaI. (**b**) For DNA shuffling, genes of interest are aligned, and fusion sites consisting of 4-nucleotide sequences (f1 to fn + 1) are defined on conserved sequences. Module fragments (core sequence, C1 to Cn, plus flanking 4-nucleotide sequences) are amplified by PCR and cloned in an intermediate cloning vector. Module fragment plasmids and the acceptor vector are assembled with a restriction-ligation using BsaI and ligase. S1 and S2, two different selectable markers

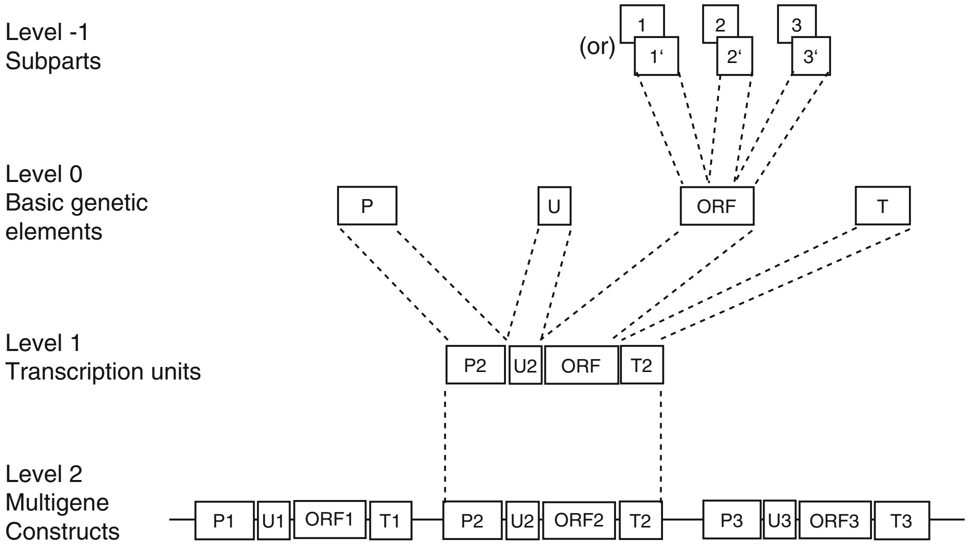


Fig. 2 Golden Gate cloning can be used for different levels of construct assembly. Cloning at each step is performed using a similar assembly reaction, except that different type IIS enzymes must be used for successive levels of assembly. This is because each cloning step results in constructs that lack restriction sites for the type IIS used. Cloning from level −1 to level 0 can be used for gene or promoter shuffling, or to make various gene fusions

2 Materials

2.1 PCR

1. Novagen KOD Hot Start DNA polymerase (Merck KGaA, Darmstadt), supplied with 10× buffer, 25 mM $MgSO_4$, and 2 mM dNTPs.
2. Custom-made primers can be ordered from many commercial vendors (e.g., Invitrogen, Karlsruhe).
3. NucleoSpin® Extract II kit (Macherey Nagel, Düren), for purification of PCR products.

2.2 Cloning

1. Restriction endonuclease SmaI (10 U/μL) (NEB, New England Biolabs Inc., Ipswich, MA, USA), supplied with 10× NEBuffer 4 (200 mM Tris–acetate pH 7.5, 100 mM magnesium acetate, 500 mM potassium acetate, 10 mM dithiothreitol).
2. Restriction endonuclease BsaI (10 U/μL) (NEB), supplied with 10× NEBuffer 4.
3. T4 DNA Ligase 3 U/μL or T4 DNA Ligase (HC) 20 U/μL (Promega, Mannheim), both supplied with 10× ligation buffer (300 mM Tris–HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP).

4. For measuring DNA concentration, we use the NanoDrop ND2000 (Peqlab, Erlangen).
5. Luria-Bertani (LB) Medium: 1 % bacto-tryptone, 0.5 % yeast extract, 1 % NaCl in deionized water, adjusted to pH 7.0 with 5 N NaOH. For plates, 1.5 % agar is added.
6. Antibiotics carbenicillin (used instead of ampicillin) and kanamycin: filter-sterilized stocks of 50 mg/mL in H_2O (stored in aliquots at −20 °C) are diluted 1:1,000 (final concentration: 50 μg/mL) in an appropriate amount of medium after the medium has been autoclaved and cooled down. For spectinomycin, a stock of 40 mg/mL is made and is used at a final concentration of 100 μg/mL (dilution 1:400).
7. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal): stock solution of 20 mg/mL in dimethylformamide (DMF). For preparation of plates, the stock is diluted 1:500 (final concentration: 40 μg/mL) in an appropriate amount of LB agar after autoclaving/melting and cooling down.

2.3 Preparation of Chemically Competent Cells

1. Solution TFB1: 30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15 % glycerol; adjust to pH 5.8 (with 1 M acetic acid), filter-sterilize, and store at 4 °C (ready to use) or at room temperature (cool down before use).
2. Solution TFB2: 100 mM MOPS (or PIPES), 75 mM $CaCl_2$, 10 mM RbCl, 15 % glycerol; adjust to pH 6.5 (with 1 M KOH), filter-sterilize, and store at 4 °C (ready to use) or at room temperature (cool down before use).
3. The OD_{600} of bacterial cultures is measured in a SmartSpec™ 3000 spectrophotometer (Biorad, Muenchen).

2.4 Screening of Colonies

1. DNA minipreps: NucleoSpin® Plasmid Quick Pure (Macherey Nagel, Dueren).
2. Restriction endonucleases (NEB or Fermentas, St. Leon-Rot), all supplied with 10× buffer and if necessary also with 100× BSA (dilute 1:10 and store in aliquots at −20 °C).
3. DNA ladder: GeneRuler™ 1 kb DNA Ladder Plus (Fermentas) is used as marker for gel electrophoresis.
4. 50× TAE buffer: 242.0 g Tris–HCl, 57.1 mL acetic acid, and 100 mL 0.5 M EDTA, pH 8.0, in 1 L of deionized water.
5. Gels: agarose (0.7–1.5 %) in 1× TAE is melted in a microwave oven and one drop of a 0.025 % ethidium bromide solution (Carl Roth GmbH, Karlsruhe) is added per 100 mL of melted agarose solution.
6. Running buffer of agarose gels is 1× TAE.

7. Gels are checked visually using a Syngene GelVue transilluminator (VWR, Darmstadt), and pictures are taken by using a Quantity one® gel analysis software (Biorad).
8. DNA maps of plasmids are made by using the Vector NTI software (Invitrogen).

2.5 Sequencing

1. DNA/constructs to sequence are sent to an external contractor (GATC Biotech, Konstanz). Sequence data are analyzed using the DNASTAR's Lasergene software.
2. Primers M13RP (CAGGAAACAGCTATGACC) and/or M13FP (TGTAAAACGACGGCCAGT) are used for sequencing of inserts cloned in pUC19-derived vectors.

3 Methods

3.1 Selection of Fusion Sites

Fusion sites consist of 3- or 4-nucleotide sequences that will be used for restriction enzyme digestion and ligation for the DNA assembly step. Fusion sites of 4 nucleotides are selected if a type IIS restriction enzyme such as BsaI is used for DNA assembly, or of 3 nucleotides if an enzyme of the type of SapI is used (*see* **Note 1**). For DNA shuffling, fusion sites are selected from sequences conserved among all parental sequences (sequence f1 to fn + 1, Fig. 1). For assembly reactions of a higher order (i.e., for assembly of transcription units or of multigene constructs, levels 1 and 2, Fig. 2), fusion sites do not necessarily consist of native sequences, since these will serve to link various genetic elements of different sequences and origins. In contrast to fusion sites selected for gene shuffling, which need to be defined for each gene individually, these fusion sites can be universally defined and are reused for all assembly reactions of the same level (*see* **Note 2**).

Fusion sites used in the same assembly reaction must follow several requirements:

1. A first requirement is to avoid selecting the same sequence twice, as this could lead to annealing of the ends of inappropriate DNA fragments during the restriction-ligation step (**step 3.5**), resulting in deletion of the sequences between these two sites. It is also important to make sure that the sequence of any site does not match the sequence of any of the other chosen sites on the complement strand. For example, choice of the sequence ATTC will preclude the choice of the sequence GAAT for any of the other fusion sites used for this shuffling experiment. Use of two such sites would sometimes lead to ligation of two inappropriate fragments, one in the opposite orientation. This would lead to formation of molecules that will not be able to form circular plasmids, but that would continue to ligate to further fragments and form linear concatemers.

2. A second requirement is to avoid the 16 palindromic sequences (for 4-nucleotide fusion sites), since any palindromic DNA end can be ligated to another copy of the same DNA fragment in the opposite orientation, and lead to the same problem as described above. For enzymes that produce a 4-nucleotide extension, 240 different sequences are, therefore, available.
3. Finally, a third but optional requirement can be defined to maximize the efficiency of DNA shuffling. We have observed that inappropriate ligation of fragments can sometimes occur between ends in which 3 out of 4 consecutive nucleotides are identical, for example, as in sequences GGTG and AGTG, or GGTG and CACT. Therefore, combination of two such sites should be avoided if possible.

The number of fusion sites, as well as their position within the gene to be shuffled, is chosen depending on the needs of the user for each specific protein. Therefore, the size (*see* **Note 3**) and number of fragments to shuffle/assemble will vary for each gene and each experiment. We have tested up to eight recombination points within a gene (for assembly of nine fragments), but a higher number is possible as well, although with reduced efficiency (*see* **Note 4**).

The following steps consist of amplifying the defined fragments by PCR, cloning them in intermediate vectors, and sequencing them. Alternatively, the fragments amplified by PCR can be used directly for gene assembly or DNA shuffling without an intermediate cloning step, but this approach may lead to reduced cloning efficiency (*see* **Note 5**). Alternatively, these steps can also be replaced by simply ordering the fragments of interest from a gene synthesis company (*see* **Note 6**).

3.2 PCR Amplification of the Modules

Modules defined by the position of the fusion sites are amplified by PCR using primers designed to add two BsaI sites flanking each module. Primers are designed such that the overhangs created by digestion of the amplified products with BsaI (or any other type IIS enzyme chosen) correspond to the sequence of the chosen fusion sites. Therefore, the sequence ttggtctca is added to each primer sequence, for example, ttggtctca CAGG nnnnn (CAGG being the fusion site, followed by 16–20 nucleotides of target sequence). For nine modules prepared from three homologous sequences, 54 primers need to be made.

A requirement for Golden Gate assembly is to not have any internal BsaI sites present within any of the DNA fragments to assemble. Indeed, the presence of a BsaI site within one of the modules would lead to redigestion of the shuffled DNA sequences containing such fragment, at the end of the assembly step. These linear molecules will of course not transform *E. coli*. Therefore, any internal site needs to be removed upon generation of the entry clones. Removal of internal BsaI sites from PCR fragments can be done easily using gene SOEing [18] (*see* **Note 7**). An alternative strategy has also been described elsewhere [11].

PCR can be performed using any source of DNA, but for small modules, template DNA is not necessary (*see* **Note 8**). For amplification, we use the enzyme KOD Hot Start DNA polymerase since it has a very low error rate and, unlike Taq polymerase or several enzyme mixes, produces DNA products with blunt ends. Blunt ends are advantageous, as the products can be easily cloned in any standard vector such as pUC19 by blunt-end cloning (see below).

1. The PCR mix is set up following the manufacturer's instructions, for example, using KOD polymerase, with the following conditions: 1 μL plasmid DNA (5–20 ng/μL), 5 μL of 10× buffer, 3 μL of 25 mM $MgSO_4$, 5 μL of 2 mM dNTPs, 1.5 μL each of 10 μM sense and antisense primers, and 1 μL of KOD Hot Start DNA polymerase (10 U/μL, final concentration 0.02 U/μL) in a total reaction volume of 50 μL.
2. PCR is performed using the following cycling conditions: (1) incubation at 95 °C for 2 min for polymerase activation, (2) denaturation at 95 °C for 20 s, (3) annealing at 58 °C for 10 s—the temperature for the annealing step can be adjusted for specific primers, but the temperature of 58 °C usually works well for primers designed as described above, (4) extension at 70 °C, the duration depends on the length of the expected fragment (from 10 s/kb for fragments smaller than 500 bp up to 25 s/kb for fragments larger than 3 kb, see manufacturer's instructions); **steps 2–4** are repeated 35 times and are followed by a final extension step at 70 °C for 20 s–2 min (depending on fragment length). The reaction is then kept at 12 °C until taken out of the thermocycler.
3. Of the PCR product obtained, 2 μL is then analyzed by gel electrophoresis to make sure that a product of the correct size has been amplified.
4. The amplified fragment is purified from remaining primers, potential primer dimers, and remaining polymerase enzyme by using the NucleoSpin® Extract II kit following the kit protocol. DNA is eluted from the column with 30–50 μL of elution buffer (5 mM Tris–HCl, pH 8.5). In case several bands were amplified rather than only the expected fragment, the same kit can also be used to cut and extract the appropriate DNA fragment from an agarose gel.

3.3 Blunt-End Cloning of the Modules

Many commercial kits are available for cloning PCR products, including the pGEM-T kit (Promega), pJET (Fermentas), and TOPO® TA kit (Invitrogen). Alternatively, PCR products can also be cloned efficiently using blunt-end cloning with a protocol that uses restriction-ligation [19, 20]. This method has the advantage that the DNA fragment of interest to be cloned does not need to be flanked by any specific sequence, and therefore primers used for amplification do not need to have special extensions (*see* **Note 9**). Cloning vectors for generating entry clones for shuffling need to fulfill preferentially two requirements: (1) they should preferably not contain any restriction site for the type IIS enzyme chosen for shuffling (*see* **Note 10**), and (2) the antibiotic resistance gene of the entry vector should preferably be different from the one in the destination vector. Since several commercial cloning vectors have a BsaI restriction site in the ampicillin resistance gene (e.g., pGEM-T or pJET), we have made our own entry cloning vector that simply consists of pUC19 lacking a BsaI restriction site (*see* **Note 11**).

1. Add 0.5 μL of vector (50 ng), 1 μL of PCR product (50–100 ng), 2 μL of 10× ligation buffer (Promega), 1 μL of SmaI enzyme (10 U; NEB), 1 μL of ligase (3 U; Promega), and 14.5 μL of water (total volume of 20 μL) into a tube. The reaction mix is incubated for 1–2 h at room temperature or in a 25 °C incubator, if one is available.
2. The entire ligation mix is transformed to DH10B chemically competent cells and plated on LB plates with X-gal and the appropriate antibiotic (the transformation protocol is described below in Subheading 3.6).
3. White colonies (or sometimes pale blue when small inserts are cloned) are picked and inoculated in 5 mL of LB medium containing the appropriate antibiotic.
4. Plasmid DNA is extracted using the NucleoSpin® Plasmid Quick Pure kit from Macherey Nagel following the manufacturer's instructions.
5. Plasmid DNA can be checked by restriction enzyme digestion using BsaI, and analysis of the digested DNA by agarose gel electrophoresis. A fragment of the size of the expected module should be visible.
6. Two minipreps are sent for sequencing using primers M13RP and/or M13FP.
7. When the correct sequence has been verified, DNA concentration of the plasmid prep is measured using the NanoDrop ND2000 (Peqlab).

3.4 Construction of the Destination Vector

A destination vector compatible with the entry modules needs to be made. The vector should respect the following criteria. It should contain two BsaI sites (or any other type IIS enzyme chosen) with cleavage sites compatible with the beginning of the first module set and the end of the last module set. The vector backbone should not contain any BsaI restriction site and should have an antibiotic resistance gene different from the one used for the entry clones. Additionally, the vector may contain a lacZα fragment (Fig. 1b) to allow blue-white selection of the resulting clones. Optionally, restriction sites for a second type IIS enzyme may be added to the vector to allow further subcloning of the sequence that will be formed by assembly of the fragments in the first assembly reaction. An example of destination vector sequence is as follows: gaagac aa abcd 1234 t gagacc, followed by LacZ gene sequences, followed by sequence ggtctc a 5678 efgh tt gtcttc followed by vector backbone sequences. In this example, fusion sites 1234 and 5678 are used for assembly of several fragments using BsaI, while fusion sites abcd and efgh are used for further subcloning of the assembled sequence using BpiI.

3.5 Golden Gate DNA Assembly

Once entry constructs and the recipient vector are made and sequenced, performing DNA shuffling/assembly only requires pipetting all components into a reaction mix, incubating the mix in a thermocycler, and transforming the ligation mix into competent cells. An important factor is to use an equimolar amount of DNA for each of the module sets and the destination vector. If a module set contains several modules, the amount of DNA for each module of a set containing *x* different modules (*x* alternative homologous sequences) should contain only 1/*x* the amount of DNA compared to the vector; for example, each module from a set containing three modules should have a third the amount of DNA compared with the recipient vector.

1. A restriction-ligation is set up by pipetting 40 fmol (or 100 ng, *see* **Note 12**) of each module set and of the vector, 2 μL 10× ligation buffer, 10 U (1 μL) of BsaI, and either 3 U (1 μL) of ligase for assembly of 2–4 module sets or 20 U (1 μL) HC ligase for assembly of more than four module sets, in a total volume of 20 μL into a tube.
2. The restriction-ligation mix is incubated in a thermocycler. For assembly of 2–4 module sets, incubation for 60–120 min at 37 °C is sufficient. If more module sets are ligated together, the incubation time is increased to 6 h, or cycling is used as following: 2 min at 37 °C followed by 3 min at 16 °C, both repeated 50 times (*see* **Note 13**).
3. Restriction-ligation is followed by a digestion step (5 min at 50 °C) and then by heat inactivation for 5 min at 80 °C. The final incubation step at 80 °C is very important and should not be omitted. Its purpose is to inactivate the ligase at the end of the restriction-ligation. Omitting this step would lead to religation of some of the insert and plasmid backbone fragments still present in the mix, when it is taken out of the thermocycler

before transformation. Such unwanted products might be ligated more efficiently than they are redigested by the type IIS enzyme at room temperature or on ice in the time interval between restriction-ligation and transformation. Therefore, a larger percentage of colonies would contain such type of undesired ligation products.

3.6 Transformation of the Library in Competent Cells

The entire ligation is transformed into chemically competent DH10B cells (*see* **Note 14**).

1. Frozen chemically competent cells (100 μL per tube) are thawed on ice.
2. The entire ligation is added to the cells and the mix is incubated on ice for 30 min.
3. The cells and DNA mix is heat shocked for 90 s at 42 °C in a water bath.
4. The cells are allowed to recover on ice for 5 min.
5. To the cells, add 1 μL of LB medium and incubate the tube at 37 °C in a shaker-incubator (150 rpm) for 45 min to 1 h.
6. After incubation, 25–100 μL of the transformation is plated on LB agar plates containing antibiotic and X-gal. Plating of an aliquot of the transformation is necessary to estimate the number of independent constructs that will be obtained. The remainder of the transformation can be inoculated into 5 mL of liquid LB with the appropriate antibiotic.
7. The plates and liquid culture are incubated overnight at 37 °C.
8. Many white and very few blue colonies should be obtained on the plate. A few white colonies from the plate can be picked for preparation of miniprep DNA. Plasmid DNA can be analyzed by restriction digestion and sequencing to estimate the number of correct clones.
9. Miniprep DNA is also prepared from the liquid culture. This DNA prep should represent a library of constructs containing shuffled DNA.

Depending on the specific goal of the shuffling experiment, clones can be functionally screened either individually or as a library. The shuffled plasmid library may be transformed in any target organism of choice for functional screening.

3.7 Preparation of Chemically Competent DH10B Cells

Chemically competent or electrocompetent *E. coli* cells can be either purchased from a commercial vendor or made in the laboratory. The protocol that we use is as follows:

1. *E. coli* strain DH10B is inoculated from a glycerol stock onto an LB plate; the inoculum is streaked on the plate using a loop

so as to obtain individual colonies. The plate is incubated overnight at 37 °C.

2. Inoculate 5 mL of LB from a single colony and incubate the flask overnight in a shaker-incubator (37 °C, shaking 150 rpm).
3. The following day, transfer 2 mL of this culture to a flask containing 200 mL of LB and incubate for around 2 h until OD_{600} reaches 0.6.
4. Cool down the cells on ice for 10 min. The cells are pelleted in a centrifuge for 5 min at 4,500 rpm (4,000 × *g*) at 4 °C. The cells are resuspended in 0.4 volume of ice-cold TFB1.
5. Repeat the centrifugation. Resuspend the pellet in 1/25 volume of ice-cold TFB2.
6. The cells are aliquoted 100 μL per tube and shock-frozen in liquid nitrogen. The aliquots are stored at −80 °C.

4 Notes

1. Several different type IIS enzymes can be used for construct assembly. We have, for example, tested the enzymes BsaI, BpiI, and Esp3I. For all three, restriction-ligation can be performed efficiently in ligase buffer from Promega. All three have a 6-bp recognition sequence and a 4-nucleotide cleavage site located 1 (BsaI and Esp3I) or 2 nucleotides (BpiI) away from the recognition sequence. Enzymes of the type of SapI such as LguI can also be used efficiently in a restriction-ligation [21]. These enzymes have a 7-bp recognition sequence, and, therefore, fewer sites will have to be removed from sequences of interest to clone (discussed in Subheading 3.2). However, these enzymes have a 3-nucleotide cleavage site, meaning that only 64 different sequences are available to choose from for use as fusion sites.
2. To standardize DNA assembly reactions, each basic module type (such as promoter, coding sequence, or terminator), is flanked by two defined fusion sites. This allows any module to be cloned interchangeably with any other module of the same type using the same assembly reaction. We and others have earlier published sets of fusion sites that can be used for assembly of transcription units [11, 22]. These sequences may be used by other groups, although they may still be changed before they are universally adopted. For example, we have selected 4-nucleotide-long fusion sites for assembly of transcription units, but the use of 3-nucleotide sequences (using SapI or LguI for assembly) would allow to have fewer nucleotides fixed in a transcription unit.

3. The minimal size of modules that we have tested is 38 bp (including the fusion sites, but excluding the flanking BsaI recognition sites). In theory, a module needs to be long enough for the two strands to remain annealed under restriction-ligation conditions, usually 37 °C. Therefore, smaller modules could theoretically be made, which would be useful if a user wants to focus his efforts on a very small region of a protein of interest. Another example of small modules would consist of modules containing protein purification or detection tags.

4. We have successfully assembled up to 17 fragments in one step, but with seriously reduced efficiency (0–3 positive clones out of 12 colonies analyzed [17]). Ligation of nine fragments can be quite efficient, but efficiency seems to decrease rapidly with a higher number of fragments. Assembly of more than nine fragments can be performed efficiently using two successive cloning steps rather than one [17].

5. Assembly of PCR products directly in the destination vector without first cloning them as intermediate constructs is possible. It is however recommended to purify the PCR products using a column to remove any DNA polymerase left in the PCR product and to remove primer dimers that may be produced during PCR amplification. Indeed some of the primer dimers are flanked by two fusion sites (these are part of the primers) and can therefore be cloned, resulting in incorrect constructs. The final constructs may also contain PCR-induced mutations.

6. For generation of entry clones, the steps consisting of PCR amplification, cloning, and sequencing can be replaced by ordering the desired sequence from a gene synthesis company. The fragment to be synthesized should be ordered directly with the appropriate type IIS restriction sites flanking the sequence of interest. Also, it is useful to make sure upon ordering that the cloning vector in which the ordered DNA fragment will be cloned does not contain additional sites for the type IIS enzyme chosen, and has an antibiotic resistance gene different from the one in the vector that will be used for assembly.

7. Basically, two primers overlapping an internal BsaI site are made, one in each orientation, with a mismatch designed to introduce a silent mutation in the type IIS restriction site. Two separate PCRs are performed with primers designed to amplify the two halves of the module. The PCR products are purified on a column, and a mix of both is used as a template for a second PCR performed using both flanking primers only (the two primers flanking the given module). This PCR is purified on a column, cloned, and sequenced.

8. For small modules of up to 80 nucleotides, PCR amplification does not necessarily require a DNA template. For example, two

complementary primers can be ordered covering the entire sequence of the module (including the flanking type IIS restriction sites). Both primers are annealed in water and directly used for blunt-end cloning in the cloning vector. For larger but still small modules, two overlapping primers can be ordered that are complementary at their 3′ end on a length of 20–25 nucleotides. A double-stranded DNA fragment can be obtained by performing a PCR with both primers without a template. In theory, one single PCR cycle should be sufficient, but using 35 cycles as for normal PCR also works well.

9. One restriction for blunt-end cloning is that the ends of the primer should not recreate a SmaI site (or any other blunt-end restriction site used for cloning) after ligation of the PCR product (i.e., the DNA fragment to be cloned should not start with the sequence GGG or finish with CCC). A second restriction is that the fragment to be cloned should not contain an internal restriction site for the enzyme used for cloning. If this is the case, another enzyme should be chosen for cloning, for example, EcoRV (a cloning vector containing a unique EcoRV site in the polylinker should be used).

10. The presence of a BsaI site in the vector backbone of the entry modules does not prevent from using them for Golden Gate shuffling, since plasmids containing the final shuffled sequences should not contain this vector backbone. However, the presence of such a site in all entry constructs would lead to continuous ligation and redigestion at this site, and would therefore unnecessarily consume some ATP from the ligation mix, at the expense of the desired ligation events.

11. The widely used pUC19 vector also contains a BsaI site in the ampicillin resistance gene. A simple strategy, enzymatic inverse PCR [23], can be used to eliminate the internal BsaI site in pUC19. The entire plasmid can be amplified with two primers overlapping with the BsaI site: primers bsarem1 (ttt ggtctc a ggtt ctcgcggtatcattgcagc) and bsarem2 (ttt ggtctc a aacc acgctcaccggctccag). These primers are designed to introduce a single silent nucleotide mutation in the BsaI recognition site in the vector. The primers are themselves flanked by two BsaI restriction sites that form two compatible overhangs after BsaI enzyme digestion. After amplification of the entire plasmid with both primers, the PCR is purified with a column (to remove remaining polymerase and nucleotides). The linear fragment is subjected to restriction-ligation using BsaI and ligase, and transformed in *E. coli*.

12. In practice, if all module plasmids and the vector have approximately the same size (4–5 kb), simply adding 100 ng of DNA of each module set and of the vector will work relatively well. However, if plasmids with widely different sizes are used,

calculating an equimolar amount should provide a higher cloning efficiency. The following formula (from the NEB catalog) can be used: 1 μg of a 1,000-bp DNA fragment corresponds to 1.52 pmol.

13. We have found that both types of programs work well when high concentration ligase is used, but both programs can be tested in parallel by the users to optimize ligation efficiency.
14. Any other *E. coli* strain can also be used. If higher transformation efficiency is required, the restriction-ligation mix can be transformed in electrocompetent *E. coli* cells. In this case, DNA from the restriction-ligation mix should first be ethanol-precipitated and resuspended in 10 μL of water.

Acknowledgments

The authors would like to thank Dr. Stefan Werner and Dr. Ernst Weber for critical reading of this manuscript.

References

1. Ellis T, Adie T, Baldwin GS (2010) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 3:109–118
2. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
3. Gibson DG, Benders GA et al (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc Natl Acad Sci U S A 105:20404–20409
4. Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:e16
5. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3: e3647
6. Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4:e5553
7. Lebedenko EN, Birikh KR, Plutalov OV, Berlin YA (1991) Method of artificial DNA splicing by directed ligation (SDL). Nucleic Acids Res 19:6757–6761
8. Szybalski W, Kim SC, Hasan N, Podhajska AJ (1991) Class-IIS restriction enzymes—a review. Gene 100:13–26
9. Berlin YA (1999) DNA splicing by directed ligation (SDL). Curr Issues Mol Biol 1:21–30
10. Lu Q (2005) Seamless cloning and gene fusion. Trends Biotechnol 23:199–207
11. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765
12. Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S (2012) Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng Bugs 3:38–43
13. Sanjana NE, Cong L et al (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7:171–192
14. Morbitzer R, Elsaesser J, Hausner J, Lahaye T (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res 39:5790–5799
15. Cermak T, Doyle EL et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82
16. Geißler R, Scholze H et al (2011) Transcriptional activators of human genes with programmable DNA-specificity. PLoS One 6:e19509
17. Weber E, Gruetzner R et al (2011) Assembly of designer TAL effectors by golden gate cloning. PLoS One 6:e19722

18. Horton RM, Ho SN et al (1990) Gene splicing by overlap extension. Biotechniques 8:528–535

19. Bolchi A, Ottonello S, Petrucco S (2005) A general one-step method for the cloning of PCR products. Biotechnol Appl Biochem 42:205–209

20. Liu ZG, Schwartz LM (1992) An efficient method for blunt-end ligation of PCR products. Biotechniques 12:28–30

21. Kotera I, Nagai T (2008) A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme. J Biotechnol 137:1–7

22. Sarrion-Perdigones A, Falconi EE et al (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 6:e21622

23. Stemmer WP, Morris SK (1992) Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. Biotechniques 13:214–220

Chapter 13

Construction of Synthetic Gene Circuits in the *Escherichia coli* Genome

Bei-Wen Ying, Yuya Akeno, and Tetsuya Yomo

Abstract

The construction of stable and functional synthetic circuits in bacteria is necessary in the areas of systems and synthetic biology. The common approach using plasmids to carry foreign genetic circuits offers convenience in genetic construction but is poor in genetic stability (e.g., variation in copy number). Genome recombination provides the stable genetic maintenance of synthetic circuits, but is often labor intensive and time consuming when the genetic circuits are complex and the DNA fragments become larger. The method introduced here is modified from that reported by Wanner's group and is available for integration of complex genetic circuits into the *Escherichia coli* chromosome.

Key words Homogenous recombination, Synthetic gene circuit, *Escherichia coli*, Genome replacement, Dual antibiotic selection

1 Introduction

Means for efficient construction of synthetic circuits that play biological roles in living cells are in great demand. Successful examples of synthetic circuits in plasmid format have been reported functional in bacteria [1, 2]. However, comparison of the cellular concentrations of proteins encoded by genes carried on plasmids or the genome indicated that the plasmid format showed larger copy number noise in synthetic circuits [3]. Reduced fluctuation in copy number could be achieved by genome integration and would lead to accurate phenotypic behavior of the synthetic switch corresponding to the design principle [3]. The classic methods for genome integration, the so-called allele replacement methods, have been used to inactivate bacterial chromosomal genes [4–7], but required creating the gene disruption on a suitable plasmid before recombining it onto the chromosome. A relatively simple method to disrupt chromosomal genes in *Escherichia coli* was developed by Wanner's group, in which PCR primers provide homology to the target genes [8]. The method was based on the Red system that

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2_13, © Springer Science+Business Media New York 2013

promoted a greatly enhanced rate of recombination over those exhibited by recBC, sbcB, or recD mutants when using linear DNA. Wanner's method has been used in the construction of single-gene knockout mutants [9] and systematic mutagenesis of the *E. coli* genome [10]. Nevertheless, the reported protocol for construction of complicated synthetic networks is limited to constructs 2–3 kb in length [11].

For the precise design and interpretation of cellular networks, it is necessary to construct relatively large fragments and complex gene circuits. However, there is no well-established methodology for connecting the designed individual components in the genome to accomplish the desired function [12]. The protocol introduced here is a modification of the original method with regard to medium, temperature, transformation, and selection, as described elsewhere [3]. Linear foreign sequences of various lengths were easily integrated into the *E. coli* genome simultaneously. The refined recombination procedure allowed efficient construction of a single copy of complex genetic circuits in cells, offering both stability of genetic circuits and accessibility for complex construction. This method has been recently used for the efficient integration of synthetic circuits into the bacterial genome for stochastic studies and the artificial engineering of genetic networks [13, 14]. The availability of long-fragment inserts facilitated the reconstruction of complex networks on the genome and provided a powerful tool for precise engineering in synthetic and systems biology.

2 Materials

2.1 Strains and Plasmids

Most *E. coli* strains can be used for genetic construction. Here, the strains DH1 and JM109 were used for genome replacement and plasmid harvest, respectively. The plasmid pKD46, carrying the genes encoding the recombinase, was used for all recombination and/or gene deletions. The plasmids of the pBR series generally carry an ampicillin resistance gene (Amp^r) and contain the DNA fragments involved in the synthetic gene circuits. Another antibiotic resistance gene (i.e., Cm^r) must be included in the DNA fragment for positive selection. The sources of the strains and plasmids were as follows:

1. *E. coli* DH1 (National BioResource Project, National Institute of Genetics, Japan).
2. *E. coli* JM109, Electro-cell (TaKaRa).
3. pKD46 (from Dr. Wanner).
4. pBR series (preliminary constructed by customs, Amp^r and Cm^r).

2.2 Media and Reagents

1. Chloramphenicol solution: 30 mg/ml stock in 100 % ethanol, sterile-filtered.
2. Ampicillin solution: 50 mg/ml stock in water, sterile-filtered.
3. LB (Luria-Bertani) medium: Dissolve 10 g of Bacto Tryptone, 5 g of Yeast Extract, and 10 g of NaCl in 900 ml of ddH_2O. Add water to 1,000 ml. Sterilize by autoclaving for 20 min.
4. LB ampicillin (50 μg/ml) plates: Dissolve 4 g of Bacto Tryptone, 2 g of Yeast Extract, 4 g of NaCl, and 6 g of agar in 400 ml of deionized water. Sterilize by autoclaving for 20 min. Let LB agar cool to about 50 °C. Add 400 μl of ampicillin solution. Pour about 20 ml into each Petri dish.
5. LB chloramphenicol (30 μg/ml) plates: Dissolve 4 g of Bacto Tryptone, 2 g of Yeast Extract, 4 g of NaCl, and 6 g of agar in 400 ml of deionized water. Sterilize by autoclaving for 20 min. Let LB agar cool to about 50 °C. Add 400 μl of chloramphenicol solution. Pour about 20 ml into each Petri dish.
6. SOB medium: Dissolve 4 g of Bacto Tryptone, 1 g of Yeast Extract, and 0.12 g of NaCl in 180 ml of ultra-pure water. Add 0.25 ml of 2 M KCl. Adjust to pH 7.0 using 8 M NaOH and add water to 200 ml. Sterilize by autoclaving for 20 min. Before using, add 20 ml of sterile 1 M $MgCl_2$.
7. SOC medium: Add 0.4 ml of sterile 2 M glucose to 20 ml of sterile SOB.
8. 10 % glycerol solution, sterile.
9. 60 % glycerol solution, sterile.
10. Go Taq® (Promega): used for colony PCR.
11. Phusion® (Finnzymes): used for high fidelity amplification of DNA templates.
12. 1 % agarose gel in 1× TAE buffer.
13. 6× Loading Buffer (TaKaRa).
14. Ethidium bromide (EtBr).
15. 1× TAE (Tris–acetate–EDTA) buffer: Make 50× TAE for stock solution; dissolve 242 g of Tris–HCl in 25 ml of 0.5 M EDTA, 57.1 ml of glacial acid, and approximately 750 ml of deionized water and add water to 1,000 ml. Dilute 50× stock solution with deionized water.
16. *Dpn*I (NEB) and NEB 4 buffer: used for digestion of plasmids.
17. 1.3 M arabinose.
18. Primers for genetic construction and genome replacement: customized and commercially ordered.

2.3 DNA Purification Kits

The plasmids and PCR products are generally purified using commercially available kits as follows. The purification kits from other manufacturers, e.g., Promega, also show good performance. The purification and extraction procedures are carried out essentially according to the manufacturer's protocols.

1. QIAprep® spin Miniprep kit (Qiagen).
2. QIAquick® PCR Purification kit (Qiagen).
3. MinElute® PCR Purification kit (Qiagen).
4. QIAquick® Gel Extraction kit (Qiagen).

2.4 Apparatus and Tools

The following instruments and equipment are required for gene transduction, electrophoresis, cell culture and incubation, PCR, sample collection and detection.

1. Gene Pulser® Xcell (Bio-Rad).
2. Gene Pulser® Cuvette 0.1 cm electrode gap, pkg. of 50 (Bio-Rad).
3. Electrophoresis apparatus: Mupid®-2plus (Advance).
4. Air incubator: IS600 (Yamato).
5. Water bath shaker: Personal 11 (Taitec).
6. Centrifugal machine: Micro Refrigerated Centrifuge 3,740 (Kubota).
7. Thermal cycler: GRADIENT PCR (TaKaRa).
8. UV light box: BioDoc-It® Imaging System (UVP).

3 Methods

The pBR plasmids carrying the parts or full structure of the synthetic circuit must be constructed in advance (*see* **Notes 1–3**). The antibiotic resistance gene, other than ampicillin resistance, should be included in the genetic parts for genome replacement. The flow chart of the complete genome replacement procedure is illustrated in Fig. 1. Following PCR amplification and purification of the linear target sequence, transformation (electroporation) for genome replacement was performed to introduce the construct into competent cells. To distinguish genomic recombinants from the original plasmid carriers (pBR series), the synthetic sequence generally encoded a different antibiotic resistance gene (e.g., Cm^r) from the original plasmid (e.g., Amp^r). False transformants (i.e., transformed colonies) carrying the plasmid grow on both antibiotic plates; genomic recombinants grow only on the plate carrying the antibiotic the resistance gene for which is encoded in the circuit, and not that encoded on the plasmid. Dual-antibiotic selection for positive transformants reduced the labor and costs of large-scale screening, and uncovered a high ratio of positive candidates on the colony PCR check.

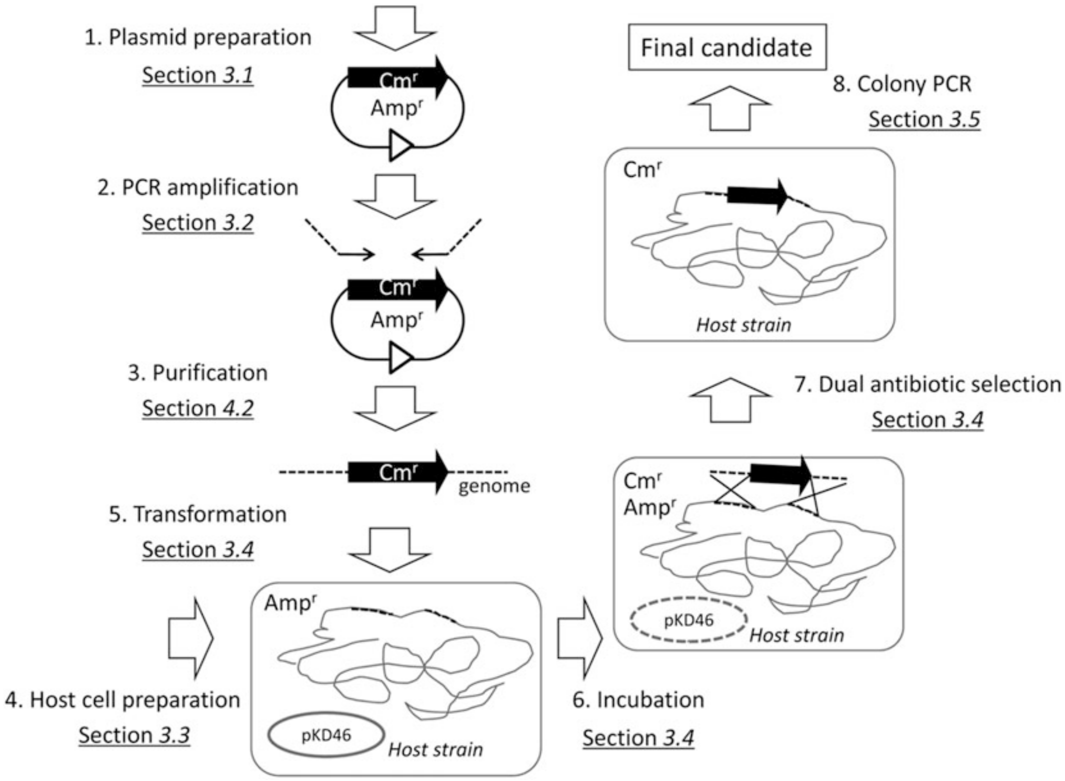


Fig. 1 Overview of the refined method. All eight steps are described in detail in Subheading 3. The *broken lines* indicate the homologous sequences. Modified from Ying et al. [3]

3.1 Plasmids Amplification and Purification

1. Add 0.2–2 μl of pBR plasmids (approximately 1–50 ng) carrying the target DNA fragment (*see* **Note 4**) to 50 μl of *E. coli* JM109 Electro-cells and mix by tapping. Then, chill on ice for 30 min.
2. Transfer to Gene Pulser® Cuvette 0.1 cm electrode and electroporate the cells at 1,800 V, 25 μF, 200 Ω, with Gene Pulser Xcell.
3. Immediately following electroporation, add 900 μl of SOC.
4. Transfer to 1.5-ml microfuge tube and incubate at 37 °C for 0.5–1 h.
5. Plate 10–100 μl on LB ampicillin plate.
6. Incubate overnight at 37 °C.
7. Pick a single colony into 5 ml of LB containing 30 μg/ml of chloramphenicol.
8. Place in a shaker at 37 °C, 160 min^{-1} and grow overnight.
9. Transfer to 1.5-ml microfuge tube and centrifuge for 1 min at 17,900 × *g*. Discard the flow-through fraction.

10. Transfer remaining culture to the same tube and centrifuge for 1 min at 17,900 × *g*. Discard the flow-through fraction. Repeat until all of the culture is used.
11. Purify plasmid DNA using the QIAprep® spin Miniprep kit, according to the QIAprep Miniprep Handbook second edition.
12. Store the purified plasmid at –30 °C.

3.2 Preparation of DNA Fragments for Homologous Recombination

1. Mix the following reagents together on ice.

   Note: Always add enzyme last. For multiple samples, make a large master mix and divide equally into each PCR tube.

5× HF Buffer	10 μl
2.5 mM dNTPs	4 μl
Forward primer (10 pmol/μl)	1.5 μl
Reverse primer (10 pmol/μl)	1.5 μl
Template (plasmid)	1–10 ng
Phusion polymerase	0.5 μl
Add ddH$_2$O to	50 μl

2. PCR amplification is performed with initial denaturation at 98 °C for 30 s, followed by 25 cycles consisting of 10 s at 98 °C for denaturation, 30 s at 60 °C for annealing, and 30 sec (per 1 kb length) at 72 °C for extension, with a final extension step of 5 min at 72 °C.
3. Add 1 μl of 6× Loading Buffer (TaKaRa) to 5 μl of PCR products and mix.
4. Electrophorese the samples through 1 % agarose gel in 1× TAE buffer at 100 V for about 40 min.
5. Soak the gel in the EtBr bath for 20 min.
6. Check DNA bands using a UV light box.
7. Purify the remaining PCR products using QIAquick PCR Purification kit, according to QIAquick Spin Handbook second edition.

   Note: If multiple bands are detected, DNA purification by QIAquick Gel extraction kit is required, according to QIAquick Spin Handbook second edition.
8. Mix 44 μl of purified samples, 5 μl NEB 4 Buffer (10×) and 1 μl of *Dpn*I. Then, incubate at 37 °C for 2 h.
9. Purify DNA fragments using MinElute® PCR Purification kit, according to MinElute® Handbook second edition.
10. Store the purified DNA at –30 °C.

3.3 Preparation of Competent Cells for Genome Recombination

1. Inoculate the *E. coli* DH1 host cells into 5 ml of LB.
2. Place in a shaker at 37 °C, 160 min^{-1} and grow overnight.
3. Inoculate fresh culture into 5–50 ml (depending on requirements) of SOB to $\text{OD}_{600} \approx 0.1$.
4. Place in a water bath shaker at 37 °C, 160 min^{-1}, for 3–4 h ($\text{OD}_{600} = 0.8$).
5. Transfer to 15-ml tube and chill on ice for 10 min.
6. Centrifuge for 3 min at 7,500 × *g* at 4 °C. Discard the flow-through fraction.
7. Gently resuspend the cells in 0.1–1 ml of ice-cold 10 % glycerol.
8. Centrifuge for 3 min at 7,500 × *g* at 4 °C. Discard the flow-through fraction.
9. Repeat **steps 5** and **6**.
10. Gently resuspend the cells in 40–400 μl of ice-cold 10 % glycerol.
11. Aliquot 40 μl into each 1.5-ml microfuge tube, and store the competent cells at –80 °C (*see* **Note 5**).

Introduce pKD46 into the Host Cells

12. Add 1 μl of pKD46 to 40 μl of competent cells from –80 °C stock (**step 11**) and mix by tapping. Then, chill on ice for 30 min.
13. Transfer to Gene Pulser® Cuvette 0.1 cm electrode gap and electroporate the cells at 1,800 V, 25 μF, 200 Ω, with Gene Pulser Xcell.
14. Immediately following electroporation, add 900 μl of pre-warmed SOC.
15. Transfer to 1.5-ml microfuge tube to an air incubator at 30 °C for 2 h.
16. Plate 10–100 μl of the cells on LB ampicillin plate.
17. Incubate overnight at 30 °C.
18. Pick a single colony into 5 ml of LB with ampicillin (50 μg/ml).
19. Place in water bath shaker at 30 °C, 160 min^{-1} and grow overnight.
20. Inoculate fresh culture into 5–50 ml of SOB with ampicillin (50 μg/ml) and 50 μl of 1.3 M arabinose to $\text{OD}_{600} \approx 0.1$.
21. Place in a shaker at 30 °C, 160 min^{-1}, for 3–4 h ($\text{OD}_{600} = 0.8$).
22. Prepare competent cells as described in **steps 5–11**.

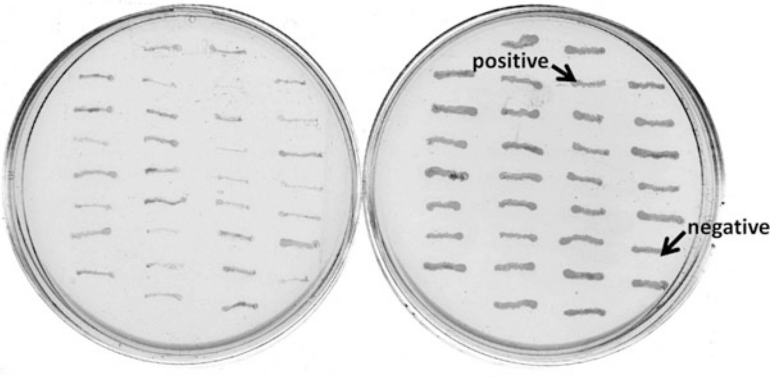


Fig. 2 Dual antibiotic selection. The transformed colonies were transferred onto LB plates containing ampicillin (*left*) or chloramphenicol (*right*). Positive colonies only grew on LB chloramphenicol, but negative colonies grew on both plates (*arrows*)

3.4 Genome Recombination and Dual Antibiotic Screening

1. Add 2 μl of DNA fragments (synthetic gene circuits) prepared in Subheading 3.2 to 40 μl of competent cells (host cells carrying pKD46) prepared in Subheading 3.3 and mix by tapping. Then, chill on ice for 30 min.
2. Transfer to Gene Pulser® Cuvette 0.1 cm electrode and electroporate the cells at 1,800 V, 25 μF, 200 Ω, with Gene Pulser Xcell.
3. Immediately following electroporation, add 900 μl of prewarmed SOC (*see* **Note 6**).
4. Transfer to 1.5-ml microfuge tube and incubate without shaking at 37 °C for 2 h.
5. Plate 100–300 μl on LB chloramphenicol plate (*see* **Note** 7).
6. Incubate overnight at 37 °C.
7. Strike the single colonies from the plate in **step 5** onto both LB ampicillin plate (as check plate) and LB chloramphenicol plate (as master plate).
8. Place the LB ampicillin plate in the air incubator at 30 °C and the LB chloramphenicol plate at 37 °C. Grow overnight.
9. Selection based on the difference in growth of colonies between the two plates: positive candidates exhibit rapid growth on the LB chloramphenicol plate and slow or no growth on the LB ampicillin plate (Fig. 2).

3.5 Colony PCR and Final Choice of Positive Clones (See Note 8)

1. Suspend the selected colonies in 100 μl ddH$_2$O.
2. Mix the following reagents together on ice.

   Note: For multiple samples, make a large master mix and divide equally into each PCR tube.

Go Taq	10 μl
Forward primer (10 pmol/μl)	2 μl
Reverse primer (10 pmol/μl)	2 μl
Cell suspension	6 μl
Total	20 μl

3. PCR amplification is performed with initial denaturation at 95 °C for 2 min, followed by 25 cycles consisting of 30 s at 95 °C for denaturation, 30 s at 60 °C for annealing, and 4 min at 72 °C for extension (1 kb/min), with a final extension step of 5 min at 72 °C.
4. Load 5 μl of PCR products on a 1 % agarose gel.
5. Run the gel at 100 V for about 40 min.
6. Soak the gel in the EtBr bath for 30 min.
7. Check DNA bands using a UV light box. The positive colonies show a PCR product of the correct size.
8. Streak the positive colonies onto a new LB chloramphenicol plate to form single colonies.
9. Incubate overnight at 37 °C.
10. Strike the single colonies from the plate in **step 8** onto both LB ampicillin and LB chloramphenicol plates for dual antibiotic selection (repeat **steps 8–11** until no growth occurs on LB ampicillin plate).
11. Pick a single colony from the chloramphenicol plate into 5 ml of LB with chloramphenicol (30 μg/ml).
12. Place in the water bath shaker at 37 °C, 160 min^{-1} and grow overnight.
13. Mix 750 μl of fresh culture and 250 μl of 60 % glycerol. Store at –80 °C.

4 Notes

1. The homologous sequences, i.e., the complementary DNA fragments up- and downstream of the target synthetic circuit for genome recombination, were usually as long as ~50 bp and could be easily generated by PCR amplification (Subheading 3.2). As discussed previously [3], longer overlaps, e.g., 100–150 bp, often provided better performance and larger numbers of positive colonies. Generally, larger fragments for genome integration preferred longer homologous overlaps. Note that overly long overlaps sometimes resulted in poorer

output. If the overlap is longer than 200–300 bp, subcloning of these fragments on the plasmid carrying the target synthetic circuit is strongly recommended, as repeated PCR amplification would cause unpredicted mutations in the target synthetic circuits.

2. As the functional behavior of the synthetic circuit is required in most studies, reducing the unknown effect from the native regulation in the host cell must be considered. Theoretically, either identical or reverse direction in gene expression framework of synthetic insertion is available. As the upstream native regulation may play a role in expression of the integrated synthetic circuit, the inverse direction relative to the native chromosomal site is suggested, except in cases where the native regulation is under investigation. Thus, the homologous overlap forward in the genome is reverse on the plasmid, and vice versa.
3. The chromosomal location for recombination can be freely decided by the research purpose, but it greatly determines the efficiency or productivity of genome integration. The native expression level of the gene site for recombination often governs the base expression level of the synthetic circuit inserted into that site. Any trace transcriptional leakage or the balance of gene expression among varied chromosomal sites could deterministically eliminate the physiological function from a genetically successful construct. In addition, it is easier to incorporate synthetic circuits into the chromosomal sites for transposons, e.g., *intC*, but the integration is unstable. The site of *galK* is highly recommended for genome recombination due to its genetic stability and moderate native expression level.
4. The pBR series plasmids introduced here have copy numbers of 20–30 per cell. Although there is no limitation in copy number of the plasmids for carrying the synthetic circuits, a relatively low copy number is recommended, particularly when a strong promoter is used; high expression level of the foreign gene circuits is sometimes toxic to the host cell and leads to failure of plasmid construction and harvest.
5. The protein level of recombinase in competent cells carrying the plasmid pKD46 is important for recombination efficiency. Although the protein concentration is difficult to evaluate, the following steps can be customized to adjust the recombination efficiency: repeated transfer of cells carrying pKD46 on LB chloramphenicol plates before preparation of competent cells; extra addition of arabinose in the preculture for preparation of competent cells.
6. In the transformation step for genome recombination, the inducer arabinose can also be added in the pre-warmed SOC.

The 2-h incubation time can be extended to overnight or 2 days. These modifications generally increase the number of transformed colonies.

7. The concentrations of antibiotics for selection of positive recombinants introduced here are in accordance with the common use. However, in the case of poor recombinant efficiency, a half dosage of the antibiotics is suggested for the transformation step. After positive selection, applying the common dosage of antibiotics in the transferred master plates is practical.
8. The antibiotic selection of transformed colonies is often used as the only experimental verification for positive transformants. However, some unknown and undetectable reactions also occur in cells during genome recombination. False positive colonies commonly occur on antibiotic selection. Thus, PCR-based genetic confirmation is required to confirm the correct synthetic construction as designed.

Acknowledgements

This work was partially supported by Grants-in-Aid for Challenging Exploratory Research 22657059 (to B.W.Y.) and the "Global COE (Centers of Excellence) program" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403:339–342
2. Kashiwagi A, Urabe I, Kaneko K, Yomo T (2006) Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. PLoS One 1:e49
3. Ying BW, Ito Y, Shimizu Y, Yomo T (2010) Refined method for the genomic integration of complex synthetic circuits. J Biosci Bioeng 110:529–536
4. Dabert P, Smith GR (1997) Gene replacement with linear DNA fragments in wild-type Escherichia coli: enhancement by Chi sites. Genetics 145:877–889
5. Kato C, Ohmiya R, Mizuno T (1998) A rapid method for disrupting genes in the Escherichia coli genome. Biosci Biotechnol Biochem 62:1826–1829
6. Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. J Bacteriol 179:6228–6237
7. Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR (1999) Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res 27:4409–4415
8. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645
9. Baba T, Ara T, Hasegawa M et al (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008
10. Kang Y, Durfee T, Glasner JD et al (2004) Systematic mutagenesis of the Escherichia coli genome. J Bacteriol 186:4921–4930
11. Serra-Moreno R, Acosta S, Hernalsteens JP et al (2006) Use of the lambda Red recombinase system to produce recombinant prophages

carrying antibiotic resistance genes. BMC Mol Biol 7:31

12. Haseltine EL, Arnold FH (2007) Synthetic gene circuits: design with directed evolution. Annu Rev Biophys Biomol Struct 36:1–19
13. Shimizu Y, Tsuru S, Ito Y et al (2011) Stochastic switching induced adaptation in a starved *Escherichia coli* population. PLoS One 6: e23953
14. Matsumoto Y, Ito Y, Tsuru S et al (2011) Bacterial cells carrying synthetic dual-function operon survived starvation. J Biomed Biotechnol 2011:e489265

Chapter 14

Shuffling of DNA Cassettes in a Synthetic Integron

David Bikard and Didier Mazel

Abstract

The complexity of even small gene networks makes them hardly amenable to rational design. Testing random combinations of genetic elements in a directed evolution procedure is thus of interest for many applications including metabolic engineering. Here we describe how the recombination machinery of class 1 integrons can be used to deliver and shuffle genetic elements at a chromosomal locus in *E. coli*.

Key words Synthetic biology, Directed evolution, Recombination, DNA shuffling, Integron, Metabolic engineering

1 Introduction

Over the past 20 years, efficient mutagenesis techniques have been developed allowing the introduction of random point mutations, as well as sequence shuffling through homologous recombination. However the efficient random recombination of heterologous sequences largely remains a challenge. It has been achieved in vitro through various techniques of randomized assembly ligation and overlap extension PCR [1–5]. However these techniques are limited in the size and number of elements they can combine.

Integrons are natural recombination platforms involved in the rapid adaptation of bacteria to changing environment. They are present in the genome of roughly 10 % of sequenced bacteria, and are largely responsible for the massive spread of multiple resistances to antibiotics. Integrons consist of an integrase gene (*intI*) able to capture, stockpile, and shuffle gene cassettes. Natural cassettes generally consist of a single promoterless ORF associated to an *attC* recombination site. Cassettes can be excised through *attC* × *attC* recombination into a circular intermediate and preferentially integrated at a primary recombination site downstream of the integrase gene called *attI*. Different classes of integrons have been identified on the base of the integrase sequence identity (for a comprehensive review, *see* ref. 6).

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_14, © Springer Science+Business Media New York 2013

We use here the class 1 integrase isolated from R388 multiresistant integron, the associated *attI1* site, and the $attC_{aadA7}$ site isolated from an *aadA7* spectinomycin resistance cassette. These elements have been extensively described and shown to efficiently recombine. We have already shown how this machinery can be used to deliver gene cassettes of the tryptophan operon and shuffle them to generate a wide variety of combination yielding improved phenotypes [7]. Here is a detailed procedure of the steps required to construct a strain with a synthetic integron integrated in the chromosome, to deliver new genetic elements in the array, and to shuffle them.

2 Materials

The minimal platform of an integron consists of the integrase gene and the *attI* recombination site. In all experiments, the IntI1 integrase is expressed in *trans* from the pBAD18::*intI1* plasmid. The *attI1* site is integrated into the chromosome at the *attB* site of bacteriophage lambda using a suicide integration vector (procedure described below). Depending on your application, you may want to directly clone cassettes in the integron array that you will integrate into the chromosome.

2.1 Biological Material: DNA

1. Plasmid pSWK carrying an *attI* site. The pSWK plasmid carries an R6K origin of replication which depends on the Π protein expressed from the *pir* gene. It also carries the *attP* site from bacteriophage lambda, allowing its integration at the lambda *attB* site in the *E. coli* chromosome. This plasmid complies with biobrick standard RFC10 and can be found on the registry of standard biological parts as BBa_J99000 [7].
2. The integron $attC_{aadA7}$ site: BBa_J99001 [7].
3. The integron *attI1* site: BBa_J99002 [7].
4. Plasmid pBAD18::*intI1* [8].
5. Plasmid pTSA29-CXI-AK [9]. A thermo-sensitive pSC101 vector carrying the lambda integrase under the control of the thermo-inducible CI857/pR promoter.

2.2 Biological Material: Strains

1. pi1 *E. coli* cells (*DH5α ΔthyA*::(*erm-pir116*) [EmR]), for the replication of the pSWK plasmid [10].
2. MG1655 *E. coli* cells (or any *E. coli* strain where you want to integrate the synthetic integron).
3. ω7249 *E. coli* cells ((*F-*) *RP4-2-Tc*::*MuΔnic35 ΔdapA*::(*erm-pir*) [KmR ErmR]). ω7249 cells express the *pir* gene allowing the replication of vectors carrying the R6K replication origin. It also expresses the RP4 conjugation machinery allowing the transfer of plasmids carrying the RP4 origin of transfer [11].

2.3 Culture Media and Reagents

1. Luria Bertani (LB) culture medium.
2. LB agar plates supplied with suitable selection agent.
3. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 μg/ml, chloramphenicol (Cm), 25 μg/ml.
4. Thymidine (Thy) and diaminopimelic acid (DAP) is supplemented when necessary to a final concentration of 0.3 mM.
5. Glucose and L-arabinose (Ara) are added at, respectively, 10 and 2 mg/ml final concentration for the repression and induction of the pBAD promoter.
6. $CaCl_2$ solution, 50 μM.
7. EcoRI, XbaI, SpeI, PstI restriction enzymes. Those are the restriction sites of the RFC10 assembly standard.
8. T4 DNA ligase.

3 Methods

The global idea is to first integrate the *attI1* site into the chromosome. This will be the landing platform of any integron cassette of your choice. Cassettes present in the array can be shuffled through the expression in *trans* of the IntI1 integrase from the pBAD18::*intI1*. Finally, new cassettes can be delivered to the array through a suicide conjugation assay.

Depending on your application, an initial cassette array can directly be integrated together with the *attI* site. Cassettes are simply created through the association of any DNA element to an $attC_{aadA7}$ recombination site and cloned downstream of the *attI* site. This can be done using the construction method of your choice. The method we use here is biobrick standard assembly. We constructed vectors carrying *attI* (BBa_J99002) and *attC* sites (BBa_J99001) complying with the RFC10 standard.

Several methods have been described for the efficient integration of DNA in the *E. coli* chromosome [12, 13]. The method we chose here is to clone the initial integron cassette array on a vector (pSWK) that can then be integrated at the chromosomal *attB* site of bacteriophage lambda. pSWK carries an *attP* site and can thus be integrated at the *attB* site by the lambda integrase. It has an R6K origin of replication and can only replicate in cells expressing the *pir* gene. This plasmid can thus be integrated into *pir*-cell expressing the lambda integrase in *trans* (from plasmid pTSA29-CXI-AK), and plating on the pSWK-resistant marker will select integration events. However, any other integration method and any other neutral chromosomal locus should be suitable.

New cassettes can be delivered to an integron array through a suicide conjugation strategy. The pSWK vector can be used to this end too. pSWK carries an RP4 origin of transfer and can be delivered to recipient cells by ω7249 cells expressing the RP4 conjugation machinery. This donor strain can be used in a conjugation assay to deliver cassettes in a recipient strain carrying an integron array and expressing the IntI1 integrase in *trans* from the pBAD18::*intI1*.

3.1 Construction of the Initial Cassette Array

1. The initial integron array can be constructed with the method of your choice. Here, we use a classic digestion/ligation/transformation protocol. The use of the RFC10 idempotent assembly standard allows the easy association of any biobrick to an $attC_{aadA7}$ site and its cloning into the integron array on the pSWK vector. For more details on biobrick assembly, see partsregistry.org. The pSWK vector can only replicate in cells expressing the *pir* gene. Transformation is thus carried out into pi1 cells, and selection on LB agar + Cm + Thy plates.

3.2 Integration of the Initial Cassette Array into the Chromosome

1. You first need to transform the pTSA29-CXI-AK plasmid carrying the lambda integrase into your strain. The method of your choice can be used to prepare competent cells. We use here a classical $CaCl_2$ chemo-competent protocol. After heat-shock, cells need to be incubated at 30 °C for the expression period and plated on LB agar + Ap at 30 °C in order for the plasmid to stably replicate.
2. In order to integrate your pSWK into these cells, you now need to prepare competent cells expressing the lambda integrase. Launch an overnight culture at 30 °C in LB + Ap from a single colony.
3. Dilute the overnight culture to the 1/100th in LB and grow at 30 °C until $OD_{600} = 0.2$.
4. Switch the incubation temperature to 42 °C until $0.6 < OD_{600} < 0.8$. During this period, the lambda integrase will be expressed from the CI857-PR promoter.
5. Prepare $CaCl_2$ competent cells (or any other type of competent cells. The integration event is very efficient, so you don't need to prepare ultra-competent cells for this to work) (*see* **Note 1**).
6. Transform with at least 20 ng of pSWK DNA and incubate for the expression time for 1H at 42 °C.
7. Plate on LB agar + Cm and grow overnight at 42 °C. Doing all this steps at 42 °C ensures that the lambda integrase is expressed and that at the same time the pTSA29 is lost.
8. The following day you should have colonies with the pSWK integrated at the *attB* site of your strain.
9. Check for proper integration by doing a colony PCR using a primer binding on the pSWK and one binding on the chromosome.

3.3 Cassette Shuffling Through the Intl1 Integrase Expression

1. Transform the pBAD18::*intI1* in your cells using the method of your choice.
2. Induce the integrase by diluting an overnight culture to the 1/100th in LB + Ara. You can prepare a negative control in LB + Glu. Grow overnight (*see* **Note 2**).
3. Plate cells on the selective media of your choice. Recombinations induced by the integron machinery happen at frequencies below 10^{-2}. In order to find recombined clones, you need a powerful screening assay. In our initial study where we shuffle the genes of the tryptophan biosynthesis operon, the selection is performed in tryptophan auxotroph strains and recombinants are selected on minimal media [7].

3.4 Delivery of New Cassettes Through Conjugation

1. Transform the pSWK carrying the cassettes of interest into ω7249 cells. We use $CaCl_2$ competent cells, but any method is suitable. Remember that ω7249 cells need DAP to grow (*see* **Note 3**).
2. This strain will be your donor strain in the conjugation assay. Depending on your application, you may want to use several donor strains to combinatorially deliver cassettes of interest in a single recipient strain. The recipient strain needs to carry the pBAD18::*intI1* plasmid and an *attI* site in the chromosome. Dilute an overnight culture of the donor strain(s) to the 1/100th in LB + DAP and of the recipient strain in LB + DAP + Ara (*see* **Note 4**).
3. Grow donor and recipient cells to $0.4 < OD_{600} < 0.6$.
4. Mix 1 ml of donor cells with 4 ml of recipient cells and filter onto a 0.22 μm membrane (*see* **Note 5**).
5. Place the filter on a fresh LB + DAP + Ara plate, the side with the cells facing up. Incubate 6H at 37 °C.
6. Put the filter in a 50 ml tube with 5 ml LB and vortex to remove the cells from the filter.
7. Plate the cells on the selective media of your choice. (If the pSWK vector you use carries a resistance that is not already present in your recipient strain, you can select for the integration of the whole vector in the integron array. It is then possible to get rid of the vector backbone using a counter selection marker, leaving all or some of the cassettes in the chromosome.)

4 Notes

1. It is important that competent cells are kept on ice throughout the whole process.

2. The longer the cells are growing with the integrase expressed, the higher the recombination frequency will be. To improve recombination frequency, several passages in LB + Ara can be done. Note however that without selection, cassettes are lost at a frequency similar to the one of reordering events [7].
3. The use of the DAP auxotroph donor (ω7249) for the conjugation delivery protocol guaranties that you get rid of the donor by simply selecting on LB + antibiotic marker of the pSWK, as DAP is completely absent in LB.
4. The recipient strain does not actually need DAP to grow, but adding DAP in the medium allows to later mix the cultures without worrying about killing the donor cells.
5. If you do not have the equipment to do this, you can mix the cells and centrifuge them gently 4 min at 2,000 × *g*, discard the supernatant, resuspend the cells in 100 μl, and pipette onto the filter.

References

1. Bittker JA, Le BV et al (2004) Directed evolution of protein enzymes using nonhomologous random recombination. Proc Natl Acad Sci U S A 101:7011–7016
2. Gertz J, Siggia ED, Cohen BA (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature 457:215
3. Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296:1466
4. Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Systems Biol 3:145
5. Tsuji T, Onimaru M, Yanagawa H (2001) Random multi-recombinant PCR for the construction of combinatorial protein libraries. Nucleic Acids Res 29:E97
6. Cambray G, Guerout AM, Mazel D (2010) Integrons. Annu Rev Genet 44:141–166
7. Bikard D, Julie-Galau S et al (2010) The synthetic integron: an in vivo genetic shuffling device. Nucleic Acids Res 38:e153
8. Demarre G, Frumerie C et al (2007) Identification of key structural determinants of the IntI1 integron integrase that influence attC x attI1 recombination efficiency. Nucleic Acids Res 35:6475–6489
9. Valens M, Penaud S et al (2004) Macrodomain organization of the Escherichia coli chromosome. EMBO J 23:4330–4341
10. Demarre G et al (2005) A new family of mobilizable suicide plasmids based on the broad host range R388 plasmid (IncW) or RP4 plasmid (IncPα) conjugative machineries and their cognate *E. coli* host strains. Res Microbiol 156:245–255
11. Babic A, Guérout A-M, Mazel D (2008) Construction of an improved RP4 (RK2)-based conjugative system. Res Microbiol 159:545
12. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645
13. Chaveroche MK, Ghigo JM, d'Enfert C (2000) A rapid method for efficient gene replacement in the filamentous fungus Aspergillus nidulans. Nucleic Acids Res 28:E97

Part IV

Computational Tools for Modelling Biological Systems

Chapter 15

Systematic Methodology for the Development of Mathematical Models for Biological Processes

Cleo Kontoravdi

Abstract

Synthetic biology gives researchers the opportunity to rationally (re-)design cellular activities to achieve a desired function. The design of networks of pathways towards accomplishing this calls for the application of engineering principles, often using model-based tools. Success heavily depends on model reliability. Herein, we present a systematic methodology for developing predictive models comprising model formulation considerations, global sensitivity analysis, model reduction (for highly complex models or where experimental data are limited), optimal experimental design for parameter estimation, and predictive capability checking. Its efficacy and validity are demonstrated using an example from bioprocessing. This approach systematizes the process of developing reliable mathematical models at a minimum experimental cost, enabling *in silico* simulation and optimization.

Key words Global sensitivity analysis, Optimal experiment design, Mathematical modelling, Parameter estimation, Model validation

1 Introduction

Biochemical networks are systems with intricate interdependencies and complex regulatory loops. Mathematical modelling can be used to interpret experimental findings in order to reveal and analyze their composite behavior, or even to design experiments that will help discriminate between different hypotheses regarding the underlying mechanisms. They can further aid the decomposition of a network into smaller sections and assess the modularity of their behavior.

A prerequisite is the development of a modelling framework supported by fundamental knowledge and experimental data. Herein, a systematic framework for developing models of biological systems is described. This consists of four steps: the formulation of mathematical equations to describe the system of interest, the analysis of the system of equations developed, the

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_15, © Springer Science+Business Media New York 2013

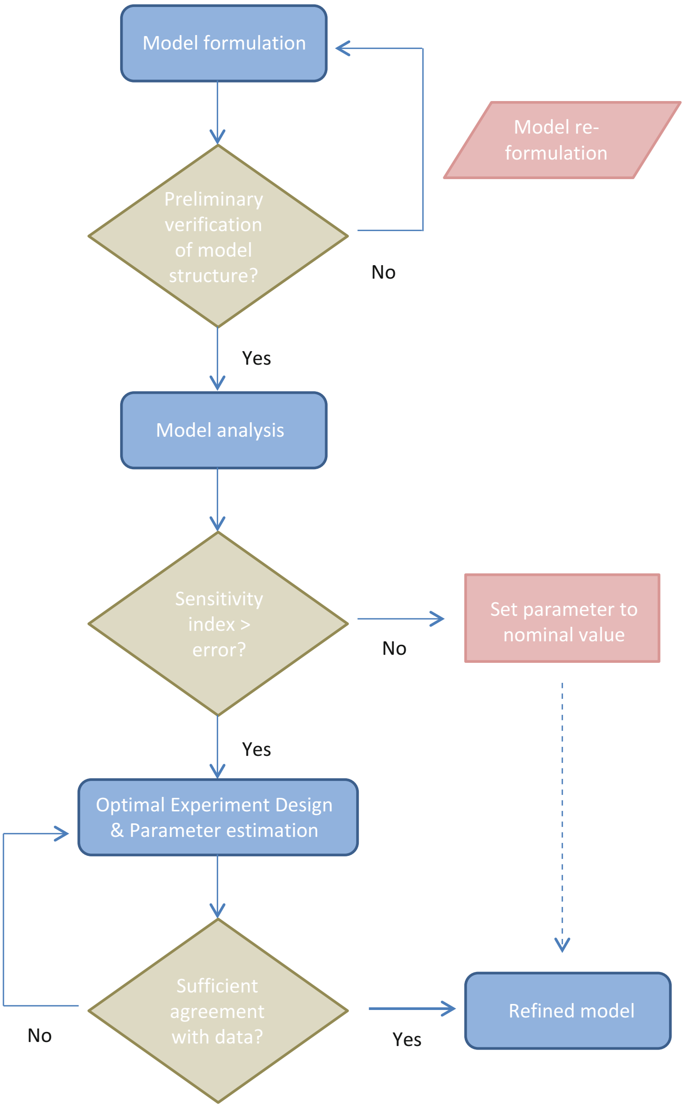

Model formulation
Model re-formulation
Preliminary verification of model structure?
No
Yes
Model analysis
Sensitivity index > error?
No
Set parameter to nominal value
Yes
Optimal Experiment Design & Parameter estimation
Sufficient agreement with data?
No
Yes
Refined model

Fig. 1 Model development framework

design of experiments with the purpose of confirming model validity within a range of conditions, and the use of the resulting data to estimate model parameters. Following model verification, it is recommended that model performance be tested against independent data sets. Even though a mathematical model cannot truly be validated, it can be confirmed to apply under certain experimental conditions. Below we present each step of the methodology, as summarized in Fig. 1, and illustrate its application through a case study on mammalian cell culture.

2 Model Development Framework

2.1 Step 1: Model Formulation

This first step of model formulation should be carried out with two considerations in mind:

- What the phenomena underlying system behavior are.
- What the intended use of the model is.

There are two approaches to model formulation, bottom-up and top-down. The bottom-up approach accounts for each known biological mechanism implicated in the system behavior under study (e.g., [1, 2]). It is an exhaustive approach that is best suited for testing hypotheses regarding the underlying mechanisms. However, it requires in-depth knowledge and availability of precise intracellular measurements under a wide range of conditions. On the other hand, the top-down modelling approach starts with a lumped description and adds mechanistic detail incrementally when warranted by experimental data. An example of the top-down approach as applied to the protein folding pathway in yeast is described in [3].

The intended use of the model also dictates the important model characteristics. For example, models used for online decision-making, such as control and optimization, are required to accurately describe a wide range of dynamic conditions, while being computationally cheap. Such models could, in fact, be reduced formats of more detailed models. This step should also include an initial verification of the model structure using preliminary experimental data. These can also be used to provide an initial estimate of parameter values when these are unknown.

2.2 Step 2: Model Analysis

Sensitivity analysis is used to increase confidence in the model and its predictions by providing an understanding of how its outputs respond to changes in the inputs, such as model parameters or model-independent variables, and is considered a prerequisite for model building in any setting. Three categories of sensitivity analysis methods exist, screening, local and global sensitivity analysis [4], which are presented below. When considering which outputs to examine, it is vital that the model developer understands which model outputs are measurable experimentally and with what degree of accuracy.

2.2.1 Factor Screening

Factor screening is suitable for models with a large number of input factors, which are therefore computationally expensive to evaluate. It is reported to be able to economically handle models with hundreds of input factors. Nevertheless, this type of method only yields qualitative results, i.e., it only ranks factors in order of importance, without quantifying each factor's impact on the model output. The principal factor screening methods include one-at-a-time design, Morris' design, Cotter's Systematic Fractional Replicate Design, Iterated Fractional Factorial Design, and Sequential Bifurcation [4].

2.2.2 Local Sensitivity

Local sensitivity analysis concentrates on the local impact of the factors on the output of the model and is usually carried out by computing partial derivatives of the output functions with respect to the input factors. In order to compute these derivatives, two methods are used, namely Finite Difference Approximation and the Direct Method. Their use for predictive models should be limited to the case of linear models or models with modest factor variations. Local sensitivity analysis methods are not suitable for nonlinear models as they are affected by uncertainties of different orders of magnitude [4].

2.2.3 Global Sensitivity Analysis

Global sensitivity analysis is the only method that provides quantitative results while examining the entire range of parameter values. This is of particular importance in biological modelling, since model parameters can vary within large intervals depending on their physiological meaning. An additional advantage of global sensitivity analysis is that the estimates of individual parameters are evaluated while varying all other parameters as well. In this way, the relative uncertainty of each input is taken into account, revealing any existing first-order interactions. Global sensitivity analysis can aid in model building by identifying the factors, i.e., the model inputs in the form of variables and parameters, which contribute the most to model variability. The interest of its application is in determining which model parameters have a significant impact on the output, and therefore require further investigation to determine their values with high precision. A number of global sensitivity analysis techniques can be used to calculate the desired sensitivity estimates, such as correlation ratios [5, 6], the Fourier Amplitude Sensitivity Test [7–10], and the Sobol' method [4, 11]. The FAST and Sobol' methods, for all intents and purposes, are equivalent.

The more recently developed Derivative-Based Global Sensitivity Measures (DGSM, [12]) is a global screening method, shown in [13] to provide results similar to the variance-based methods' total sensitivity index (TSI). In fact, it has been shown to have a direct correlation with the Sobol' TSI in most cases [12]. The main benefit of using DGSM over a variance-based method is the significant gain in terms of computational time. DGSM provides only TSI information, and not first or higher order information.

The dimensionality of the sensitivity analysis problem is ultimately defined by the number of model parameters; therefore, a computational constraint regarding the maximum possible number of individually scanned parameters is imposed implicitly. This constraint is unavoidable due to the number of model evaluations required for the Monte Carlo integrals to converge, which increases with model dimension. Researchers in the field of GSA often resort to parameter grouping in order to reduce the dimensionality of the problem, thus solving a more tractable version of the original problem. A detailed discussion on parameter grouping and various methods for grouping can be found in the work of [13].

The higher the sensitivity index of a parameter with respect to an output, the more influence that parameter exerts on the value of that output. Typically, a global sensitivity analysis will yield a table of numerical values associating inputs to outputs. Each of these values reflects how much the value of the particular input influences the value of a specific output. However, when analyzing the results and deciding on which parameters are significant, it is important to take into account the confidence with which the value of an output can be determined experimentally. In other words, it is important to consider the experimental error associated with measuring each output. If the sensitivity index of a parameter is greater than that error, then it means that variation in that parameter value will yield a change in the output that is experimentally observable. The experimental error therefore sets a threshold for parameter significance.

2.3 Step 3: Optimal Experiment Design

Parameter estimation for models of biological systems is usually carried out using already existing data. However, unique identification of the parameter set is only possible if the available data are sufficiently rich [14]. Furthermore, measurement noise and sampling frequency also influence parameter accuracy. The use of model-based experimental design can result in significant improvements in parameter confidences [14, 15].

Optimal experimental design targets the determination of input profiles that generate informative experimental data, which then enable accurate parameter estimation. This comes as a result of observations that highly dynamic occurrences benefit the estimation process. For example, it has been shown that a fed-batch experiment with time-varying feed rate leads to more accurate parameter estimates that a batch experiment of the same system [16]. Optimal experimental design tools have found application mainly in the food microbiology sector, e.g., [17, 18].

Optimal experiments are designed to improve model accuracy by reducing parameter uncertainty. The latter is expressed through the size of the variance-covariance matrix

$$V = V(\hat{\theta}, \varphi)$$

where, $\hat{\theta}$ is the vector of the best parameter estimates and φ is the vector of the experiment decision variables and contains the variances of individual parameters and covariance of pairs of parameters. The aim of optimal experimental design is to minimize some metric of this matrix or maximize the same metric of the inverse of V, which is known as the Fisher Information Matrix, M. Several criteria exist that can achieve this:

1. D-optimal design minimizes the volume of the confidence ellipsoid, i.e., the determinant of V, and is independent of parameter scales [19].
2. E-optimal design minimizes the length of the principle axis of the confidence ellipsoid, i.e., the length of the largest eigenvalue of V [19].
3. A-optimal design minimizes the average variance of the parameter estimates, i.e., the trace of V, and is affected by parameter scales [19].
4. M-optimal design [15] is based on the implementation of D-optimal design and the minimization of the angle between the indifference ellipsoid axis and the reference axis in the parameter space. The latter seeks to reduce parameter correlation [20].

D-optimal design is the most widely used [19].

2.4 Step 4: Parameter Estimation and Model Validation

The data generated from the designed experiments are analyzed to obtain the variance model and related parameters before being used for parameter estimation. A key consideration when deciding if parameter estimation is successful is the breadth of the confidence interval for each estimated parameter and how this compares to its optimized value. Confidence intervals of the same order of magnitude or larger than the optimized parameter value indicate insufficient data to estimate that parameter and, thus, validate the model. Hence, it is important to follow the process of designing an experiment, conducting it and reestimating parameter values iteratively, until sufficient agreement between the data and the model has been reached.

If it is not possible to obtain a satisfactory parameter estimate, then the model developer needs to consider (a) if all possible control variables, i.e., inputs we can manipulate, have been taken into account when designing the experiments; (b) the quality of the experimental data in terms of measurement error, and (c) whether the value of that parameter stays constant throughout the changing conditions of the experiment, or if it is likely to change due to possible regulation effects. The latter is a consideration that would typically be taken into account in Step 1. However, dynamic experiments often provide additional insight into the underlying mechanisms and the model may need to be reformulated to account for this.

3 Case Study: Application of Model Development Platform to Bioprocessing

Step 1: A hybrid model of antibody-producing mammalian cell cultures describing cell growth limited by glucose and glutamine availability and inhibited by lactate and ammonia accumulation, cell death depending on ammonia accumulation, and synthesis of an IgG1 antibody product is described in [21]. The mathematical description can be found in Table 1. This was confirmed to be structurally correct by comparison to batch cell culture data.

Step 2: The Sobol' method of global sensitivity analysis supplemented by a one-at-a-time method were used to identify the seven important model parameters to which the model outputs were most sensitive. The threshold for sensitivity was 10 %, which is the largest error associated to the analytical technique used to quantify the outputs, specifically the antibody concentration in the extracellular medium. Parameters with sensitivity indices above 10 % required accurate estimation from experimental data and were the focus of the optimal experimental design. The significant parameters were $Y_{\mathrm{x,glc}}$, $Y_{\mathrm{x,gln}}$, $Y_{\mathrm{lac,glc}}$, $Y_{\mathrm{amm,gln}}$, K_{lysis}, μ_{max}, and N_{H}.

Step 3: D-optimal experimental design was then applied to design information-rich experiments for the estimation of the values of these "sensitive" parameters and extend the range of validity to fed-batch conditions. D-optimal experimental design was used to maximize the information content of experimental data, specifically for the estimation of the seven parameters identified in the sensitivity analysis and one-at-a-time screening. This is done by minimizing the volume of the confidence ellipsoid, i.e., minimizing the determinant of the variance-covariance matrix, V, of the parameters to be estimated.

The optimization problem seeks to determine the initial conditions, experiment duration, variation of controls, and timing of samples such that the maximum amount of information is generated with the given measurements. These measurements were nutrient and metabolite concentrations, the viable and total cell concentrations, and the extracellular MAb concentration.

The design was conducted in the gPROMS [22] modelling environment, which has a dedicated function, namely "experiment design for parameter precision", and uses an SRQPD sequential quadratic programming code. The amount of feed, feeding interval, and experiment duration were treated as degrees of freedom within certain operational limits. The initial glucose concentration was limited between 5.5 and 25 mM. The feed was allowed to vary between 0 and 12.5 ml/h. Concentrated medium was supplied to the culture vessel in pulses and the maximum total volume of feed was fixed at 8.75 ml, which represented less than 5 % of the total culture volume (200 ml), so as to avoid dilution effects. The amount of feed supplied at each feeding interval was optimized by the design, as was the timing of these intervals. The earliest measurement time was set at 12 h and the minimum time between measurements at 6 h. The duration of the experiment was allowed to vary between 4 and 8 days. The optimal duration was determined at 168 h (7 days), during which a pulse feed was introduced once a day, at an interval of 24 h.

Table 1
Summary of model equations

Description	Equation
Reactor working volume	$\frac{dV}{dt} = F_{\mathrm{in}} - F_{\mathrm{out}}$
Viable cell concentration	$\frac{d(VX_v)}{dt} = \mu VX_v - \mu_d VX_v - F_{\mathrm{out}} X_v$
Total cell concentration	$\frac{d(VX_t)}{dt} = \mu VX_v - F_{\mathrm{out}} X_t$
Specific cell growth rate	$\mu = \mu_{\mathrm{max}} f_{\mathrm{lim}} f_{\mathrm{inh}}$
Growth limitation function	$f_{\mathrm{lim}} = \left(\frac{[\mathrm{GLC}]}{K_{\mathrm{glc}}+[\mathrm{GLC}]}\right)\left(\frac{[\mathrm{GLN}]}{K_{\mathrm{gln}}+[\mathrm{GLN}]}\right)$
Growth inhibition function	$f_{\mathrm{inh}} = \left(\frac{\mathrm{KI}_{\mathrm{lac}}}{\mathrm{KI}_{\mathrm{lac}}+[\mathrm{LAC}]}\right)\left(\frac{\mathrm{KI}_{\mathrm{amm}}}{\mathrm{KI}_{\mathrm{amm}}+[\mathrm{AMM}]}\right)$
Cell death rate	$\mu_d = \frac{\mu_{d,\,\mathrm{max}}}{1+\left(\frac{K_{d,\,\mathrm{amm}}}{[\mathrm{AMM}]}\right)^n}, n > 1$
Glucose material balance	$\frac{d(V[\mathrm{GLC}])}{dt} = -Q_{\mathrm{glc}} VX_v + F_{\mathrm{in}}[\mathrm{GLC}]_{\mathrm{in}} - F_{\mathrm{out}}[\mathrm{GLC}]$
Specific glucose uptake rate	$Q_{\mathrm{glc}} = \frac{\mu}{Y_{x,\mathrm{glc}}} + m_{\mathrm{glc}}$
Glutamine material balance	$\frac{d(V[\mathrm{GLN}])}{dt} = -Q_{\mathrm{gln}} VX_v - K_{d\mathrm{gln}} V[\mathrm{GLN}] + F_{\mathrm{in}}[\mathrm{GLN}]_{\mathrm{in}} - F_{\mathrm{out}}[\mathrm{GLN}]$
Specific glutamine uptake rate	$Q_{\mathrm{gln}} = \frac{\mu}{Y_{x,\mathrm{gln}}} + m_{\mathrm{gln}}$
Maintenance energy of glutamine	$m_{\mathrm{gln}} = \frac{\alpha_1[\mathrm{GLN}]}{\alpha_2+[\mathrm{GLN}]}$
Lactate material balance	$\frac{d(V[\mathrm{LAC}])}{dt} = Q_{\mathrm{lac}} VX_v - F_{\mathrm{out}}[\mathrm{LAC}]$
Specific lactate production rate	$Q_{\mathrm{lac}} = Y_{\mathrm{lac,glc}} Q_{\mathrm{glc}}$
Ammonia material balance	$\frac{d(V[\mathrm{AMM}])}{dt} = Q_{\mathrm{amm}} VX_v + K_{d\mathrm{gln}} V[\mathrm{GLN}] - F_{\mathrm{out}}[\mathrm{AMM}]$
Specific ammonia production rate	$Q_{\mathrm{amm}} = Y_{\mathrm{amm,\,gln}} Q_{\mathrm{gln}}$
Heavy chain mRNA balance	$\frac{dm_H}{dt} = N_H S_H - Km_H$
Light chain mRNA balance	$\frac{dm_L}{dt} = N_L S_L - Km_L$
IntraER heavy chain balance	$\frac{d[H]}{dt} = T_H m_H - R_H$
IntraER light chain balance	$\frac{d[L]}{dt} = T_L m_L - R_L$
Rate of heavy chain consumption for assembly	$R_H = \frac{2}{3} K_A[H]^2$
Rate of light chain consumption for assembly	$R_L = 2K_A[H_2][L] + K_A[H_2L][L]$
Assembly intermediate balance in the ER	$\frac{d[H_2]}{dt} = \frac{1}{3} K_A[H]^2 - 2K_A[H_2][L]$
Assembly intermediate balance in the ER	$\frac{d[\vert H_2L]}{dt} = 2K_A[H_2][L] - K_A[H_2L][L]$
IntraER antibody balance	$\frac{d[H_2L_2]_{ER}}{dt} = K_A[H_2L][L] - K_{ER}[H_2L_2]_{ER}$

(continued)

Table 1
(continued)

Description	Equation
Golgi antibody balance	$\frac{d[H_2L_2]_G}{dt} = \varepsilon_1 K_{\text{ER}}[H_2L_2]_{\text{ER}} - K_G[H_2L_2]_G$
Antibody balance in the extracellular medium	$\frac{d(V[\text{MAb}])}{dt} = (\gamma_2 - \gamma_1\mu)Q_{\text{MAb}}VX_v - F_{\text{out}}[\text{MAb}]$
Specific antibody production rate	$Q_{MAb} = \varepsilon_2\lambda K_G[H_2L_2]_G$

Table 2
Parameter estimation results for fed-batch culture

		Confidence intervals		
Parameter	**Final value**	**90 %**	**95 %**	**99 %**
$\Upsilon_{\text{x,glc}}$	2.6×10^8	2.224×10^7	2.654×10^7	3.494×10^7
$\Upsilon_{\text{x,gln}}$	8×10^8	2.134×10^7	2.546×10^7	3.352×10^7
$\Upsilon_{\text{lac,glc}}$	2.0	1.768×10^{-1}	2.109×10^{-1}	2.777×10^{-1}
$\Upsilon_{\text{amm,gln}}$	4.5×10^{-1}	6.564×10^{-2}	7.832×10^{-2}	1.031×10^{-1}
K_{lysis}	$3. \times 10^{-2}$	2.520×10^{-3}	3.007×10^{-3}	3.959×10^{-3}
μ_{max}	5.4×10^{-2}	8.306×10^{-4}	9.909×10^{-4}	1.305×10^{-3}
N_{H}	101	5.067	6.045	7.959

Step 4: The sequence of designing an experiment, performing it, and using the generated data for model validation was done iteratively, until model predictions and experimental data were in satisfactory agreement. The results of the designed experiments were used to estimate the values of model parameters using the "parameter estimation" entity in gPROMS based on the same SRQPD sequential quadratic programming code. Parameter estimation was based on the maximum likelihood formulation, which attempts to determine values for the uncertain physical and variance model parameters that maximize the probability that the model will predict the measurement values obtained from the experiments. The statistical variance model of constant variance was used in this case. The 95 % confidence intervals of estimated parameters were deemed satisfactory at ±10 % of the final parameter values (Table 2), except for $\Upsilon_{\text{amm,gln}}$ (±17 %).

The complete set of parameter values for fed-batch cultures can be seen in Table 3, where they are compared to the corresponding values for batch culture operation. The major differences are in the parameters describing nutrient utilization, as expected. Increased

Table 3
Parameter values for batch and fed-batch culture operation

Parameter	Description	Unit	Batch	Fed-batch
K	Heavy- and light-chain mRNA decay rate	h^{-1}	10^{-1}	10^{-1}
K_A	Assembly rate constant	cell/molecule-L	10^{-6}	10^{-6}
$K_{d,amm}$	Ammonia constant for cell death	mM	1.76	1.76
$K_{d,gln}$	Constant for glutamine degradation	h^{-1}	9.6×10^{-3}	9.6×10^{-3}
K_{ER}	Rate constant for ER-to-Golgi transport	h^{-1}	6.9×10^{-1}	6.9×10^{-1}
K_G	Rate constants for Golgi-to-medium antibody transport	h^{-1}	1.4×10^{-1}	1.4×10^{-1}
K_{glc}	Monod constant for glucose	mM	7.5×10^{-1}	7.5×10^{-1}
K_{gln}	Monod constant for glutamine	mM	7.5×10^{-2}	7.5×10^{-2}
KI_{amm}	Monod constant for ammonia	mM	28.48	28.48
KI_{lac}	Monod constant for lactate	mM	171.76	171.76
K_{lysis}	Cell lysis rate	h^{-1}	5.5×10^{-2}	3.0×10^{-2}
m_{glc}	Maintenance coefficient of glutamine	mmol/cell-h	4.9×10^{-14}	4.9×10^{-14}
n	Cell death by ammonia coefficient	–	2	2
N_H	Heavy chain gene copy number	gene/cell	1.4×10^{2}	1.0×10^{2}
N_L	Light chain gene copy number	gene/cell	1.2×10^{2}	1.0×10^{2}
S_H	Heavy chain gene-specific transcription rate	mRNA/gene-h	3×10^{3}	3×10^{3}
S_L	Light chain gene-specific transcription rate	mRNA/gene-h	4.5×10^{3}	4.5×10^{3}
T_H	Heavy chain-specific translation rate	chain/mRNA-h	17	17
T_L	Light chain-specific translation rate	chain/mRNA-h	11.5	11.5
$\Upsilon_{amm,gln}$	Yield of ammonia from glutamine	mmol/mmol	4.3×10^{-1}	4.5×10^{-1}
$\Upsilon_{lac,glc}$	Yield of lactate from glucose	mmol/mmol	1.4	2.0
$\Upsilon_{x,glc}$	Yield of cells on glucose	cell/mmol	1.1×10^{8}	2.6×10^{8}
$\Upsilon_{x,gln}$	Yield of cells on glutamine	cell/mmol	5.6×10^{8}	8×10^{8}
α_1	Constant of glutamine maintenance coefficient	mM L/cell-h	3.4×10^{-13}	3.4×10^{-13}
α_2	Constant of glutamine maintenance coefficient	mM	4	4
γ_1	Constant for antibody production	–	10^{-1}	10^{-1}
γ_2	Constant for antibody production	h	2	2

(continued)

Table 3 (continued)

Parameter	Description	Unit	Batch	Fed-batch
ε_1	ER glycosylation efficiency factor	–	9.9×10^{-1}	9.95×10^{-1}
ε_2	Golgi glycosylation efficiency factor	–	1	1
μ_{max}	Maximum specific death rate	h^{-1}	5.8×10^{-2}	5.8×10^{-2}
$\mu_{\mathrm{d,max}}$	Maximum specific growth rate	h^{-1}	6×10^{-2}	6×10^{-2}

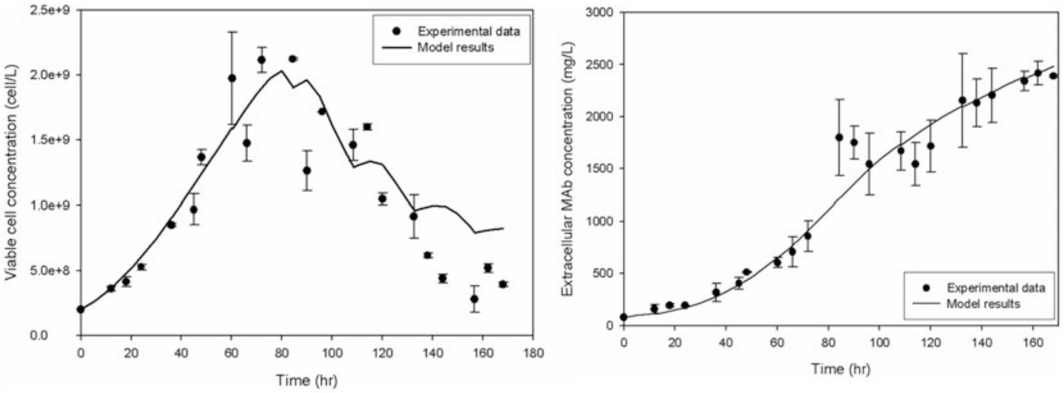


Fig. 2 Agreement of model simulation results with experimental data for cell (*left panel*) and antibody (*right panel*) concentration for fed-batch cultures of HFN 7.1 hybridoma cells (adapted from [23])

uptake of both nutrients in fed-batch culture also results in prolonged cell growth and higher viable (and hence, total) cell concentration, and is reflected in the higher values in cell yield on glucose and glutamine (2.6×10^8 and 8×10^8, respectively) compared to values for batch culture (1.1×10^8 and 5.6×10^8, respectively). The yield of lactate on glucose is at its maximum value of 2 in the case of fed-batch culture, while it is 1.4 in batch culture. This is due to the increased availability and utilization of glucose, which leads to higher conversion into lactate, and agrees with previous experimental observations [24, 25]. Finally, there is some difference in the values of the heavy and light chain gene copy number. As the values for batch cultures represent initial estimates, these are estimated using the data for fed-batch operation. The estimated value of 100 genes per cell for both the heavy and the light chain is therefore considered to be the most accurate estimate since these figures should be independent of culture operation mode.

The resulting model is in good agreement with the results of this first experiment (Figs. 2 and 3). Simulation results for viable cell concentration closely match the data during the first 60 h, as

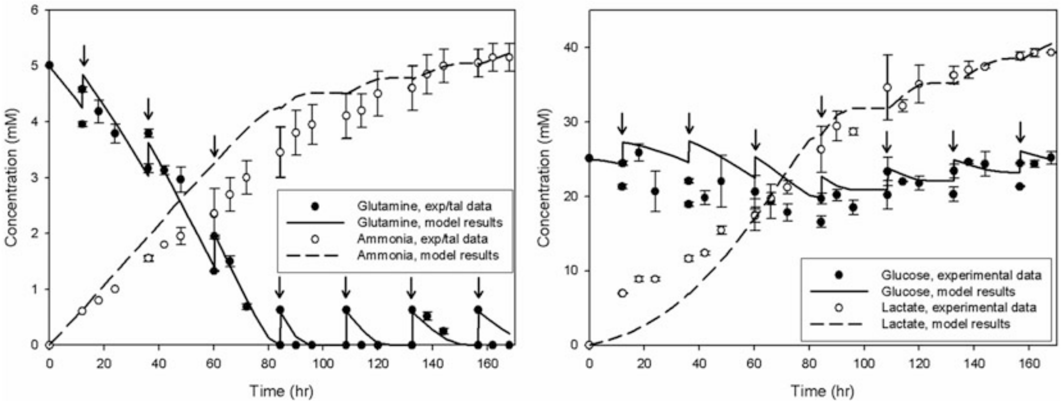


Fig. 3 Agreement of model simulation results with experimental data for glutamine, ammonia (*left panel*), glucose, and lactate (*right panel*) concentrations in the extracellular medium

shown in Fig. 2. Thereafter, model predictions follow the trend of the experimental data correctly predicting the plateau in viable cell concentration, and during the last 40 h the model over-predicts the concentration of viable cells in the culture. This is probably because the viable cell concentration in vitro is too low for the cells to recover viability, but the model predicts that the cells will continue to grow on the additional glucose supplied through the feed. The model closely tracks the data throughout the duration of the culture (Fig. 2).

Figure 3 shows good agreement between model simulation results and experimental data for glutamine and ammonia. Glutamine concentration is correctly predicted throughout the duration of the culture. Discrepancies only occur once glutamine has reached zero concentration (after 80 h), when additional glutamine fed to the culture is metabolized more quickly than predicted. Ammonia concentration is also closely tracked by model results. It is correctly predicted during the initial lag phase, over-predicted over the following 70 h, and closely matched during the remainder of the culture. Overall, the model captures the profiles of all measure variables successfully.

Following the successful estimation of parameter values, the validity of the resulting model under fed-batch conditions was confirmed by comparison to an independent set of fed-batch culture data.

4 Conclusions

We outlined a systematic step-by-step methodology for model development. A demonstration of its application to formulate and confirm the reliability of a simple, hybrid model of animal cell cultures under batch and fed-batch culture conditions was also presented. Overall, the framework aims to build high-fidelity models of biological systems at a minimum experimental cost, while avoiding trial-and-error practices. This systematic combination of modelling and experimentation through process systems engineering tools can Synthetic Biologists develop mathematical models of existing systems for analysis, or models of systems to be designed to test their performance against desired characteristics.

References

1. Jimenez Del Val I, Nagy JM, Kontoravdi C (2011) A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. Biotechnol Prog 27:1730–1743
2. Selvarasu S, Ho YS et al (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnol Bioeng 109:1415–1429
3. Hildebrandt S, Raden D et al (2008) A top-down approach to mechanistic biological modeling: application to the single-chain antibody folding pathway. Biophys J 95:3535–3558
4. Saltelli A, Chan K, Scott EM (2000) Sensitivity analysis. Wiley, New York
5. Kendall M, Stuart A (1979) The advanced theory of statistics, vol 2. Macmillan, New York
6. Mckay MD (1995) Evaluating prediction uncertainty. US Nuclear Regulatory Commission and Los Alamos National Laboratory, Washington, DC
7. Cukier RI, Fortuin CM et al (1973) Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 1. Theory. J Chem Phys 59:3873–3878
8. Cukier RI, Schaibly JH, Shuler KE (1975) Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 3. Analysis of approximations. J Chem Phys 63:1140–1149
9. Cukier RI, Levine HB, Shuler KE (1978) Nonlinear sensitivity analysis of multi-parameter model systems. J Comput Phys 26:1–42
10. Schaibly JH, Shuler KE (1973) Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 2. Applications. J Chem Phys 59:3879–3888
11. Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55:271–280
12. Sobol IM, Kucherenko S (2009) Derivative based global sensitivity measures and their link with global sensitivity indices. Math Comput Simul 79:3009–3017
13. Kiparissides A, Kucherenko SS et al (2009) Global sensitivity analysis challenges in biological systems modeling. Ind Eng Chem Res 48:7168–7180
14. Versyck KJ, Claes JE, Vanimpe JF (1997) Practical identification of unstructured growth kinetics by application of optimal experimental design. Biotechnol Prog 13:524–531
15. Nathanson MH, Saidel GM (1985) Multiple-objective criteria for optimal experimental design—application to ferrokinetics. Am J Physiol 248:R378–R386
16. Munack A (1989) Optimal feeding strategy for identification of monod-type models by fed-batch experiments. Computer applications in fermentation technology: modelling and control of biotechnological processes. Elsevier Applied Science Publishers Ltd, Barking Essex
17. Van Derlinden E, Bernaerts K, Van Impe JF (2008) Accurate estimation of cardinal growth temperatures of Escherichia coli from optimal dynamic experiments. Int J Food Microbiol 128:89–100
18. Bernaerts K, Gysemans KPM et al (2006) Optimal experiment design for cardinal values estimation: guidelines for data collection (vol 100, pg 153, 2005). Int J Food Microbiol 110:112–113
19. Jacques JA (1998) Design of experiments. J Franklin Inst 335:259–279
20. Sidoli FR, Mantalaris A, Asprey SP (2004) Modelling of Mammalian cells and cell culture processes. Cytotechnology 44:27–46
21. Kontoravdi C, Asprey SP et al (2005) Application of global sensitivity analysis to determine goals for design of experiments: an example study on antibody-producing cell cultures. Biotechnol Progress 21:1128–1135
22. Process Systems Enterprise (2002) gPROMS Advanced User Guide, London, UK
23. Kontoravdi C, Pistikopoulos EN, Mantalaris A (2010) Systematic development of predictive

mathematical models for animal cell cultures. Comput Chem Eng 34:1192–1198

24. Glacken MW, Fleischaker RJ, Sinskey AJ (1986) Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. Biotechnol Bioeng 28:1376–1389

25. Miller WM, Blanch HW, Wilke CR (1988) A kinetic analysis of hybridoma growth and metabolism in batch and continuous suspension culture: Effect of nutrient concentration, dilution rate, and pH Biotechnology and Bioengineering 32(8):947–965

Chapter 16

SOBOLHDMR: A General-Purpose Modeling Software

Sergei Kucherenko

Abstract

One of the dominant approaches in synthetic biology is the development and implementation of minimal circuits that generate reproducible and controllable system behavior. However, most biological systems are highly complicated and the design of sustainable minimal circuits can be challenging. SobolHDMR is a general-purpose metamodeling software that can be used to reduce the complexity of mathematical models, such as those for metabolic networks and other biological pathways, yielding simpler descriptions that retain the features of the original model. These descriptions can be used as the basis for the design of minimal circuits or artificial networks.

Key words Model reduction, Global sensitivity analysis, Metamodeling, Software

1 Introduction

Model-based simulation of complex processes is an efficient approach of exploring and studying systems whose experimental analysis is costly or time-consuming. Good modeling practice requires sensitivity analysis (SA) to ensure the model quality by analyzing the model structure, selecting the best type of model, and effectively identifying the important model parameters.

Sensitivity analysis (SA) could be defined as the study of how the uncertainty in the output of a mathematical model can be apportioned, quantitatively or qualitatively, to different sources of uncertainty in the model input [1, 2]. There are two different approaches to sensitivity analysis: local and global. Local sensitivity analysis is concerned with the derivative of the output with respect to an input parameter while holding the other inputs fixed to a central value. It is based on partial derivatives evaluated at a nominal point; therefore, it is only informative at the nominal point in the input space. On the other hand, global sensitivity analysis (GSA) methods evaluate the effect of a factor or a set of factors while other

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology,
vol. 1073, DOI 10.1007/978-1-62703-625-2_16, © Springer Science+Business Media New York 2013

factors are also varied, thus exploring the whole input space and accounting for interactions between the inputs [2, 3]. GSA can be used to rank variables, fix unessential ones, and decrease the dimensionality of a high dimensional problem.

Although variance-based methods are one of the most efficient and popular GSA techniques, these methods generally require a large number of function evaluations to achieve reasonable convergence and can become impractical for large engineering problems. One of the very important and promising developments of model analysis is the replacement of complex models and models which need to be run repeatedly online with equivalent operational metamodels. One major problem associated with traditionally used parameterized polynomial expansions and interpolative look-up tables is that the sampling efforts grow exponentially with respect to the number of input variables. A more efficient way of deducing high dimensional input–output system behavior is based on a technique known as High Dimensional Model Representation (HDMR). HDMR rests on the generic assumption that only low order variable correlations play a significant role in system behavior [4].

In the cases of computationally expensive models the metamodeling technique, which maps inputs and outputs, is a very useful and practical way of making computations tractable. SobolHDMR is a general-purpose metamodeling software. It is based on a number of new techniques which improve the efficiency of the Random Sampling-High Dimensional Model Representation (RS-HDMR) method. The original RS-HDMR method is presented in [5–7] while the improved techniques are described in [8, 9]. In particular, they include determination of an optimal polynomial order and a required number of sampled points to achieve a given tolerance. All techniques implemented in SobolHDMR make use of Quasi-Monte Carlo sampling based on Sobol sequences. SobolHDMR can be used to construct metamodels either from explicitly known models or directly from data given by "black-box" models. Unlike other techniques, methods implemented in SobolHDMR can deal with models with independent and dependent input variables as described in [9]. SobolHDMR can also be used for GSA. GSA methods evaluate the effect of a factor while all other factors are varied as well and thus they account for interactions between variables and do not depend on the choice of a nominal point like local sensitivity analysis methods. GSA methods implemented in SobolHDMR are presented in a number of papers (e.g., [3, 10–12]).

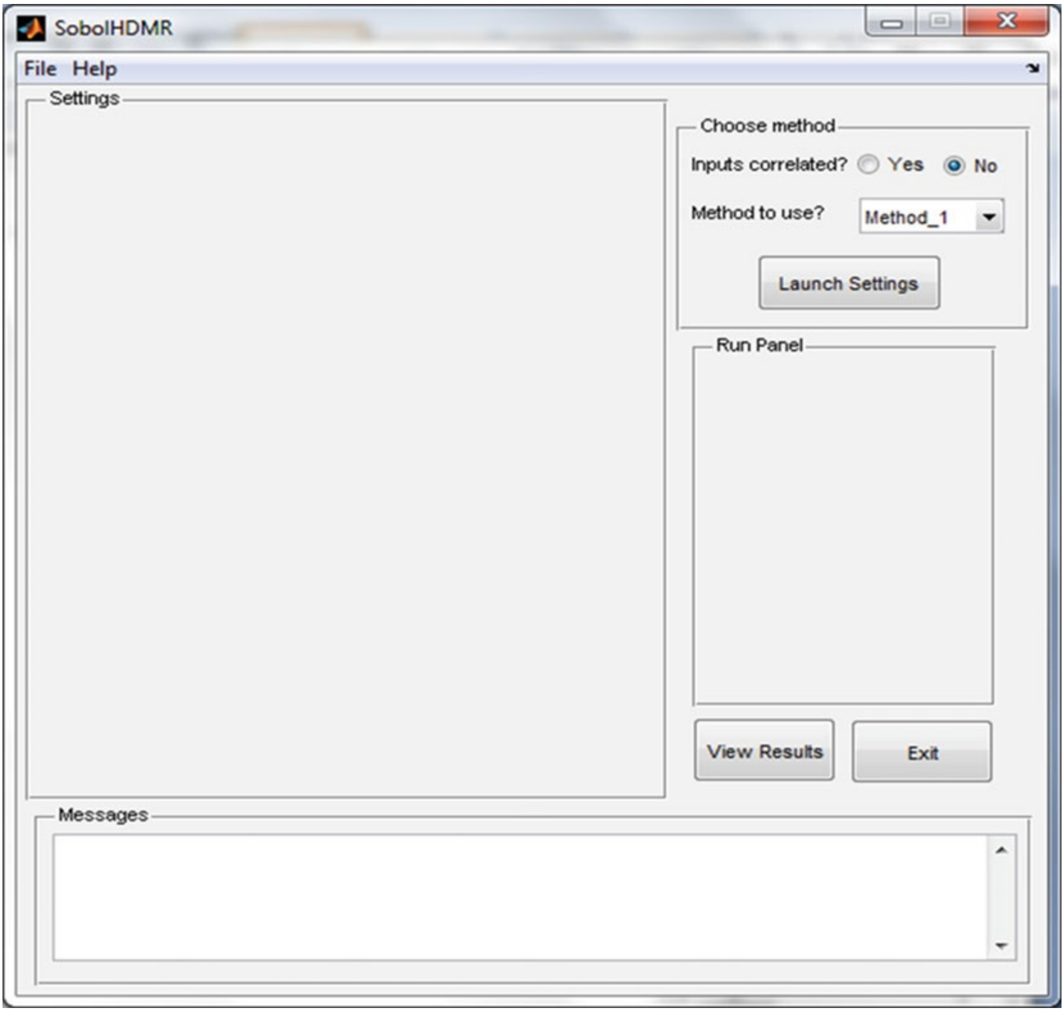
This chapter provides a step-by-step guide for the SobolHDMR methodology and the accompanying software (also in [13]). Researchers interested in using the software can contact Dr. Sergei Kucherenko (s.kucherenko@imperial.ac.uk). The theoretical background of this method can be found in Appendix 2 [9, 14].

2 Software Overview

2.1 "Choose Method" Panel

Depending on the installation package the HDMR GUI can be started by either running the "Hdmr_1.m" file from the MATLAB command prompt (for the full MATLAB and MEX file version) or by running "Start_SobolHDMR.bat" file (for the SobolHDMR.exe version, please note that there are 32-bit and 64-bit versions of this program). This opens the window shown in Fig. 1.

The "File" menu can be used to browse through any directory in order to select a file to open and view or to edit, make changes, and to save. The "Help" menu contains two submenus: one option opens the documentation (window "View manual") and the other ("About Sobol HDMR") shows version information of the SobolHDMR software.


Inputs correlated? ○ Yes ● No

Method to use? Method_1

Fig. 1 Main window for SobolHDMR

The "Choose method" panel is used to specify whether the inputs to be studied are correlated or not. Two methods (options "Method_1" and "Method_2") are available for models with independent inputs, while "Method_2" can deal with both independent inputs and correlated inputs. "Method_1" is implemented in C++ linked with MATLAB engine. "Method_2" is implemented in MATLAB. Generally "Method_1" runs faster than "Method_2" for the same model.

2.2 Launch Settings

The "Launch Settings" button launches the appropriate menus in the "Settings" panel corresponding to the method chosen; Fig. 2a, b shows the menus launched for "Method_1" and "Method_2," respectively.

2.2.1 Settings Panel: Method_1

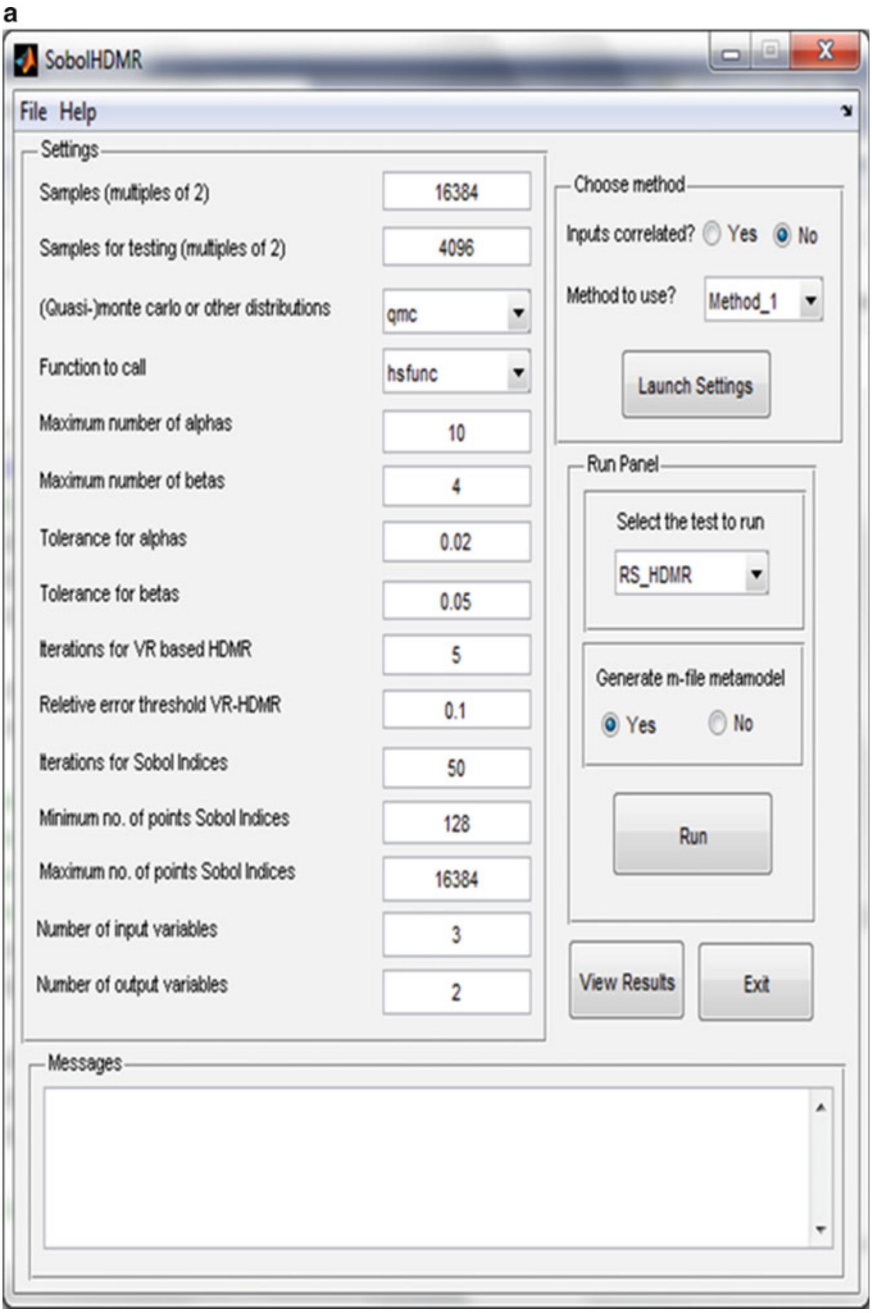
The first input parameter, which we denote N, in the Settings Panel of Fig. 2a (named "Samples") represents the total number of samples to be used for building a metamodel. This should be a multiple of 2 to utilize additional uniformity properties of the Sobol sequences, if they are chosen for sampling.

The second parameter ("Samples for testing") sets the number of points to be used for testing the quality of fit of the HDMR. Typically, it should be less than N.

The third parameter ("Quasi-(monte) carlo or other distributions") enables the choice of whether the input samples (total and that for testing) should be from a standard Monte Carlo ("mc"), Quasi-Monte Carlo ("qmc") sampling strategy based on the Sobol sequences, or a sampling from a user-specified distribution ("dist"). If the option "dist" is chosen, the individual distributions of the inputs must be supplied in the text file "inputDistribution.txt." Any unspecified input in the distribution file is taken by default to be a uniform Quasi-Monte Carlo sample. The only currently in-built distributions are the uniform, normal, and exponential density functions. An example of such an input distribution file is shown in Fig. 3.

The fourth input parameter ("Function to call") offers the option of specifying the function or data to be analyzed; it has five options. The options "hsfunc," "gfunc," and "afunc" are in-built functions representing the Homma-Saltelli (also known as Ishigami) function, the Sobol g-function, and the test function A1 used in Kucherenko et al. [15], respectively. A fourth option, "user_defined," allows the user to supply the function to be evaluated in a MATLAB "user_defined.m" file and must be named as such. This "user_defined.m" file can however call any other file (e.g., it can be linked with gProms®). It simply represents the entry point for a user-defined model function. The inputs to the user-defined function are normalized to $[0, 1]$ but can be converted back to their real values in the user_defined.m file. The syntax for a sample case (the Homma-Saltelli-Ishigami function) is shown in Fig. 4.

The fifth option ("tabulated_data") in the "Function to call" cell provides the capability for the user to supply the input–output mapping in form of text data files. The input and output data sample files for analysis must be named "InputData.txt" and "OutputDataX.txt," respectively, where "X" is an integer beginning from 1 to m, with "m" being the total number of output files;



Samples (multiples of 2) 16384

Samples for testing (multiples of 2) 4096

(Quasi-)monte carlo or other distributions qmc

Function to call hsfunc

Maximum number of alphas 10

Maximum number of betas 4

Tolerance for alphas 0.02

Tolerance for betas 0.05

Iterations for VR based HDMR 5

Reletive error threshold VR-HDMR 0.1

Iterations for Sobol Indices 50

Minimum no. of points Sobol Indices 128

Maximum no. of points Sobol Indices 16384

Number of input variables 3

Number of output variables 2

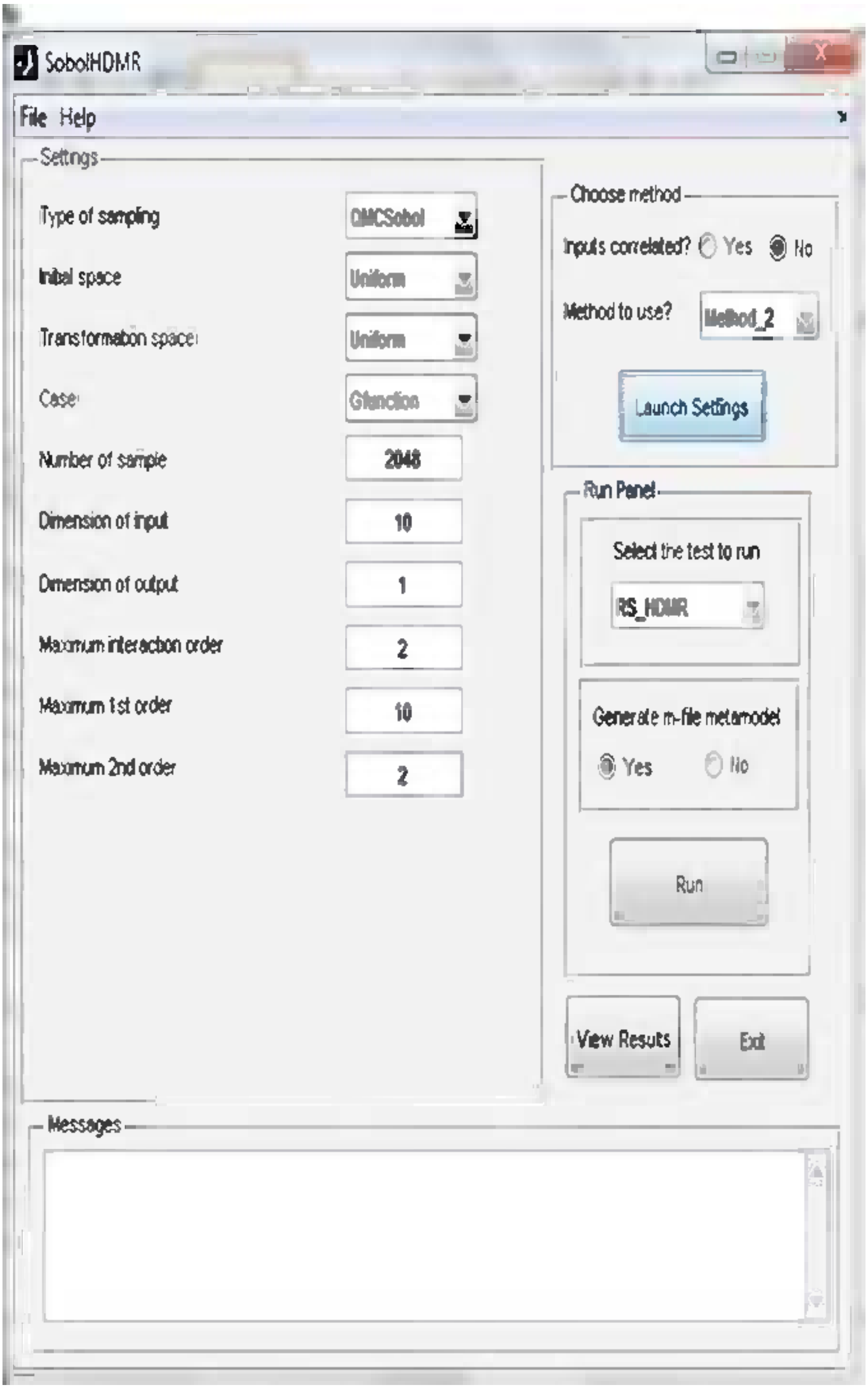
Inputs correlated? ○ Yes ● No

Method to use? Method_1

Select the test to run RS_HDMR

Generate m-file metamodel ● Yes ○ No

Fig. 2 Main window for settings of (**a**) Method_1 and (**b**) Method_2



Type of sampling QMCSobol

Initial space Uniform

Transformation space Uniform

Case Gfunction

Number of sample 2048

Dimension of input 10

Dimension of output 1

Maximum interaction order 2

Maximum 1st order 10

Maximum 2nd order 2

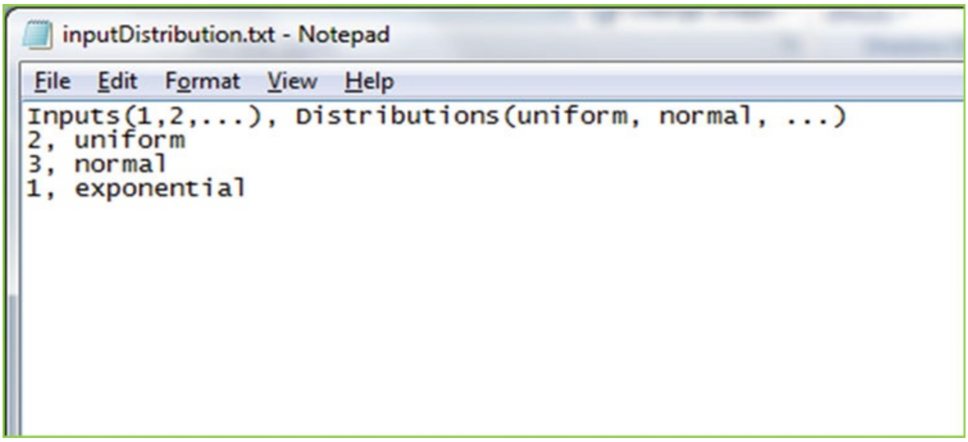
Inputs correlated? ○ Yes ● No

Method to use? Method_2

Select the test to run RS_HDMR

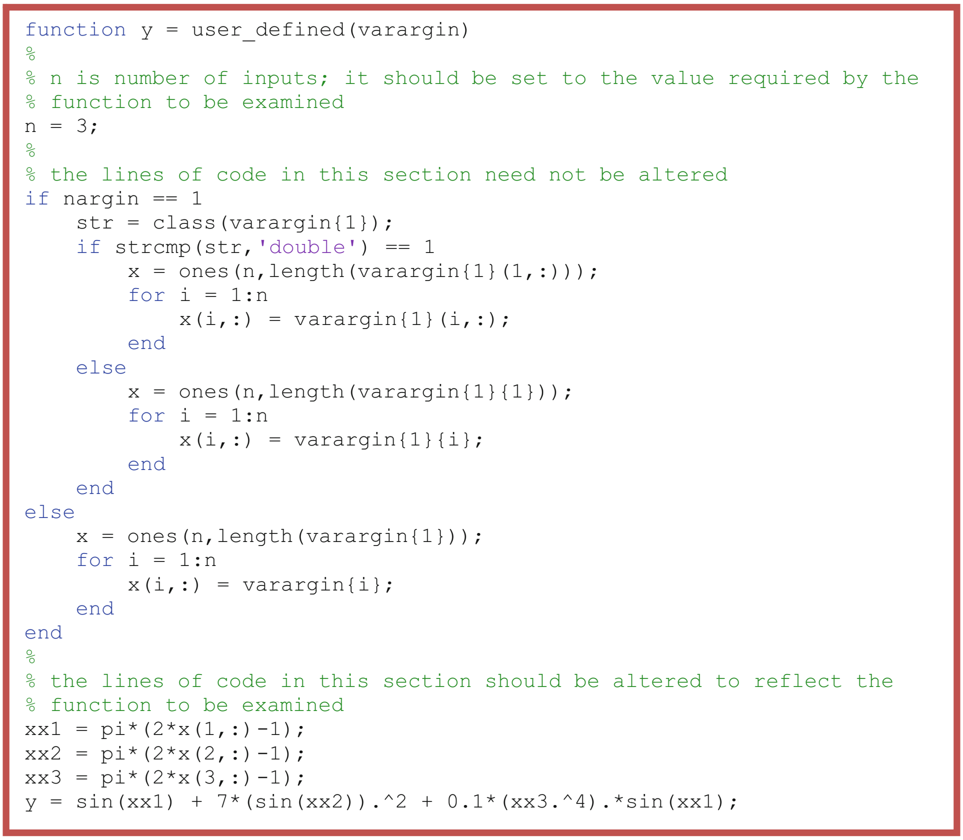
Generate m-file metamodel ● Yes ○ No

Fig. 2 (continued)



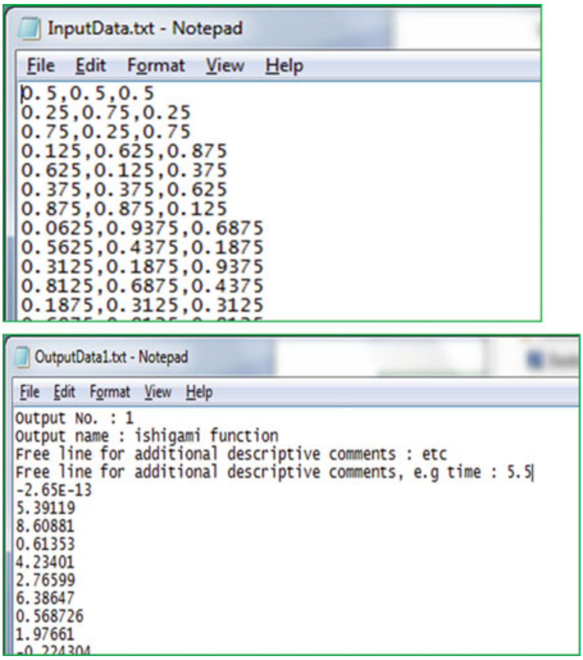
```
Inputs(1,2,...), Distributions(uniform, normal, ...)
2, uniform
3, normal
1, exponential
```

Fig. 3 Example of InputDistribution.txt file



```
function y = user_defined(varargin)
%
% n is number of inputs; it should be set to the value required by the
% function to be examined
n = 3;
%
% the lines of code in this section need not be altered
if nargin == 1
    str = class(varargin{1});
    if strcmp(str,'double') == 1
        x = ones(n,length(varargin{1}(1,:)));
        for i = 1:n
            x(i,:) = varargin{1}(i,:);
        end
    else
        x = ones(n,length(varargin{1}{1}));
        for i = 1:n
            x(i,:) = varargin{1}{i};
        end
    end
else
    x = ones(n,length(varargin{1}));
    for i = 1:n
        x(i,:) = varargin{i};
    end
end
%
% the lines of code in this section should be altered to reflect the
% function to be examined
xx1 = pi*(2*x(1,:)-1);
xx2 = pi*(2*x(2,:)-1);
xx3 = pi*(2*x(3,:)-1);
y = sin(xx1) + 7*(sin(xx2)).^2 + 0.1*(xx3.^4).*sin(xx1);
```

Fig. 4 Example of the user_defined function. File “user_defined.m”


InputData.txt - Notepad
File Edit Format View Help

```
0.5,0.5,0.5
0.25,0.75,0.25
0.75,0.25,0.75
0.125,0.625,0.875
0.625,0.125,0.375
0.375,0.375,0.625
0.875,0.875,0.125
0.0625,0.9375,0.6875
0.5625,0.4375,0.1875
0.3125,0.1875,0.9375
0.8125,0.6875,0.4375
0.1875,0.3125,0.3125
```

OutputData1.txt - Notepad
File Edit Format View Help

```
Output No. : 1
Output name : ishigami function
Free line for additional descriptive comments : etc
Free line for additional descriptive comments, e.g time : 5.5
-2.65E-13
5.39119
8.60881
0.61353
4.23401
2.76599
6.38647
0.568726
1.97661
-0.224304
```

Fig. 5 Example of "InputData.txt" and "OutpuData.txt" files

these files are also used in testing the HDMR approximation. This option can be used to build a metamodel using input data provided by either SobolHDR or by any other source (e.g., the third party software) and the matching output data given by any other source (e.g., the third party software). Multiple outputs can be used for the same inputs. In particular, this methodology can be used to build the so-called fully operational metamodels (FEOM), e.g., time-dependent metamodels (see Appendix 1 for further details).

Sample data files are provided with the GUI package to indicate the format required for such files—an extract is shown in Fig. 5. The first line in the output file should contain the output number corresponding to the output file name (e.g., "Output No. : 1" for "OutputData1.txt" and "Output No. : 2" for "OutputData2.txt"); the second/third line in the output file could contain more descriptive comments about the particular output; the fourth line should contain a numeric number representing the output, for example, the time-step corresponding to the output for a time-dependent output. The input samples must be normalized to [0, 1].

The sixth option ("genRandomNum") in the "Function to call" cell can be used if the user is interested in only generating random or quasi-random numbers for using them as an input after corresponding normalization for a third party software.

The fifth ("Maximum number of alfas") and sixth ("Maximum number of betas") parameters allow to set the maximum order of polynomials for approximation of the first- and the second-order terms in the HDMR decomposition.

The seventh ("Tolerance for alfas") and eighth ("Tolerance for betas") parameters represent the tolerance used for calculating the optimum order of polynomials for approximation of the first-order and the second-order terms in the HDMR decomposition. Note that smaller values of tolerances can result in the higher CPU time.

The ninth ("Iterations for VR based HDMR") and tenth ("Relative error threshold VR-HDMR") parameters are used in the variance reduction method of approximation for HDMR (*see* ref. [14]), representing the number of iterations and the relative error threshold for the variance reduction method. These parameters are used if the "VR-HDMR" option is chosen in the "Select the test to run" cell (*see* Subheading 2.3). The "Relative error threshold VR-HDMR" parameter is also made use of in the random sampling method of the HDMR approximation.

The eleventh ("Iterations for Sobol indices") to thirteenth ("Maximum number of points Sobol indices") parameters are used if the "Sobol_SI" option is chosen in the "Select the test to run" cell (*see* Subheading 2.3). Currently, Sobol indices using direct Sobol formulas can be used for the in-built functions only.

The fourteenth parameter ("Number of input variables") is the number of inputs in the model to be used. The default values for the in-built Homma-Saltelli, g-function, and the test function A1 used in Kucherenko et al. [15] is 3, 8, and 10, respectively. For user-defined functions, this value would represent the number of inputs in the "user_defined" function (see "user_defined.m" file in Fig. 5). For tabulated data, this parameter should be equal to the number of columns in the input data file.

The fifteenth parameter ("Number of output variables") is the number of outputs; this option is only active when the "tabulated_data" option is chosen as the function to call. Different metamodels are built corresponding to each output.

2.2.2 Settings Panel: Method_2

The first input parameter in the Settings Panel of Fig. 2b (named "Type of sampling") enables the choice of the sampling strategy for sampling inputs. Two options are available: a standard uniform Monte Carlo ("mc") sampling and the Quasi-Monte Carlo ("QMCSobol") sampling based on the Sobol sequences.

The second ("Initial Space") and the third ("Transformation space") parameters enable to specify the type of the probability distribution of the input variables in the initial and transformed spaces. There are two options: "Uniform" and "Gaussian."

The fourth input parameter ("Case") offers the option of specifying the model to be used. It has five options: the three options "Gfunction," "Ishigami," and "Hyperplane" are in-built functions representing the Sobol g-function, the Ishigami function, and the hyperplane function. The fourth option ("user_defined") allows the user to supply the function to be evaluated in a MATLAB "user_defined.m" file and must be named as such. The "user_defined.m" file can also call any other file (e.g., it can be linked with gPROMS®). It simply represents the entry point for a user-defined function.

The inputs to the user-defined function are normalized to [0, 1] but can be converted back to their real values in the "user_defined.m" file. The syntax for a sample case (the Ishigami function) is shown in Fig. 4. The fifth option ("tabulated_data") in the "Function to call" cell provides the capability for the user to supply the input–output mapping in a form of text data files. The input and output data sample files for analysis must be named "InputData.txt" and "OutputDataX.txt," respectively, where "X" is an integer beginning from 1 to m, with "m" being the total number of output files; these files are also used in testing the HDMR approximation. Sample data files are provided with the GUI package to indicate the format required for such files—an extract is shown in Fig. 5.

The first line in the output file should contain the output number corresponding to the output file name (e.g., "Output No. : 1" for "OutputData1.txt" and "Output No. : 2" for "OutputData2.txt"); the second/third line in the output file could contain more descriptive comments about the particular output; the fourth line should contain a numeric number representing the output, for example, the time-step corresponding to the output for a time-dependent output. The input samples must be normalized to [0, 1].

The fifth parameter ("Number of samples") represents the total number of samples to be used. Its value should be a multiple of 2.

The sixth parameter ("Dimension of Input") is the number of inputs in the model. The default value for the in-built Ishigami and Sobol g-functions is 3 and 8, respectively. For the Hyperplane in-built function the dimension can be specified by the user. For user-defined functions, this value would represent the number of inputs in the model given in the "user_defined.m" file. For the tabulated data, this parameter should be equal to the number of columns in the input data file.

The seventh parameter ("Dimension of Output") is the number of outputs; this option is only active when the "tabulated_data" option is chosen as the function to be used. Different metamodels are built corresponding to each output.

The eighth parameter ("Maximum interaction order") enables to choose between the first- and second-order HDMR decomposition.

The ninth ("Maximum 1st order") and tenth ("Maximum 2nd order") parameters allow to set the maximum order of polynomials for approximation of the first- and second-order terms in the HDMR decomposition.

2.3 "Run" Option

The "Run" panel includes the option to select which test to carry out. The "RS_HDMR" option is tool for building metamodels with subsequent sensitivity analysis. The "VR_HDMR" option applies a variance reduction approach for building the HDMR metamodels. The "SOBOL_SI" option is based on the direct GSA method proposed by Sobol' for evaluating sensitivity indices. The "DGSA" option results in evaluating derivative-based global sensitivity analysis measures (DGSA) based on previously reported work [3, 11, 12]. The "RBF_MODEL" option creates a metamodel based on radial basis functions using MATLAB toolbox. Once the metamodel is built, sensitivity indices are then evaluated using the direct Sobol' method [16].

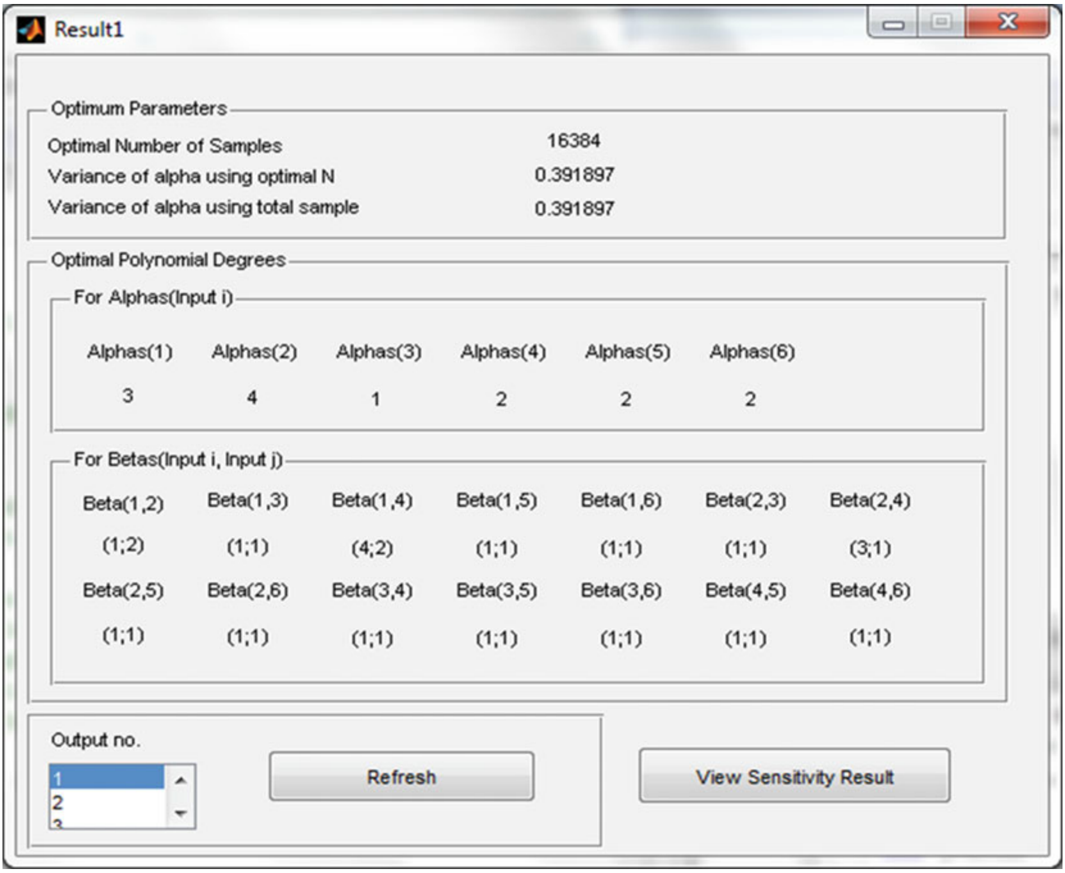
The "Run" panel also offers the option ("Yes/No") of writing the HDMR-generated metamodel to an m-file called "Parabola. m." It can take in the 0–1 inputs and produce the HDMR approximation output(s). It also generates a self-contained metamodel to a C++ Parabola.cpp in the "CPP_Code" directory. For tabulated data, this option (if "Yes" is chosen) creates a file "outParam.m" containing information about each output extracted from the corresponding output file.

The "Run" button writes the GUI inputs to a text file ("userGUIparam.txt"), builds a metamodel, calculates Sobol sensitivity indices or DGSA (depending on the chosen option), reports the tests carried out and the time elapsed in the "Message" box, and produces a result folder named "Results" (or overwrites files in any existing one) in the same directory as SobolHDMR GUI files. This result folder contains *.csv result files depending on the tests carried out.

2.4 "Results1" Panel

If "RS_HDMR" option was selected, then clicking on "View Results" after the run has executed (or typing "Result1" at the MATLAB command workspace) would produce a new figure window such as that shown in Fig. 6. This window makes use of the result files: "Results\test1_ConvAlphaNew.csv," "Results\test1_ConvAlphaOld.csv," and "Results\test2_OptPolyDeg.csv." The top panel displays the optimal number of sampled points which was obtained as a result of convergence monitoring (see Feil et al. [8] for details), the variance of the alpha-coefficients monitored as functions of the number of sampled points using the optimal number of samples and that using the total number of samples.

The middle panel displays the optimum polynomial order for each input, needed for the approximation of the first-order terms in the HDMR decomposition. A maximum of seven input-alphas optimum numbers can be displayed. The bottom panel displays the optimum polynomial order for each input pair, needed for the approximation of the second-order terms in the HDMR decomposition. A maximum of 14 pairs of input-betas can be displayed. The "Refresh" button is used to update the panel and



Result1

Optimum Parameters

Optimal Number of Samples	16384
Variance of alpha using optimal N	0.391897
Variance of alpha using total sample	0.391897

Optimal Polynomial Degrees

For Alphas(Input i)

Alphas(1)	Alphas(2)	Alphas(3)	Alphas(4)	Alphas(5)	Alphas(6)
3	4	1	2	2	2

For Betas(Input i, Input j)

Beta(1,2)	Beta(1,3)	Beta(1,4)	Beta(1,5)	Beta(1,6)	Beta(2,3)	Beta(2,4)
(1;2)	(1;1)	(4;2)	(1;1)	(1;1)	(1;1)	(3;1)
Beta(2,5)	Beta(2,6)	Beta(3,4)	Beta(3,5)	Beta(3,6)	Beta(4,5)	Beta(4,6)
(1;1)	(1;1)	(1;1)	(1;1)	(1;1)	(1;1)	(1;1)

Output no.

1
2
3

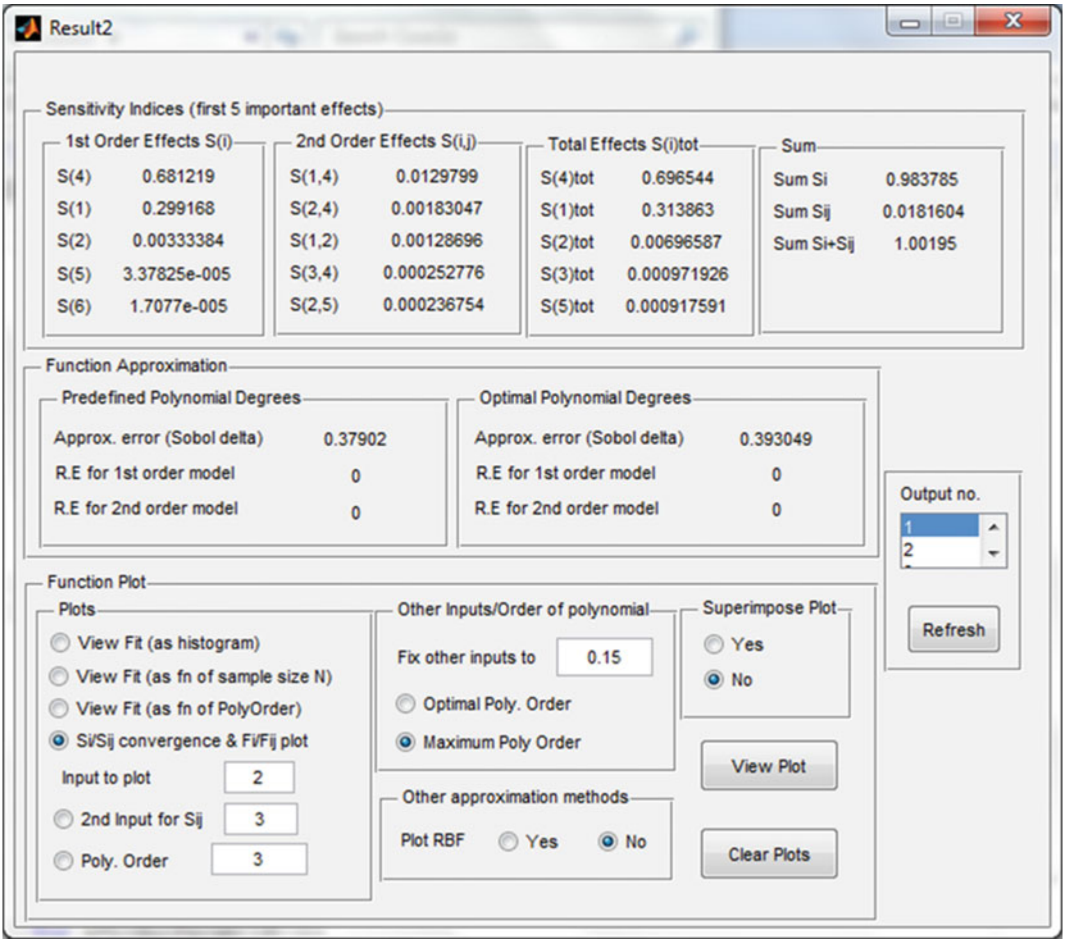
Fig. 6 "Result1" panel

view the results for the chosen output in the list-box "Output no." A number of outputs correspond to the value of the parameter "Number of output variables."

2.5 Sensitivity Analysis Results and Plots

Clicking the "View Sensitivity Result" (or typing "Result2" at the MATLAB command workspace if the RS_HDMR option was selected) launches a new window such as that shown in Fig. 7. The window makes use of the result files: "Results\test2_OptPolyDeg.csv," "Results\test3_SensIndHDMR.csv," "Results\test4a_FuncApp.csv," and "Results\test4b_FuncApp.csv" (these *.csv files are produced only when the RS_HDMR option is selected).

The "Refresh" button is used to update the panel and view the results for the chosen output in the list-box "Output no." The top panel shows the sensitivity indices; this panel has four subpanels with three showing the first-order, second-order, and total effects of a maximum of five inputs arranged in descending order, while the fourth subpanel shows the sum of the first-order and second-order effects as well as their combined sum.



Sensitivity Indices (first 5 important effects)

1st Order Effects S(i)		2nd Order Effects S(i,j)		Total Effects S(i)tot		Sum	
S(4)	0.681219	S(1,4)	0.0129799	S(4)tot	0.696544	Sum Si	0.983785
S(1)	0.299168	S(2,4)	0.00183047	S(1)tot	0.313863	Sum Sij	0.0181604
S(2)	0.00333384	S(1,2)	0.00128696	S(2)tot	0.00696587	Sum Si+Sij	1.00195
S(5)	3.37825e-005	S(3,4)	0.000252776	S(3)tot	0.000971926		
S(6)	1.7077e-005	S(2,5)	0.000236754	S(5)tot	0.000917591		

Function Approximation

Predefined Polynomial Degrees		Optimal Polynomial Degrees	
Approx. error (Sobol delta)	0.37902	Approx. error (Sobol delta)	0.393049
R.E for 1st order model	0	R.E for 1st order model	0
R.E for 2nd order model	0	R.E for 2nd order model	0

Output no.
1
2

Function Plot

Plots
○ View Fit (as histogram)
○ View Fit (as fn of sample size N)
○ View Fit (as fn of PolyOrder)
● Si/Sij convergence & Fi/Fij plot
Input to plot: 2
○ 2nd Input for Sij: 3
○ Poly. Order: 3

Other Inputs/Order of polynomial
Fix other inputs to: 0.15
○ Optimal Poly. Order
● Maximum Poly Order

Other approximation methods
Plot RBF ○ Yes ● No

Superimpose Plot
○ Yes
● No

Fig. 7 "Result2" window

The middle panel shows the approximation error ("Sobol' delta"), and the relative error for the first- and second-order HDMR metamodel both using the predefined (maximum) polynomial degrees specified in the HDMR GUI (Fig. 2a) and the optimal polynomial degrees (shown in Fig. 6).

The bottom panel can be used to view several plots. The "View Plot" button is used to view the specified plot(s).

2.6 Viewing Plots

Here we detail the "View Plot" option available at the bottom of the panel of the "Result2" window (Fig. 7).

2.6.1 "View Fit (as Histogram)" Option

The "View Fit (as histogram)" radio-button specifies a plot showing the quality of the HDMR fit in the form of two histograms; one for the original function/data and the other for the HDMR approximation output. The HDMR approximation histogram is generated using either the optimal or maximum

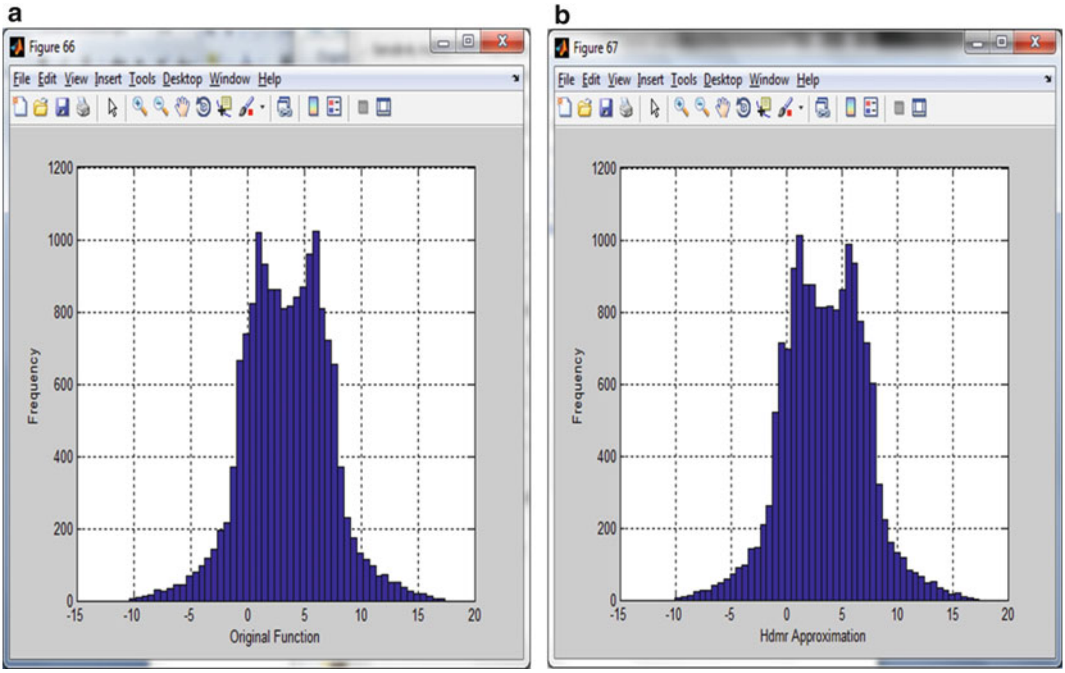


Fig. 8 Histogram of model output for (**a**) original model/data; (**b**) HDMR metamodel

polynomial degrees depending on which of "Optimal Poly. Order" or "Maximum Poly. Order" radio-buttons is chosen. An example of such plots is shown in Fig. 8a, b. This option also produces a histogram of the residual (the difference between the original output and the HDMR metamodel output), an example of which is shown in Fig. 9a. A plot of the residual versus a chosen input (as specified in the text box "Input to plot") is also produced as shown in Fig. 9b.

2.6.2 "View Fit (as fn of Sample Size N)" Option

The radio-button marked "View Fit (as fn of sample size N)" can be used to specify a plot showing the fit of the HDMR metamodel as a function of a specified input (given in the editable box "Input to plot") for different number of sampled points N for in-built functions or user-defined functions; the other inputs are fixed to the normalized value given in the editable text field "Fix other inputs to" (Fig. 10a). For tabulated data, the "View Fit (as fn of sample size N)" radio-button specifies that a scatter plot be displayed showing the HDMR metamodel versus the original data (Fig. 10b). The "View Fit (as fn of sample size N)" radio-button is used in conjunction with the "Optimal Poly. Order" and "Maximum Poly. Order" radio-buttons; these buttons specify whether the fit is constructed using the optimal polynomial degrees (also known as "optimal alpha/beta") or the maximum polynomial degrees (also known as "maximum number of alpha/beta") for all inputs.

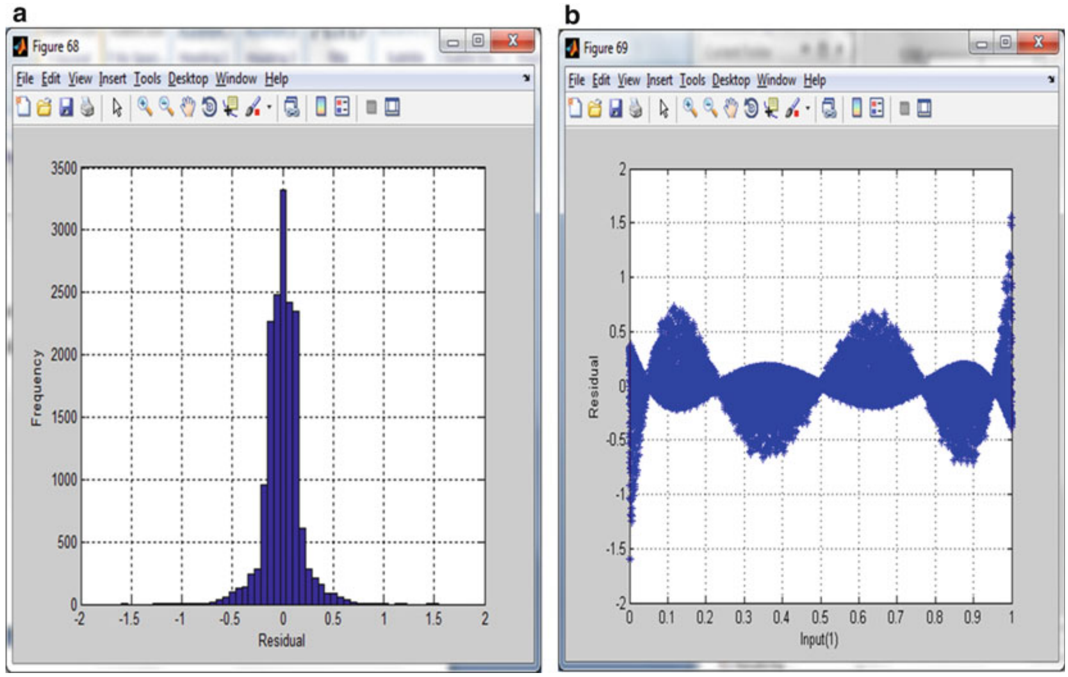


Fig. 9 (**a**) Histogram of the residual; (**b**) residual for a specific input

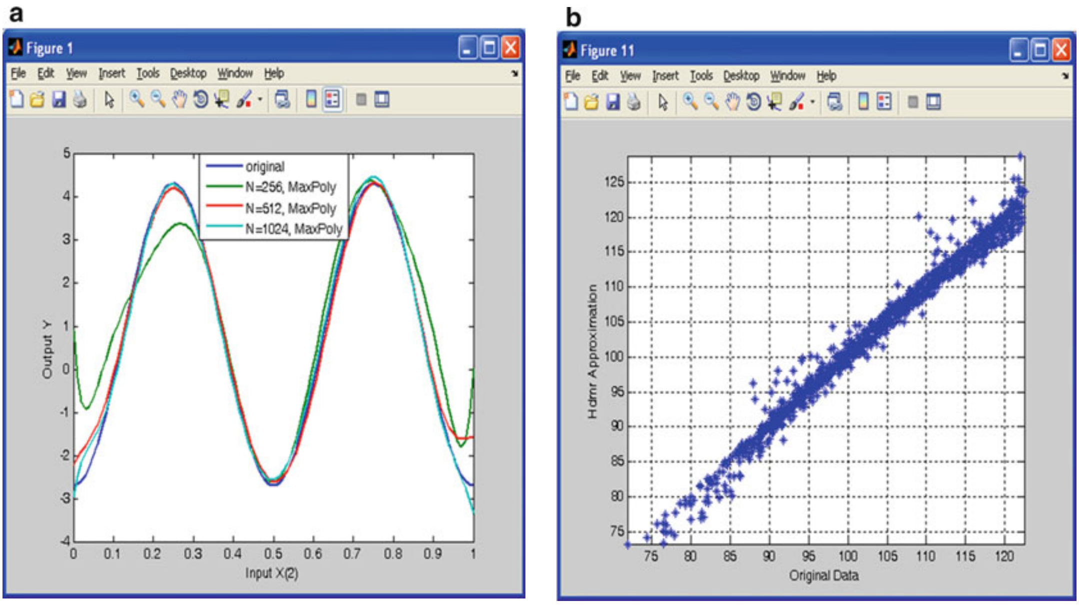


Fig. 10 (**a**) Metamodel output at different number of sampled points for a specified input when other inputs are fixed; (**b**) scatter plot showing original data versus HDMR approximation

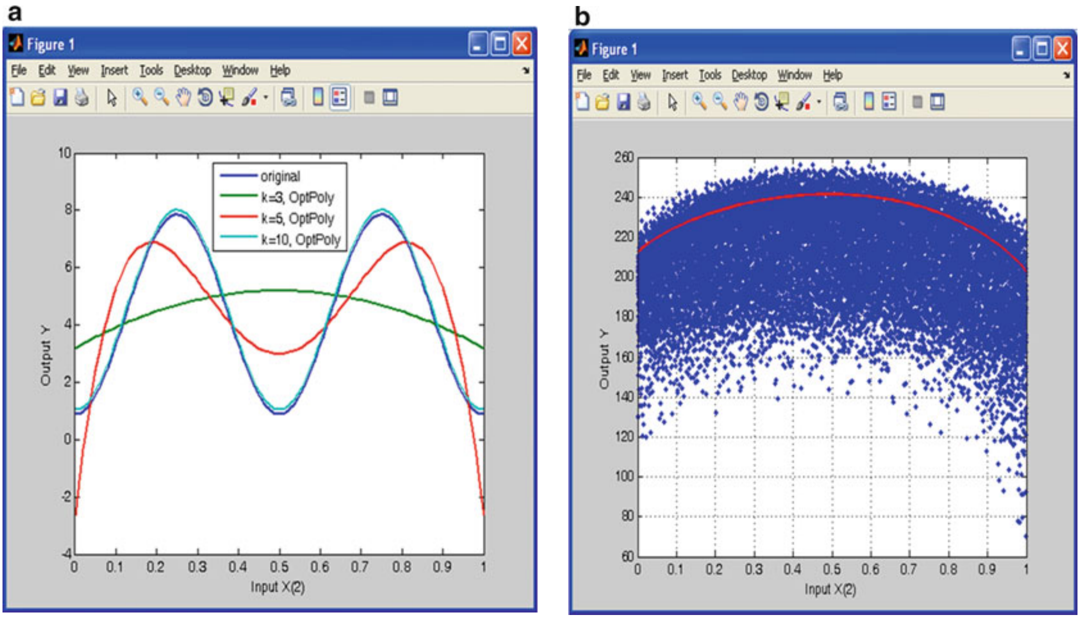


Fig. 11 (**a**) Metamodel output at different polynomial orders for a specified input when other inputs are fixed; (**b**) scatter diagram of original output data and a line diagram of the HDMR metamodel for a specified input (*red line*) when other inputs are fixed

2.6.3 "View Fit (as fn of PolyOrder)" Option

The "View Fit (as fn of PolyOrder)" radio-button, for in-built functions or user-defined functions, specifies a plot showing the quality of the HDMR fit for a specified input (given in the editable box "Input to plot") as a function of its maximum first-order polynomial degree, which can be set in the editable box "Poly. Order." For this plot, the input specified is plotted over the range of normalized (0, 1) values; the other inputs are fixed to the normalized value given in the editable text field "Fix other inputs to," using either their optimal or maximum polynomial degrees depending on which of "Optimal Poly. Order" or "Maximum Poly. Order" radio-buttons is chosen. An example of such a plot is shown in Fig. 11. This option also shows a scatter diagram of the output and a line diagram showing the HDMR output for a specified input when all other inputs are fixed to the value set in the "Fix other inputs to" field.

2.6.4 Superimpose Plot

The "Superimpose Plot" subpanel is used to specify whether or not to superimpose plots (e.g., to view the HDMR metamodel output as a function of different number of samples or different polynomial degrees for a specified input). To view the HDMR output as a function of different number of samples, it is necessary to click "Refresh" anytime a new result for a different sample size is generated, before clicking on "View Hdmr" to see the superimposed plot.

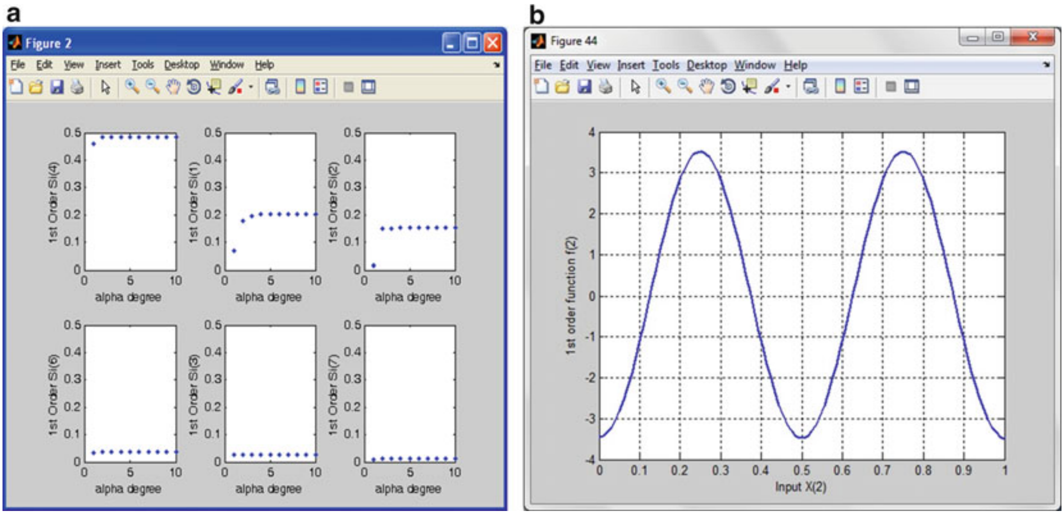


Fig. 12 (**a**) Values of the first-order sensitivity indices versus polynomial degree; (**b**) first-order component function for a specified input

2.6.5 Si/Sij Convergence & Fi/Fij Plot

The "Si/Sij convergence & Fi/Fij plot" radio-button is used to specify two classes of plots namely a plot of the convergence of the first- and second-order sensitivity indices, and a plot of the first- and second-order component functions. The second-order sensitivity indices and the second-order component functions plots are only displayed if the radio-button "second Input for Sij/Fij" is selected. The values of the first-order sensitivity indices are plotted as a function of the polynomial degree ("alpha degree") in the decomposition of the first-order terms in the HDMR decomposition (Fig. 12a). Figure 12b shows the values of the second-order indices as a function of polynomial degrees in the decomposition of the second-order terms in the HDMR decomposition for two specified inputs ("number of betas"). The component functions for the specified inputs are plotted over the whole range of such inputs (Fig. 13).

When the number of outputs is greater than 1, plots of the first and total sensitivity indices as functions of the fourth parameter specified in the output data files (typically this parameter can be time) are also produced. Figure 14 shows an example of this type of plots. Also, a plot of the HDMR output corresponding to the first three data input points in "InputData.txt" file as a function of the output file fourth parameter (e.g., time) is produced (Fig. 15).

2.7 Calculation of Sobol Sensitivity Indices or Derivative-Based Global Sensitivity Analysis Measures (DGSA)

From the main window (Fig. 2a), if "SOBOL_SI" or "DGSA" was selected, then clicking on "View Results" on GUI (Fig. 2a) after the run executes (or running "Result3.m" from the MATLAB command prompt) would produce either of the two screen-shots shown in Fig. 16.

This window provides the options of specifying which input at what number of the sampled points ("Sample size N") to plot.

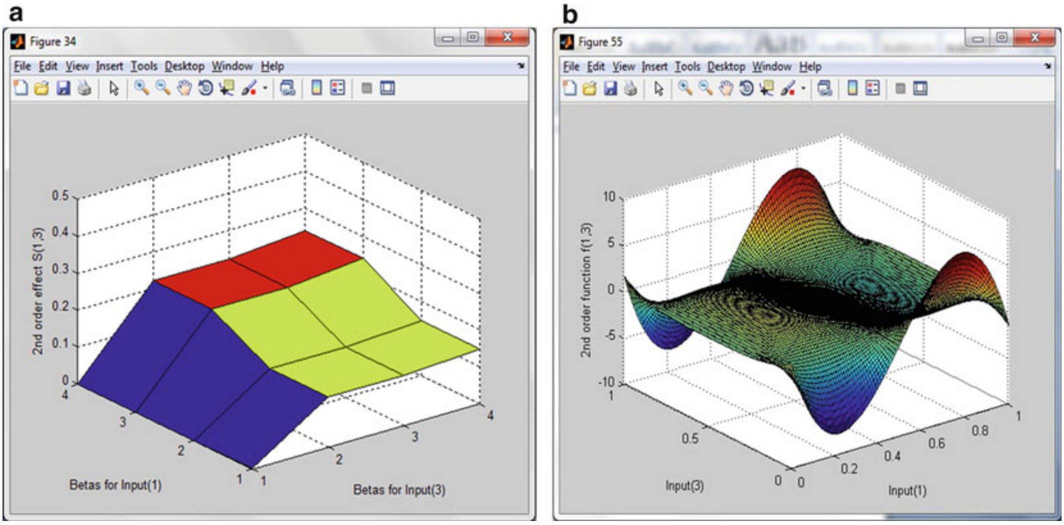


Fig. 13 (**a**) Values of the second-order sensitivity indices versus polynomial degrees for two specified inputs; (**b**) profile of the second-order component function for two specified inputs

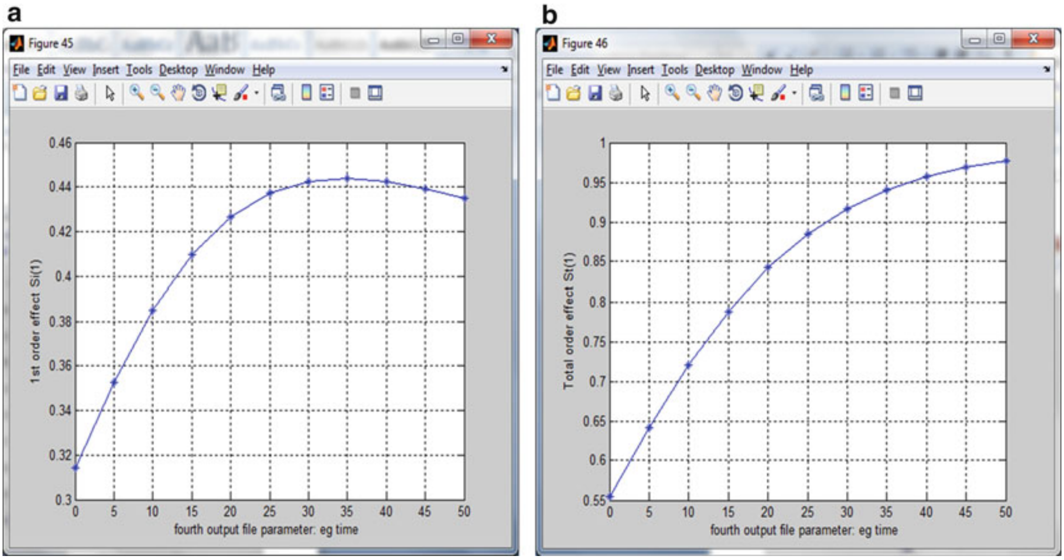


Fig. 14 (**a**) First-order sensitivity indices versus time for a specified input; (**b**) total-order sensitivity indices versus time for a specified input

There is also the provision to choose any two of four different types of plot. For the "SOBOL_SI" option, the "Si versus Input" specifies a plot of first-order sensitivity indices for all of input parameters for the chosen samples size (the values of sensitivity indices are linked by a solid line for comparison purposes).

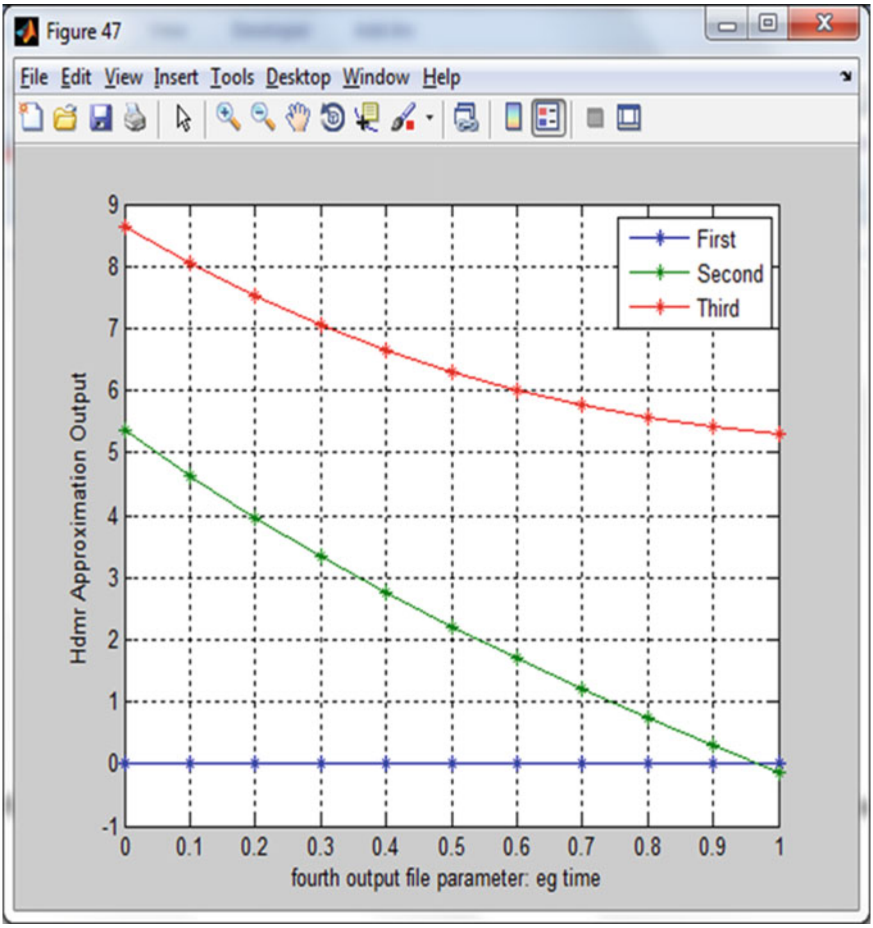
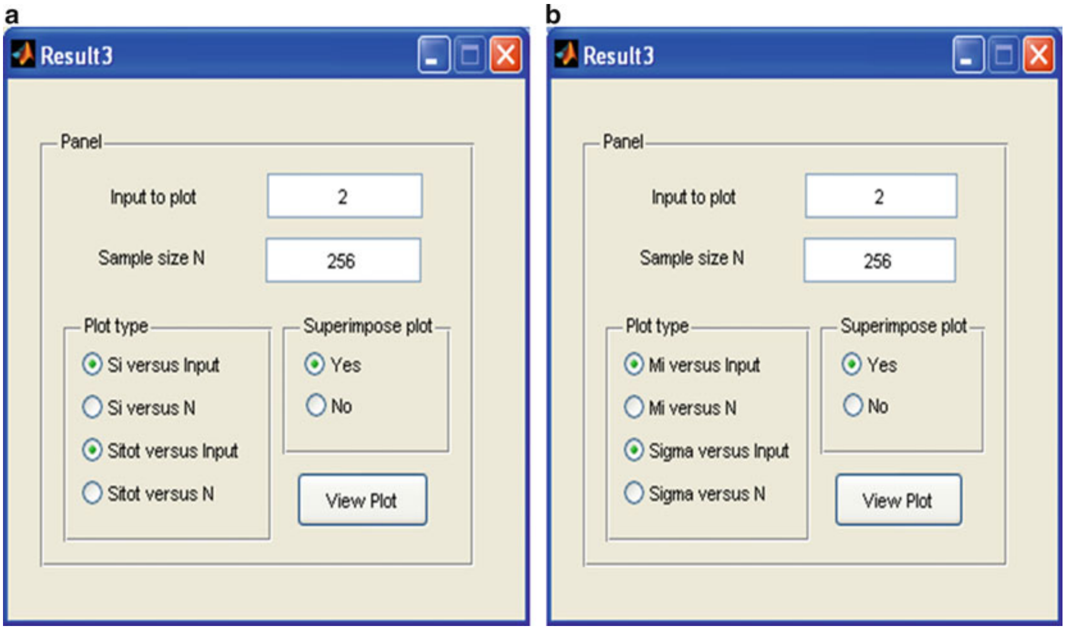


Fig. 15 HDMR Output as a function of time (for first three input data points)

The corresponding option for DGSA ("Mi versus Input") specifies the mean measures ("Mi") for all of input parameters for the chosen samples size. Examples of both plots are shown in Fig. 17a, b. Similarly, "Si versus N" or "Mi versus N" plots first-order sensitivity indices or mean effects versus the number of samples for the specified input parameter, respectively. These are in fact convergence plots. Figure 17c, d shows such plots. "Sitot versus Input," "Sitot versus N," "Sigma versus Input," and "Sigma versus N" specify plots of the total sensitivity indices or sigma-measure as a function of input parameter or number of samples depending on the chosen option. Figures 18 are examples of such plots. The "Yes/No" buttons indicate whether the plots should be superimposed (such as that shown in Figs. 17 and 18) or not.



(a)

Panel

Input to plot: 2

Sample size N: 256

Plot type

● Si versus Input
○ Si versus N
● Sitot versus Input
○ Sitot versus N

Superimpose plot

● Yes
○ No

(b)

Panel

Input to plot: 2

Sample size N: 256

Plot type

● Mi versus Input
○ Mi versus N
● Sigma versus Input
○ Sigma versus N

Superimpose plot

● Yes
○ No

Fig. 16 Window produced when (**a**) SOBOL_SI is chosen, (**b**) DGSA is chosen

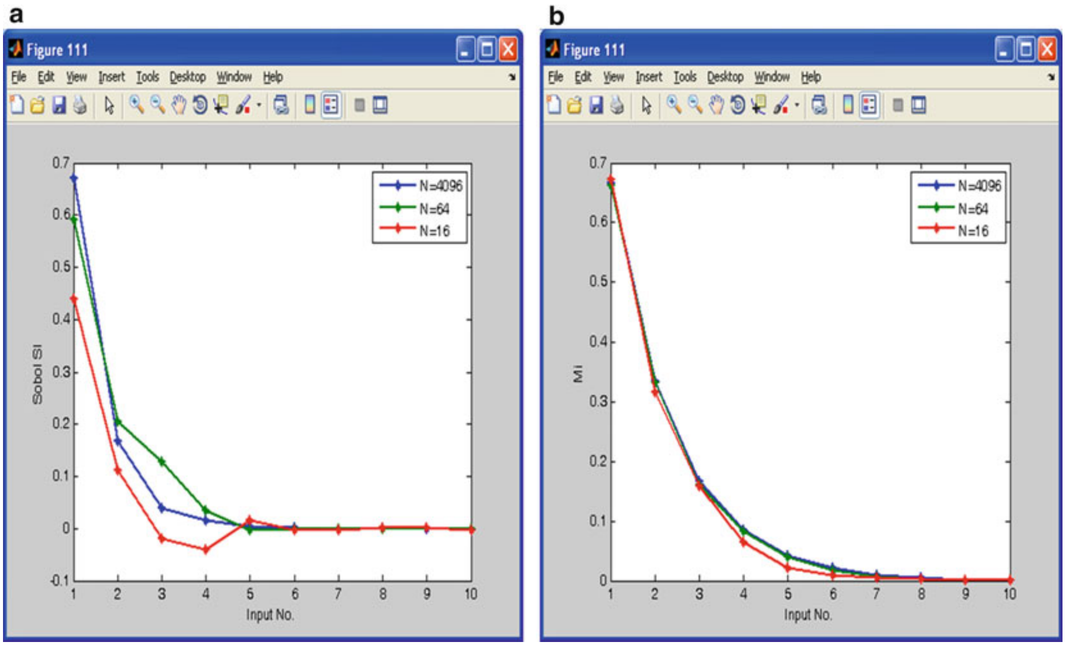


Fig. 17 (**a**) First-order sensitivity indices for all inputs; (**b**) mean measure for all inputs, (**c**) first-order sensitivity indices versus *N*, (**d**) mean measure versus *N*

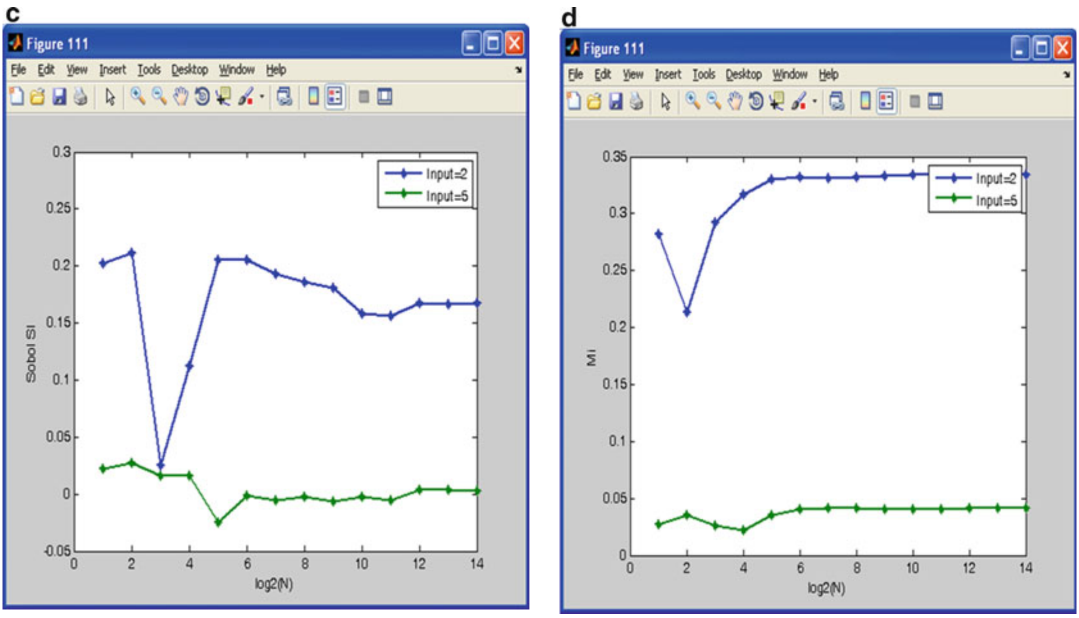


Fig. 17 (continued)

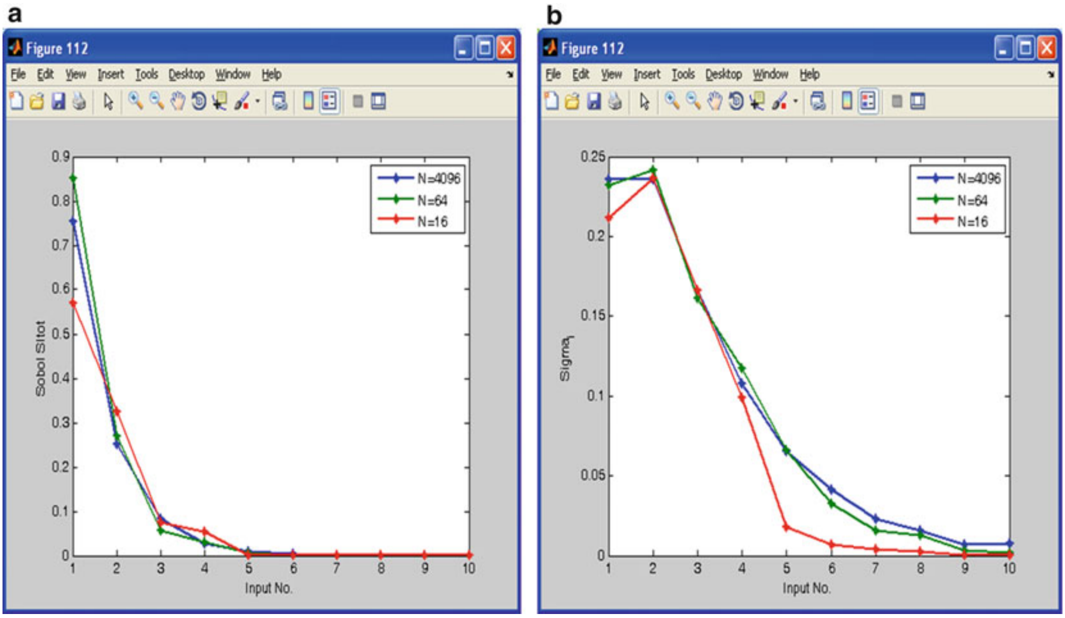


Fig. 18 (**a**) Total sensitivity indices for all inputs, (**b**) sigma-measure for all inputs versus Input, (**c**) total sensitivity indices versus *N*, (**d**) sigma-measure versus *N*

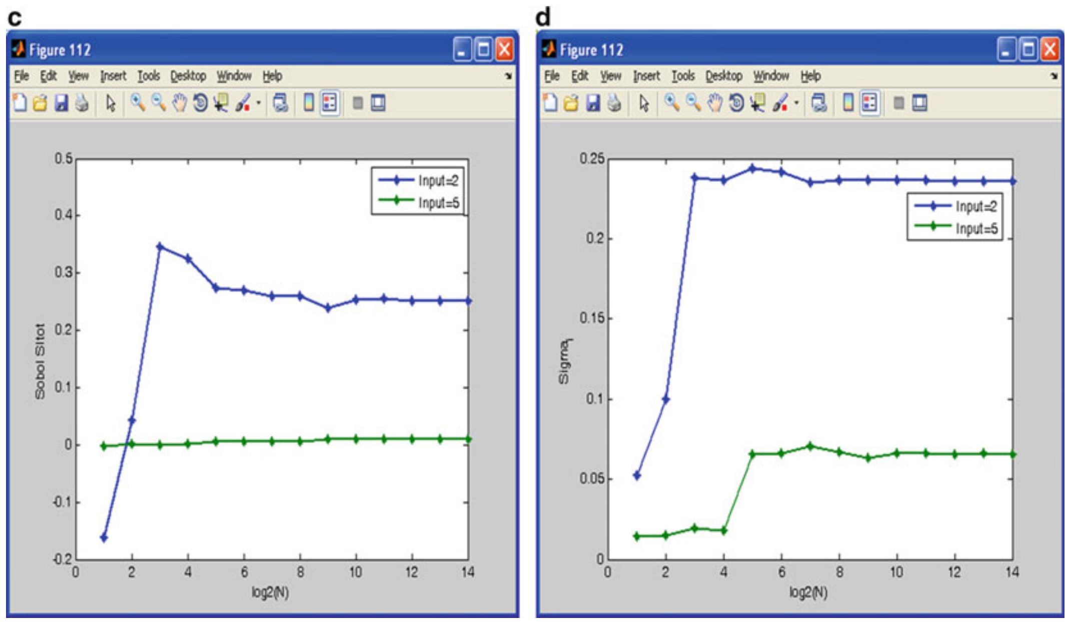


Fig. 18 (continued)

Appendix 1: Fast Equivalent Operational Model

Problem Statement

Consider a system of ordinary differential equations (ODE) with uncertain parameters:

$$\frac{d\mathbf{y}}{dt} = F(\mathbf{y}, \mathbf{p}, t)$$
$$\mathbf{y}(t = 0) = \mathbf{y}0(\mathbf{p}) \qquad (1.1)$$

Here $\mathbf{p}$ is the vector of uncertain static parameters.

The objective is to approximate $\mathbf{y}(t_i^*, \mathbf{p}),\ i = 1, \ldots, n$ at specific time t_i^* points with Quasi Random Sampling-High Dimensional Model Representation (QRS-HDMR) models. The original model can be expensive to run while the set of QRS-HDMR models also known as Fast Equivalent Operational Model (FEOM) can be run in milliseconds.

Solution Procedure

Sample N points of the vector $\{\mathbf{p}_j\}$, $j = 1, \ldots, N$ (we recall that vector $\mathbf{p}$ is an input of the HDMR model), for all $\mathbf{p}_j$ solve ODE Eq. 1.1 and obtain $K \times N$ outputs $\mathbf{y}(t_k^*, \mathbf{p}_j)$, $k = 1, \ldots, K$, $j = 1, \ldots, N$. Using these data as the input–output samples, build a set of the HDMR models (FEOM).

Test case:

Consider an ODE:

$$\frac{df}{dt} = a'_t \sin^2 X_2 + b'_t\ X_3^4 \sin X_1 \tag{1.2}$$

with initial conditions given by the Ishigami function:

$$f_{t=0} = \sin X_1 + a_{t=0} \sin^2 X_2 + b_{t=0} X_3^4 \sin X_1$$

where

$$a_t = 7\exp(-t), \quad b_t = 0.1\exp(t)$$

$$\frac{da_t}{dt} = a'_t, \quad \frac{db_t}{dt} = b'_t$$

Here $X_1,\ X_2,\ X_3$ are random variable with a probability distribution function given by:

$$p_i(X_i) = \begin{cases} \frac{1}{2\pi}, & \text{if } -\pi \le X_i \le \pi \\ 0, & \text{if } X_{i<-\pi, X_i>}\ \pi \end{cases}, \quad \text{for } i = 1, 2, 3$$

The explicit solution to the above ODE is:

$$f_t = \sin X_1 + a_t \sin^2 X_2 + b_t X_3^4 \sin X_1$$

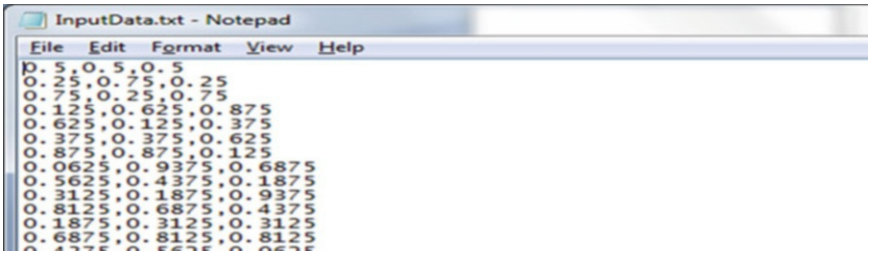
At each moment of time the total variance and partial variances can be calculated explicitly [22]:

$$\begin{aligned} D &= \frac{a_t^2}{8} + \frac{b_t \pi^4}{5} + \frac{b_t^2 \pi^8}{18} + \frac{1}{2} \\ D_1 &= \frac{b_t \pi^4}{5} + \frac{b_t^2 \pi^8}{50} + \frac{1}{2} \\ D_2 &= \frac{a_t^2}{8} \\ D_3 &= 0 \\ D_{12} &= 0 \\ D_{13} &= \frac{b_t^2 \pi^8}{18} - \frac{b_t^2 \pi^8}{50} \\ D_{23} &= 0 \\ D_{123} &= 0 \end{aligned}$$

For each time-step $t = 0.0,\ 0.1,\ 0.2, \ldots$ the ODE Eq. 1.2 is solved and a HDMR model is built using the corresponding output. FEOM can then be compiled by combining HDMRs for all time-steps.

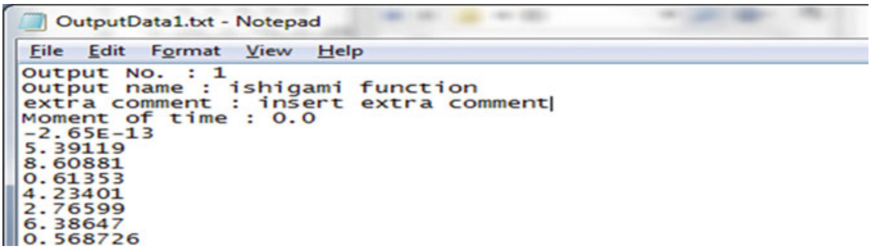
Steps:

1. Generate N Sobol (or random for the MC method) points for the input variables X_1, X_2, X_3 and store in the file “IO/InputData.txt” (subdirectory “IO”). Its content looks like this (for the QMC method):



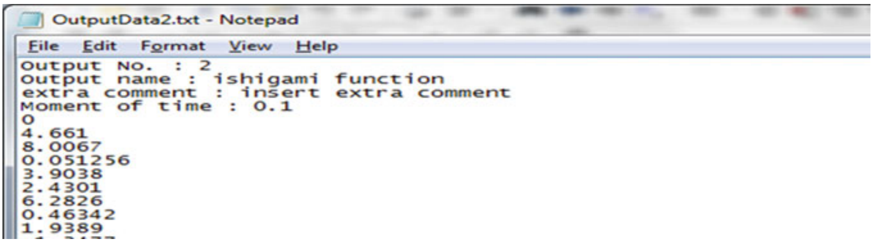
```
0.5,0.5,0.5
0.25,0.75,0.25
0.75,0.25,0.75
0.125,0.625,0.875
0.625,0.125,0.375
0.375,0.375,0.625
0.875,0.875,0.125
0.0625,0.9375,0.6875
0.5625,0.4375,0.1875
0.3125,0.1875,0.9375
0.8125,0.6875,0.4375
0.1875,0.3125,0.3125
0.6875,0.8125,0.8125
```

2. For the time-step $t = 0.0$ solve ODE Eq. 1.2 for each of N random or Sobol points to obtain the corresponding output. Store the output in the file “IO/OutputData1.txt.” The fourth line in this output file should contain the value of the time-step that was used.



```
Output No. : 1
Output name : ishigami function
extra comment : insert extra comment|
Moment of time : 0.0
-2.65E-13
5.39119
8.60881
0.61353
4.23401
2.76599
6.38647
0.568726
```

3. For the time-step $t = 0.1$ solve ODE Eq. 1.2 for each of the N random or Sobol points to obtain the corresponding output. Store the output in the directory “IO/OutputData2.txt.” The fourth line in this output file should contain the value of the time-step that was used.



```
Output No. : 2
Output name : ishigami function
extra comment : insert extra comment
Moment of time : 0.1
0
4.661
8.0067
0.051256
3.9038
2.4301
6.2826
0.46342
1.9389
```

4. Repeat Step 3 for $t = 0.2,\ 0.3, \ldots,$ and store outputs files “IO/OutputData3.txt,” “IO/OutputData4.txt,” etc. Note that the fourth line in each output file should contain the contain the value of the corresponding time-step.

Running SobolHDMR with options “Function to call = tabulated_data” and “Number of outputs = 11” (currently there are 11

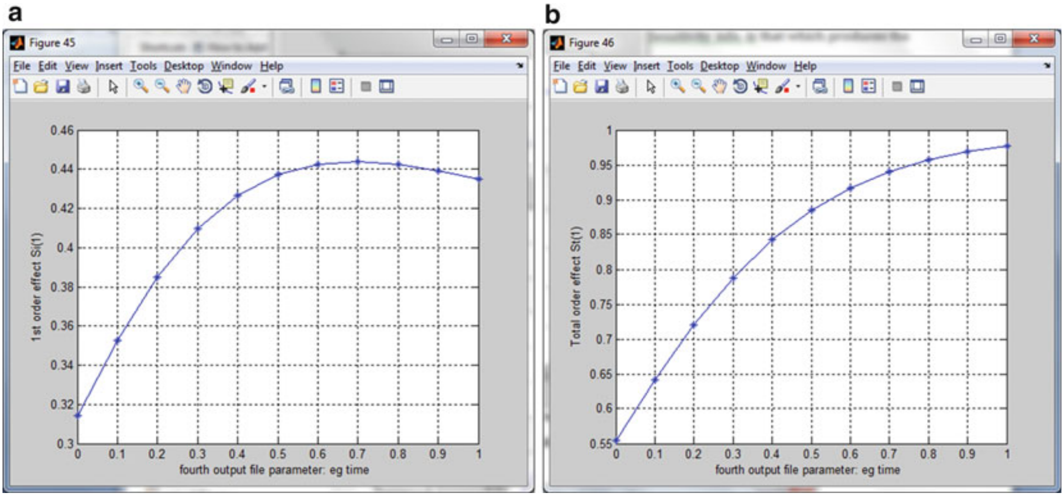


Fig. 19 (**a**) First-order sensitivity indices versus time for Input 1; (**b**) total-order sensitivity indices versus time for Input 1

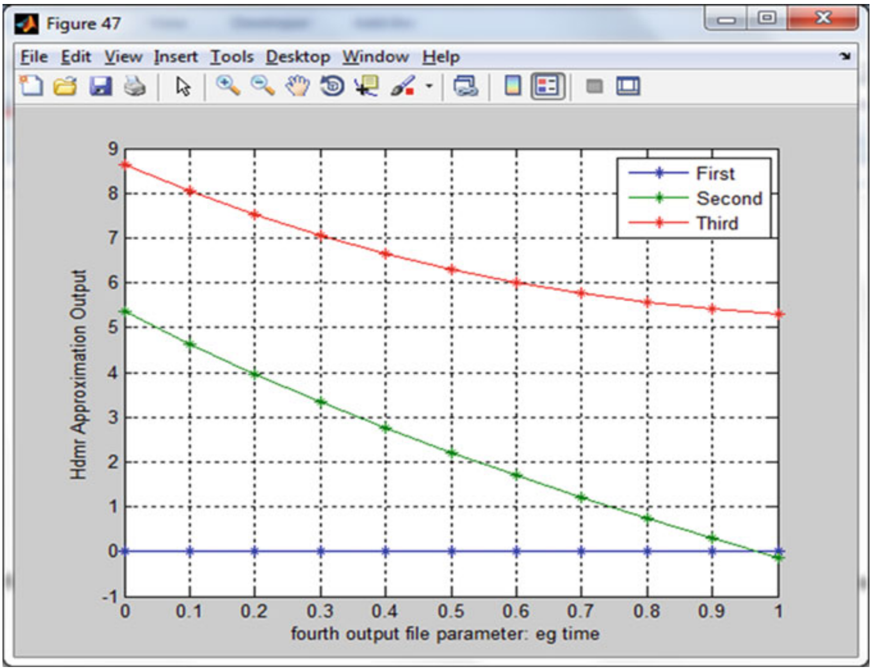


Figure 47
File Edit View Insert Tools Desktop Window Help
9
8
7
6
5
4
3
2
1
0
-1
Hdmr Approximation Output
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
fourth output file parameter: eg time
First
Second
Third

Fig. 20 HDMR output as a function of time (for the first three input data points)

output files in the "IO" folder) will create the FEOM for the above ODE example with $t = 0.0,\ 0.1, \ldots, 1.0$

The same GUI page that is used to view sensitivity results produces the following plots (Fig. 19):

Figure 20 presents the FEOM using the first three input points from the file "IO/InputData.txt"

Appendix 2: Theoretical Background

The deletion of high order members arises from a metamodeling context. Quite often in mathematical models, only relatively low order interactions of input variables have the main impact on the model output. For such models, the computation of the sensitivity terms of Eq. 2.29 is best carried out by the RS-HDMR technique as proposed by Li et al., which, as a metamodeling technique, has the more general utility of providing a representation of the input–output mapping over the whole input space [5].

Metamodels, also known as surrogate models, are resorted to when the underlying mathematical structure of a model is complex and contains many input variables. These approximate models are cheaper to evaluate than the original functions that they mimic. For black-box models or laboratory observations where mechanistic models do not exist, metamodels help provide a better understanding of the underlying relationship between inputs and outputs.

The underlying framework of all metamodeling techniques consists of a data collection strategy, the selection of a model type, and the fitting of the model to the data [17]. The fitting of the model is usually done by finding optimal values of certain model parameters that minimize an approximation-error function; such methods include least squares, best linear predictor, log-likelihood, and so on.

There are a variety of model types for approximating complex, multivariate functions. *Response surface methodologies* usually approximate the complex multivariate function by low order polynomials, such as

$$\tilde{Y} = \beta_o + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{i=1}^{k-1} \sum_{j>i}^{k} \beta_{ij} X_i X_j \qquad (1.3)$$

where $\tilde{Y}$ is an approximation to Eq. 2.5 with the parameters β_o, β_i, . . . being determined by some form of least square regression [18]. *Kriging* is an interpolative approximation method based on the weighted sum of the sampled data. It is a combination of a polynomial model plus departures which are realization of a random function [2]. *Neural networks* have been used to approximate a multivariate function as a nonlinear transformation of multiple linear regression models [18]. Although these techniques are useful for particular applications, they fall short in certain areas. *Response surface methodologies* are not accurate enough to approximate complex nonlinear multimodal profiles as they are based on simple quadratic models. *Kriging* models are difficult to obtain or even use [17]. Training of *neural networks* usually takes a lot of computing time [18]. A promising metamodeling tool for approximating complex, multivariate functions is the *HDMR*.

High Dimensional Model Representation

HDMR can be regarded as a tool for capturing high dimensional input–output system behavior. It rests on the generic assumption of only low order input correlations playing a significant role in physical systems. The HDMR expansions can be written in the following form for $f(x) \equiv f(x_1,\ x_2, \ldots, x_n)$ as

$$f(x) = f_o + \sum_i f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \cdots$$
$$+ f_{12\ldots k}(x_1, \ldots, x_n) \tag{1.4}$$

This decomposition is unique, called an ANOVA-HDMR decomposition, if the mean of each term with respect to its variable is zero as given in Eq. 2.25, resulting in pairs of terms being orthogonal [4]. Each term of the ANOVA-HDMR decomposition tells of the contribution of the corresponding group of input variables to the model output $f(x)$. The determination of all the terms of the ANOVA-HDMR requires the evaluation of high dimensional integrals, which would be carried out by Monte Carlo integration. For high accuracy, a large number of sample points would be needed. This represents a serious drawback of the ANOVA-HDMR. For most practical applications, rarely are terms beyond three-order significant [4]. Rabitz and coworkers proposed a Random Sampling-HDMR (RS-HDMR) which involves truncating the HDMR expansions up to the second or third order, and then approximating the truncated terms by orthonormal polynomials [5, 19].

Consider a piecewise smooth and continuous component function. It can be expressed using a complete basis set of orthonormal polynomials:

$$f_i(x_i) = \sum_{r=1}^{\infty} \alpha_r^i \varphi_r(x_i) \tag{1.5}$$

$$f_{ij}(x_i, x_j) = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j) \tag{1.6}$$

$$\cdots$$

Here $\varphi_r(x_i),\ \ \varphi_{pq}(x_i, x_j)$ are sets of one- and two-dimensional basis functions (Legendre polynomials) and α_r^i and β_{pq}^{ij} are coefficients of decomposition which can be determined using orthogonality of the basis function:

$$\alpha_r^i = \int_0^1 f_i(x_i)\varphi_r(x_i)dx_i, \quad r = 1, \ldots, k \tag{1.7}$$

$$\beta_{pq}^{ij} = \int_0^1 \int_0^1 f_i(x_i)\varphi_p(x_i)\varphi_q(x_j)dx_i dx_j$$
$$p = 1, \ldots, l, \quad q = 1, \ldots, l' \tag{1.8}$$

In practice the summation in Eqs. 5 and 6 is limited to some maximum orders k, l, l':

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i \varphi_r(x_i) \tag{1.9}$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j) \tag{1.10}$$

The first few Legendre polynomials are:

$$\begin{aligned}
\varphi_1(x) &= \sqrt{3}(2x-1) \\
\varphi_2(x) &= 6\,\sqrt{5}\left(x^2 - x + \frac{1}{6}\right) \\
\varphi_3(x) &= 20\,\sqrt{7}\left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}\right) \\
&\dots
\end{aligned} \tag{1.11}$$

Coefficients of the decomposition and are related to the first-, second-, and third-order sensitivity indices by [6, 8]:

$$S_i \approx \frac{\sum_{r=1}^{k} \left(\alpha_r^i\right)^2}{V}$$

$$S_{ij} \approx \frac{\sum_{p=1}^{l} \sum_{q=1}^{l'} \left(\beta_{pq}^{ij}\right)^2}{V}$$

$$S_{ijk} \approx \frac{\sum_{p=1}^{m} \sum_{q=1}^{m'} \sum_{r=1}^{m''} \left(\gamma_{pqr}^{ijk}\right)^2}{V}$$

where V, the total variance, is given by Eq. 2.28. The optimal values of α_r^i, β_{pq}^{ij}, and γ_{pqr}^{ijk}, determined by a least squares minimization criteria, are given by [5].

Typically, the higher the number of component functions in the truncated expansion, the higher will be the number of sampled points N needed to evaluate the polynomial coefficients with sufficient accuracy. Li and Rabitz proposed the use of ratio control variate methods to improve the accuracy of Eq. 2.35 in estimating α_r^i, β_{pq}^{ij}, and γ_{pqr}^{ijk} [14]. Feil et al. used quasi-random points instead of pseudorandom numbers for improving the accuracy of Eq. 2.35; they proposed determining an optimal number of points N_{opt}, such that the variance in α_r^i as a function of N in two consecutive simulations was within some tolerance [3].

Integers κ, l, l', m, m', and m'' in Eq. 2.33 are the polynomial orders, and an important problem is the choice of the optimal values for these integers. Ziehn and Tomlin proposed using a least squares minimization technique in determining the optimal polynomial order between [0, 3] for each component function [20]. Feil et al. proposed the use of the convergence of the sensitivity indices of Eq. 2.34 in defining optimal polynomial orders for each component function [3].

For a model with a high number of input parameters and significant parameter interactions, Ziehn and Tomlin recommend first applying a screening method such as the Morris method to reduce the dimensionality of the problem and thus improve the accuracy of the estimation of high order component functions for smaller sample sizes [20].

The error of the model approximation can be measured, similarly to Eq. 2.30, by the scaled distance:

$$\delta(f,\tilde{f}) = \frac{1}{V}\int \left[f(x) - \tilde{f}(x)\right]^2 dx \tag{1.12}$$

where $f(x)$ is the original function and $\tilde{f}(x)$ the approximation. This scaling serves as a benchmark to distinguish between good and bad approximation; for if the mean f_0 is used as the approximant, that is, if $\tilde{f}(x) = f_0$, then $\delta = 1$. Thus a good approximation is one with $\delta \ll 1$ [19, 21].

Metamodels play an important role in the analysis of complex systems. They serve as an effective way of mapping input–output relationships and of assessing the impact of the inputs on outputs. Metamodels can also be applied to solve various types of optimization problems that involve computation-intensive functions.

One of the very important and promising developments of model analysis is the replacement of complex models and models which need to be run repeatedly online with equivalent "operational metamodels."

There are a number of techniques for approximating complex, multivariate functions. *Response surface methodologies* usually approximate the complex multivariate function by low order polynomials, such as

$$\tilde{Y} = \beta_o + \sum_{i=1}^{k}\beta_i X_i + \sum_{i=1}^{k}\beta_{ii}X_i^2 + \sum_{i=1}^{k-1}\sum_{j>i}^{k}\beta_{ij}X_iX_j \tag{1.13}$$

where $\tilde{Y}$ is an approximation to Eq. 2.5 with the parameters β_o, β_i, ... being determined by some form of least square regression [18]. *Kriging* is an interpolative approximation method based on the weighted sum of the sampled data. It is a combination of a polynomial model plus departures which are realization of a random function [2]. *Neural networks* have been used to approximate a multivariate function as a nonlinear transformation of multiple linear regression models [18]. Although these techniques are useful for particular applications, they fall short in certain areas. *Response surface methodologies* are not accurate enough to approximate complex nonlinear multimodal profiles as they are based on simple quadratic models. *Kriging* models are difficult to obtain or even use [17]. Training of *neural networks* usually takes a lot of computing time [18]. A promising metamodeling tool for approximating complex, multivariate functions is the *HDMR*.

One major problem associated with traditionally used parameterized polynomial expansions and interpolative look-up tables is that the sampling efforts grow exponentially with respect to the number of input variables. For many practical problems only low order correlations of the input variables are important. By exploiting this feature, one can dramatically reduce the computational time for modeling such systems. An efficient set of techniques called HDMR was developed by Rabitz and coauthors [6, 14]. A practical form of HDMR, Random Sampling-HDMR (RS-HDMR), has recently become a popular tool for building metamodels [20]. Unlike other input–output mapping methods, HDMR renders the original exponential difficulty to a problem of only polynomial complexity and it can also be used to construct a computational model directly from data.

Variance-based methods are one of the most efficient and popular global SA techniques. However, these methods generally require a large number of function evaluations to achieve reasonable convergence and can become impractical for large engineering problems. RS-HDMR can also be used for GSA. This approach to GSA is considerably cheaper than the traditional variance-based methods in terms of computational time as the number of required function evaluations does not depend on the problem dimensionality. However, it can only provide estimates of the main effects and low order interactions.

ANOVA: High Dimensional Model Representation

Recall, that an integrable function $f(\boldsymbol{x})$ defined in the unit hypercube H^n can be expanded in the following form:

$$f(\boldsymbol{x}) = f_o + \sum_{i=1}^{n} \sum_{i_1 < \ldots < i_s}^{s} f_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s})$$

This expansion is unique if

$$\int_0^1 f_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) dx_{i_k} = 0, \quad 1 \leq k \leq s \qquad (1.14)$$

in which case it is known as the ANOVA-HDMR decomposition. It follows from condition that the ANOVA-HDMR decomposition is orthogonal.

Rabitz argued (in [6]) that for many practical problems only the low order terms in the ANOVA-HDMR decomposition are important and $f(\boldsymbol{x})$ can be approximated by

$$\hat{f}(\boldsymbol{x}) = f_o + \sum_{i=1}^{d} \sum_{i_1 < \ldots < i_s}^{s} f_{i_1 \ldots i_s}(x_{i_1}, \ldots, X_{i_s})$$

Here d is a truncation order, which for many practical problems can be equal to 2 or 3.

Approximation of ANOVA-HDMR Component Functions

The RS-HDMR method proposed in Li and Rabitz and Li et al. aims to reduce the sampling effort by approximating the component functions by expansions in terms of a suitable set of functions, such as orthonormal polynomials [6, 14].

Consider piecewise smooth and continuous component functions. Using a complete basis set of orthonormal polynomials they can be expressed via the expansion:

$$f_i(x_i) = \sum_{r=1}^{\infty} \alpha_r^i \varphi_r(x_i)$$

$$f_{ij}(x_i, x_j) = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j)$$

Here $\varphi_r(x_i)$, $\varphi_{pq}(x_i, x_j)$ are sets of one- and two-dimensional basis functions and α_r^i, β_{pq}^{ij} are coefficients of decomposition which can be determined using orthogonality of the basis functions:

$$\alpha_r^i = \int_0^1 f_i(x_i)\varphi_r(x_i)dx_i, \quad r = 1, \ldots, k \tag{1.15}$$

$$\beta_{pq}^{ij} = \int_0^1 \int_0^1 f_i(x_i)\varphi_p(x_i)\varphi_q(x_j)dx_i dx_j \tag{1.16}$$

$$p = 1, \ldots, l, \quad q = 1, \ldots, l'$$

In practice the summation in Eqs. 1.15 and 1.16 is limited to some maximum orders k, l, l':

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i \varphi_r(x_i)$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j)$$

The question of how to find maximum orders is discussed in the following sections. Shifted Legendre polynomials are orthogonal in the interval [0,1] with unit weight and they are typically used for uniformly distributed inputs. The higher dimensional polynomials can be expressed as the product of one-dimensional ones.

The first few Legendre polynomials are:

$$\varphi_1(x) = \sqrt{3}(2x - 1)$$

$$\varphi_2(x) = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right) \tag{1.17}$$

$$\varphi_3(x) = 20\sqrt{7}\left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}\right)$$

Decomposition coefficients are usually evaluated using Monte Carlo integration, which can be inaccurate especially at small number of sampled points N. It was shown that in the approximation of higher order polynomial expansions with a small number of sampled points N, oscillations of the component functions may occur around the exact values [6]. The integration error can be reduced either by increasing the sample size N or by applying the variance reduction techniques proposed in Li and Rabitz [14]. Feil et al. suggested using QMC sampling to reduce oscillations and the integration error [3].

Evaluation of Global Sensitivity Indices Based on RS-HDMR

For a continuous function with piecewise derivatives the following relationship exists between the square of the function and the coefficients of its decomposition c_r with respect to a complete set of orthogonal polynomials (Parseval's theorem):

$$\int_0^1 f(x)^2 dx = \sum_{r=1}^{\infty} (c_r)^2$$

Application of Parseval's theorem to the component functions of ANOVA-HDMR and definitions of SI yield the following formulas for SI:

$$S_i = \frac{\sum_{r=1}^{\infty} (a_r^i)^2}{D}$$

$$S_{ij} = \frac{\sum_{p=1}^{\infty} (\beta_{pq}^{ij})^2}{D}$$

where D is the total variance.

For practical purposes, function decompositions truncated at some maximum order of polynomials are used:

$$S_i \approx \frac{\sum_{r=1}^{k} (\alpha_r^i)^2}{V}$$

$$S_{ij} \approx \frac{\sum_{p=1}^{l} \sum_{q=1}^{l'} \left(\beta_{pq}^{ij}\right)^2}{V}$$

$$S_{ijk} \approx \frac{\sum_{p=1}^{m} \sum_{q=1}^{m'} \sum_{r=1}^{m''} \left(\gamma_{pqr}^{ijk}\right)^2}{V}$$

The total number of function evaluations required for calculation of a full set of main effect and total SI using the general Sobol' formulas is $N_F = N(n+2)$ [10]. To compute SI using RS-HDMR or QRS-HDMR only $N_F = N$ function valuations is required, which is $n + 2$ times less than for the original Sobol method for the same number of sampled points. However, in practice

RS-HDMR or QRS-HDMR can only provide sets of first- and second-order (up to third) SI.

How to Choose the Maximum Order of Polynomials

An important problem is how to choose an optimal order of the orthogonal polynomials. In majority of published works by Rabitz and coauthors the fixed order polynomials (up to the second or third order) were used. However, in some cases polynomials up to the tenth order were used, although no explanation for the choice of such a high order polynomials were given.

This problem of optimal maximum order polynomials was considered by Ziehn and Tomlin [20]. They proposed to use an optimization method to choose the best polynomial order for each component function. They also suggested excluding any component function from the HDMR expansion which does not contribute to the HDMR expansion. The overheads for using an optimization method can be considerable. We suggest a different approach to define optimal polynomial orders based on the estimated convergence of SI calculated by RS(QRS)-HDMR.

Typically the values of decomposition coefficients, α_r^i, β_{pq}^{ij}, etc., rapidly decrease with increasing the order of r and (p, q). As a result the truncation error is dominated by the first few truncation coefficients.

Another important issue is how to define a sufficient number of sampling points in MC or QMC integration of the polynomial coefficients, α_r^i, β_{pq}^{ij}. Although in a limit

$$\lim_{\substack{k\to\infty \\ N\to\infty}} \sum_{r=1}^{k} \left(\hat{\alpha}_r^i(N)\right)^2 = S_i$$

(the same asymptotic rule apply for other coefficients) but practically the accuracy of coefficients approximation depends on the number of sampled points N: $\hat{\alpha}_r^i = \hat{\alpha}_r^i(N)$. Typically, the higher the order of the component function the higher the number of sampled points required to evaluate the polynomial coefficients with sufficient accuracy [6].

To determine an optimal number of points N_{opt} it is sufficient to examine the variance of α_r^i, $r = 1, 2$ as a function of N. N is increased sequentially and N at which a required tolerance of the variance is reached is taken as N_{opt}.

After a sufficient number of function evaluations N_{opt} is made, the convergence of the estimated sensitivity indices with respect to the polynomial orders is monitored. For the first-order component functions the contribution of the subsequent α_{k+1}^i coefficient is analyzed by monitoring its relative or absolute (in the case of small values of S_i) contribution:

$$\frac{\left(\alpha_{k+1}^i\right)^2}{\sum_{r=1}^{k+1}\left(\alpha_r^i\right)^2} < \epsilon_1 \text{ if } \sum_{r=1}^{k+1}\left(\alpha_r^i\right)^2 > 10^{-4}$$

$$\left(\alpha_{k+1}^i\right)^2 < \epsilon_1, \text{ otherwise}$$

For the second-order component functions the procedure is more complex because it requires monitoring convergence in a two-dimensional space of p and q polynomial orders.

References

1. Saltelli A, Ratto M et al (2008) A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. Wiley, West Sussex
2. Sathyanarayanamurthy H, Chinnam RB (2009) Metamodels for variable importance decomposition with applications to probabilistic engineering design. Comput Ind Eng 57:996–1007
3. Kucherenko S, Fernandez MR, Pantelide C, Shah N (2009) Monte Carlo Evaluation of derivative-based global sensitivity measures. Reliab Eng Syst Saf 94:1135–1148
4. Rabitz H, Alis OF et al (1999) Efficient input–output model representations. Comput Phys Commun 117:11–20
5. Li G, Wang S, Rabitz H (2002) Practical approaches to construct RS-HDMR component functions. J Phys Chem 106:8721–8733
6. Li G, Wang S et al (2002) Global uncertainty assessment by high dimensional model representation (HDMR). Chem Eng Sci 57:4445–4460
7. Li ZQ, Xiao YG, Li ZMS (2006) Modeling of multi-junction solar cells by Crosslight APSYS. http://lib.semi.ac.cn:8080/tsh/dzzy/wsqk/SPIE/vol6339/633909.pdf. Accessed 18 June 2010
8. Feil B, Kucherenko S, Shah N (2009) Comparison of Monte Carlo and Quasi-Monte Carlo sampling methods in High Dimensional Model Representation. In: Proc First International Symposium Adv System Simulation, SIMUL 2009, Porto, Portugal, 20–25 September 2009
9. Zuniga MM, Kucherenko S, Shah N (2013) Metamodelling with independent and dependent inputs. Comput Phys Commun 184(6):1570–1580
10. Sobol' IM, Tarantola S et al (2007) Estimating the approximate error when fixing unessential factors in global sensitivity analysis. Reliab Eng Syst Saf 92:957–960
11. Sobol IM, Kucherenko S (2009) Derivative based global sensitivity measures and their link with global sensitivity indices. Math Comput Simul 79:3009–3017
12. Sobol IM, Kucherenko S (2010) A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. Comput Phys Commun 181:1212–1217
13. Kucherenko S, Zaccheus O, Munoz ZM (2012) SobolHDMR User manual. Imperial College London, London
14. Li G, Rabitz H (2006) Ratio control variate method for efficiently determining high-dimensional model representations. J Comput Chem 27:1112–1118
15. Kucherenko S, Feil B, Shah N, Mauntz W (2011) The identification of model effective dimensions using global sensitivity analysis. Reliab Eng Syst Saf 96:440–449
16. Sobol' IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Comput Simul 55:271–280
17. Wang GG, Shan S (2006) Review of metamodeling techniques in support of engineering design optimization. http://74.125.155.132/scholar?q=cache:_NCLv92moGkJ:scholar.google.com/ &hl=en&as_sdt=2000. Accessed 14 Jan 2010
18. Simpson TW, Peplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17:129–150
19. Wang SW, Georgopoulos PG, Li G, Rabitz H (2003) RS-HDMR with nonuniformly distributed variables: application to integrated multimedia/multipathway exposure and dose model for trichloroethylene. J Phys Chem 107:4707–4716
20. Ziehn T, Tomlin AS (2008) Global sensitivity analysis of a 3D street canyon model—part I: the development of high dimensional model representations. Atmos Environ 42:1857–1873
21. Sobol' IM (2003) Theorems and examples on high dimensional model representation. Reliab Eng Syst Saf 79:187–193
22. Homma T, Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. Reliab Eng Syst Safety 52:1–17

Index

Karen M. Polizzi and Cleo Kontoravdi (eds.), *Synthetic Biology*, Methods in Molecular Biology, vol. 1073, DOI 10.1007/978-1-62703-625-2, © Springer Science+Business Media New York 2013