Methods in Molecular Biology 834

Springer Protocols

Qiong Cheng *Editor*

Microbial Metabolic Engineering

Methods and Protocols

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

Microbial Metabolic Engineering

Methods and Protocols

Edited by

Qiong Cheng

DuPont Central Research and Development, Wilmington, DE, USA

Humana Press

Editor
Qiong Cheng
DuPont Central Research and Development
Wilmington, DE, USA
qiong.cheng@usa.dupont.com

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-61779-482-7 e-ISBN 978-1-61779-483-4
DOI 10.1007/978-1-61779-483-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011943294

Printed on acid-free paper

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

Preface

Classic strain improvement is a long process that relies on screening of naturally occurring or chemically induced mutations for improved strain performance. Rapid technology advancement has accelerated strain improvement by opening the field of metabolic engineering. Metabolic engineering is the practice of genetically optimizing metabolic and regulatory networks within cells to increase production and/or recovery of certain substance from cells. The approaches may be as simple as manipulation of a single gene, and it could also be as complex as reconstruction of the cell's networks.

The objective of *Microbial Metabolic Engineering* is to provide an overview of strategies and techniques of metabolic engineering within the scope of microbial applications mainly focused on bacteria and yeasts. The first part of the book describes methods to engineer genes, pathways, or the whole genome in the production host. The complementary approaches of rational design and random screening to tap into natural diversity and engineered diversity are both illustrated. The second part of the book describes the use of modern biotechnology tools in microbial metabolic engineering. It includes use of genetic tools, omics tools, FACS analysis, and flux balance analysis for identification of targets of genes/pathways for metabolic engineering. It also includes use of microfermentors and fermentation control techniques for rapid evaluation of engineered strains. The third part of the book describes several successful examples of metabolic engineering for real-world applications such as whole-cell biosensors and acetate control in large-scale fermentation. It also addresses several challenges in commercial production such as using biomass hydrolysates as feedstock and minimizing phage contamination.

Microbial Metabolic Engineering is intended both for researchers (molecular biologists, biochemists, microbiologists, physiologists, and bioinformaticians) in academia who are interested in understanding the gene functions and cellular network in microbes, and for those in industry who are interested in developing commercial products from microbial fermentations. It also provides fermentation engineers and process engineers an illustration of what could be achieved by metabolic engineering of the microbe that may have a significant impact on fermentation and downstream processing. The book provides step-by-step instructions, and could also be used as a text book in teaching undergraduate labs. In fact, some strains described in the book have been used in undergraduate labs in Cambridge University, UK. The chapters were written by renowned investigators in the field who practice the method on a regular basis. Procedures are described with enough details so that users can carry out the method without further reference to other sources. Advice and suggestions from the experienced investigators and troubleshooting of possible roadblocks are captured in the Notes section at the end of the chapters. In all, this book is intended to provide an overview of the key topics in microbial metabolic engineering and to be used as the guidebook for researchers who practice metabolic engineering for microbial applications.

Wilmington, DE, USA ***Qiong Cheng***

Preamble

Metabolic engineering is, at its simplest, the purposeful alteration of metabolism to produce molecules of interest. The science and technology in this field continue to advance at a remarkable pace, and improvements in DNA sequencing, DNA synthesis, enzyme engineering, analytical chemistry, and protocols for understanding and utilizing large and increasingly accessible databases of enzymes and metabolic pathways are all contributing to making the practice of metabolic engineering increasingly robust. Researchers have been emboldened by their new, hard-won capabilities, and are tackling ever more difficult problems, developing microbial catalysts for everything from high value pharmaceutical products through to nutritional compounds, high volume chemicals, all the way to very high volume fuel molecules. A number of commercially successful outcomes of detailed metabolic engineering programs have started to appear, and industrial companies old and new are looking to metabolic engineering as a tool to help deliver an increasingly broad range of molecules from renewable resources.

Putting it all together requires working across multiple scales with a broad range of tools and disciplines. In practical terms, creating microbes that produce desired products at commercially relevant rates, titers, and yields—usually under conditions that put severe stress on the producing organism—can challenge the most sophisticated of experimental approaches and often requires working at the very limits of our understanding of cellular and metabolic processes. Molecular tools and classical and accelerated strain selection and improvement protocols are all critically important elements in successful programs. With engineering scale-up, many previously "solved" problems can appear in a new light, and require iterative engineering/metabolic engineering revisions. Doing all this in timeframes consistent with commercial practice adds another significant challenge in designing optimum experimental protocols. As a result, metabolic engineers require access to numerous tools, and often need to pursue multiple, simultaneous approaches to a single end—iteratively combining deterministic, theory-based approaches with empirical, evolutionary-based approaches to optimize productivity.

While this can be daunting, the tools and approaches are being streamlined and simplified, and are becoming more accessible. It was not long ago, for example, that the use of metabolic flux analysis and related mathematical analyses to guide metabolic engineering approaches was exclusively reserved to a small handful of expert practitioners, and underlying data sets were difficult to access and use. As a sign of the times, I was amazed to find as I was writing this preamble that my son—a third year chemical engineering student—had a holiday problem set for one of his courses that required him to perform a constraint-based analysis of a genome-scale metabolic network. To proceed, he used databases and software tools that he downloaded from the internet—and set about solving his problem of finding a triple mutant in *E coli* that would maximize the anaerobic production of lactic acid from glucose. It was quite remarkable to see the next generation pursuing this level of sophistication and inquiry in so facile a manner. As more and more scientists and engineers are able to access ever more powerful methodologies, the field of metabolic engineering will become increasingly able to deliver timely and relevant outcomes.

This volume brings together contributions from a wide-ranging group of expert academic, institute, and industrial practitioners. The range of methodological approaches and tools that are exemplified give a good sense of the creativity and multidisciplinarity that are required to proceed from an initial idea to the large-scale production of a commercial product. Using the types of approaches and protocols outlined in this volume, we can look forward to being able to produce—with increasing speed and confidence—an increasingly large number of products from renewable resources through the use of microbial metabolic engineering.

London, UK ***John Pierce***

Contents

Contributors

David L. Alexander • *Department of Microbiology and Environmental Toxicology, University of California Santa Cruz, Santa Cruz, CA, USA*

Jennifer M. Allen • *Department of Microbiology and Environmental Toxicology, University of California Santa Cruz, Santa Cruz, CA, USA*

Hal S. Alper • *Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA*

Frances H. Arnold • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA*

Sabine Bastian • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA*

Steven D. Brown • *Biosciences Division and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Manel Camps • *Department of Microbiology and Environmental Toxicology, University of California Santa Cruz, Santa Cruz, CA, USA*

Qi Chen • *DuPont Central Research and Development, Wilmington, DE, USA*

Qiong Cheng • *DuPont Central Research and Development, Wilmington, DE, USA*

Kevin M. Croker • *DuPont Central Research and Development, Wilmington, DE, USA*

Nathan C. Crook • *Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA*

Aitor de las Heras • *Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, Madrid, Spain*

Tony Ho • *Life Technologies Corporation, Carlsbad, CA, USA*

Nele Ilmberger • *Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Hamburg, Germany*

Kaizhi Jia • *Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*

Julia Jüergensen • *Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Hamburg, Germany*

Federico Katzen • *Life Technologies Corporation, Carlsbad, CA, USA*

Dawn M. Klingeman • *Biosciences Division and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Alexander D. Kopatsis • *DuPont Central Research and Development, Wilmington, DE, USA*

Wieslaw Kudlicki • *Life Technologies Corporation, Carlsbad, CA, USA*

Ke Li • *Life Technologies Corporation, Carlsbad, CA, USA*

Yin Li • *Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*

Xiquan Liang • *Life Technologies Corporation, Carlsbad, CA, USA*

Marcin Los • *Phage Consultants, Inc, Gdansk, Poland*

Víctor de Lorenzo • *Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, Madrid, Spain*

Shaoming Mao • *Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*
Isabelle Meynial-Salles • *Université de Toulouse, Toulouse, France*
Samuel Million-Weaver • *Department of Microbiology and Environmental Toxicology, University of California Santa Cruz, Santa Cruz, CA, USA*
Kathleen A. Curran • *Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA*
Nikhil U. Nair • *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
Lansha Peng • *Life Technologies Corporation, Carlsbad, CA, USA*
Todd Peterson • *Life Technologies Corporation, Carlsbad, CA, USA*
John Pierce • BP, *London, UK*
Jason Potter • *Life Technologies Corporation, Carlsbad, CA, USA*
Peter J. Punt • *TNO Microbiology and Systems Biology, Zeist, The Netherlands*
Andrew R. Reeves • *Coskata, Inc, Warrenville, IL, USA*
Kristin Ruebling-Jass • *DuPont Central Research and Development, Wilmington, DE, USA*
Josh Shirley • *Life Technologies Corporation, Carlsbad, CA, USA*
Ronald Slomp • *TNO Microbiology and Systems Biology, Zeist, The Netherlands*
Philippe Soucaille • *Université de Toulouse, Toulouse, France; Metabolic Explorer, Biopôle Clermont-Limagne, Saint-Beauzire, France*
Wolfgang R. Streit • *Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Hamburg, Germany*
Wonchul Suh • *DuPont Central Research and Development, Wilmington, DE, USA*
Jun Sun • *DuPont Central Research and Development, Wilmington, DE, USA*
Yong Tao • *Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*
Billyana Tsvetanova • *Life Technologies Corporation, Carlsbad, CA, USA*
Tina K. Van Dyk • *DuPont Central Research and Development, Wilmington, DE, USA*
Johan van Groenestijn • *TNO Microbiology and Systems Biology, Zeist, The Netherlands*
J. Mark Weber • *Fermalogic, Inc, Chicago, IL, USA*
Dongming Xie • *DuPont Central Research and Development, Wilmington, DE, USA*
Liewei Xu • *Life Technologies Corporation, Carlsbad, CA, USA*
Jian-Ping Yang • *Life Technologies Corporation, Carlsbad, CA, USA*
Shihui Yang • *Biosciences Division and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA; National BioEnergy Center, National Renewable Energy Laboratory, Golden, CO, USA*
Ying Zha • *TNO Microbiology and Systems Biology, Zeist, The Netherlands*
Jianzhong Zhang • *DuPont Central Research and Development, Wilmington, DE, USA*
Yanping Zhang • *Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*
Huimin Zhao • *Departments of Chemistry, Biochemistry, and Bioengineering, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

Chapter 1

Screening for Cellulases with Industrial Value and Their Use in Biomass Conversion

Julia Jüergensen, Nele Ilmberger, and Wolfgang R. Streit

Abstract

Cellulose is an easily renewable and highly occurring resource. To take advantage of this great potential, there is a constant need of new cellulose degrading enzymes. In industrial applications enzymes have to function under extreme conditions like high temperature, very acidic or basic pH and different solvents.

Cellulases have a huge area of application, for example the textile and food industry as well as the generation of bioethanol as an alternative energy source. They have the ability to yield a great energetic potential, but there is still a lack of economical technologies to conquer the stability of the cellulose structure. Via metagenomic research and well-directed screening, it is possible to detect new cellulases, which are active under tough industrial conditions.

Key words: Cellulase, Metagenome, Renewable energy, Biotechnology, Screening

1. Introduction

Cellulose is not only the most occurring organic compound in nature but also the most abundant renewable energy source, next to chitin. Plants usually contain 35–50% (dry weight) cellulose. Lignin, hemicelluloses, and other back-up substances provide support against cellulolytic activities (1).

It can be used as a valuable source for bioethanol and other products. Therefore cellulose (Fig. 1), consisting of very stable β-1,4-linked glucose subunits, must be hydrolysed into fermentable sugar. Hundreds to ten thousand cellobiose molecules (dimer of β-D-glucose) form an unbranched chain. A protofibril contains several of these chains and in turn a microfibril contains many protofibrils. These crystalline filaments have the same tension stability as steel and are insoluble in water and diluted acids or bases (2). Breakdown of cellulose can be performed by chemical treatment or enzymatic hydrolysis. Chemical breakdown has the disadvantage of

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_1,

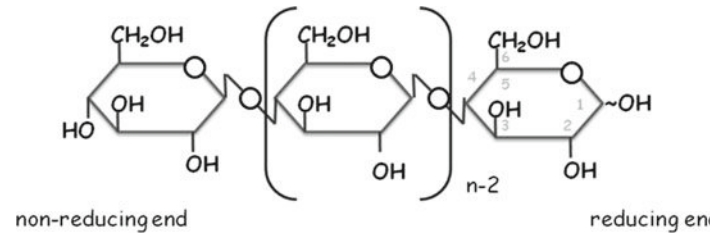


Fig. 1. Cellulose structure D-glucose; linked to large polymers via the β-1,4 glycosidic linkage.

cost-intensive pollutants. For large-scale enzymatic hydrolysis, some problems occur like the water insolubility and the poor accessibility for the enzymes.

Cellulases are glykosylhydrolases that degrade oligo- or polysaccharides (3). They are distinguished from other glykosylhydrolases by their ability to hydrolyze β-1,4-glycosidic bonds between glucosyl residues. Three major types of enzymatic activity are necessary for complete degradation of cellulose: endoglucanases (1,4-β-D-glucan-4-glucanohydrolases; EC 3.2.1.4) cut randomly within amorphous areas and generate oligosaccharides of different lengths and therefore new chain endings. Exoglucanases including cellodextrinases (1,4-β-D-glucan glucanohydrolases; EC 3.2.1.74) divide up cellobiase molecules of the new chain endings, and they are not limited to the amorphous regions but can also cut crystalline cellulose. The cellobiohydrolases (1,4-β-D-glucan cellobiohydrolases; EC 3.2.1.91) and β-glucosidases (β-glucoside glucohydrolases; EC 3.2.1.21) dissolve the β-glycosidic bonds between the glucose monomers (4, 5). The most recent nomenclature describes more than 100 families of glycosylhydrolases, which are organized into 14 clans as listed at the CAZy server (http://afmb.cnrs-mrs.fr/CAZY/) and that are involved in the hydrolysis of cellulose molecules.

The enzymatic breakage of the β-1,4-glycosidic bonds in cellulose proceeds through an acid hydrolysis mechanism using a proton donor and nucleophile or base. The hydrolysis can either result in the inversion or retention (double replacement mechanism) of the anomeric configuration of carbon-1 at the reducing end (1, 6, 7).

Cellulases have many industrial applications; next to the generation of bioethanol, a realistic long-term energy source is for example the finishing of textiles (2, 8). These industrial processes require cellulolytic activity under a variety of temperature, pH and ionic conditions, and they are usually carried out by mixtures of cellulases, because of their ability to synergise (2). Investigation of the broad diversity of cellulolytic enzymes involved in the natural degradation of cellulose is necessary for the optimisation of these processes. While there remains much interest in the isolation of cellulases from fungal sources, there has been a recent increase in the isolation of diverse novel cellulases from prokaryotic organisms (7).

Aerobic bacteria and fungi secrete their cellulases (non-complexed) as exoenzymes. In most cases free cellulases contain cellulose binding domains (CBD) which bring the catalytic domain in direct contact to the substrate (9). CBDs play a role in binding the cellulase to insoluble cellulose (10, 11). Cellulases lacking a CBD show reduced activities against crystalline cellulose while retaining the capacity to depolymerise soluble cellulosic substrates (10, 12–14).

Some anaerobe microorganisms arrange their cellulases in fist-like multi-enzyme complexes, known as cellulosome, which is linked to the cell surface (4, 15). A cellulosome comprises different cellulases organized on a non-catalytic scaffolding protein which mediates the attachment to the substrate (15). The cellulosomes are bound strongly to the cell surface, but flexible enough to bind crystalline cellulose as well, thus the diffusion of the originated sugar is prevented (16).

The majority of the till now investigated prokaryotic cellulases have been isolated from cultured microorganisms. Cellulases from specific organisms tend to be active in the pH and temperature conditions corresponding to their environment such as the endoglucanase from an alkalophilic Bacillus species which has a pH range from 7.0 to pH 12.0 (17) and the β-1,4-endoglucanase from the gut bacterium *Cellulomonas pachnodae* which has a pH range between pH 4.8 and pH 6.0 (18). Many of the cellulases with the industrially relevant characteristics are obtained from extremophile microorganisms (8, 19), because of their extended tolerance to temperature, ionic strength and pH.

Cultivation of microbes from extreme or other specific environments is particularly problematic, what results in a large proportion of uncultured bacteria, especially in these habitats. Ninety-nine per cent of all microorganisms cannot be isolated or cultivated.

Metagenomics is a cultivation-independent analysis of the microbial DNA of a specific habitat and involves direct isolation of DNA from the environment followed by cloning and expression of the metagenome in a heterologous host (20). For this reason, metagenomics has recently become a very powerful tool to search for novel enzymes that are useful for biotechnological applications, and there is already a wide range of new biocatalysts found by this technique (21, 22). The technology has been summarized a few times (21–23) and since its first publication (24) a remarkable number of reports were published providing new enzymes with a high potential for industrial applications (25–28). Because cellulose is such a valuable biopolymer, there have been a significant number of publications about the isolation of metagenome-derived cellulases. One of the earliest articles presented the detection of cellulases from a thermophilic, anaerobic digester fuelled by lignocelluloses (29), and a recent study detected seven cellulases with novel features (30). Metagenomic screening of extreme environments, soda-lakes in Egypt (Africa), detected more than a dozen

cellulases, some of which displayed habitat-related halotolerant characteristics (31, 32). Anyhow, the highly genetically diverse of non-extreme environments contain a wide range of cellulases which are highly stable and suitable for industrial applications (33, 34). Further examples of successful isolation of metagenome-derived cellulases have been described (35, 36). Sequences-based approaches have led to the identification of several putative cellulose clones (37); the functionality of these enzymes has to be verified.

Here, we offer some easy-to-follow protocols for constructing a metagenome library and for screening microbial cellulases.

2. Materials

2.1. Sampling

1. Sterile vessel, for example falcon tube.
2. Gloves.

2.2. Isolation of Metagenomic DNA from Elephant Faeces

1. QIAamp® DNA Stool Mini Kit (Qiagen GmbH, Hilden, Deutschland).
2. Fresh elephant faeces or frozen in glycerine.

2.3. Construction of a Metagenomic Large Insert Library

1. CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, WI).
2. TE buffer: 10 mM Tris pH 8.0, 1 mM EDTA.
3. LMP (low melting point) agarose.
4. LB medium containing 10 mM $MgSO_4$ and 0.2% Maltose for growth of EPI300-T1R host cells.
5. Phage-dilution buffer: 10 mM Tris–HCl pH 8.3, 100 mM NaCl, 10 mM $MgCl_2$.
6. LB plates supplemented with 12.5 μg/mL chloramphenicol.
7. Microtiter plates (96 wells) containing 200 μL LB medium supplemented with 12.5 μg/mL chloramphenicol.
8. 86% glycerine, 100 μL per each well.

2.4. Screening Metagenomic Libraries for Cellulose Degrading Enymatic Activity

1. CMC agar plates: 2 g carboxymethyl cellulose (CMC), 5 g NaCl, 10 g peptone, 5 g yeast extract, 15 g agar.
2. 48er steel stamps.
3. Congo red solution: 0.2% congo red.
4. 1 M NaCl solution.

2.5. DNSA-Assay

1. Sonicator.
2. 50 mM Tris–HCl pH 8.0.

3. LB+CMC (1 L): 10 g tryptone, 5 g yeast extract, 5 g NaCl, 2 g carboxymethylcellulose (CMC).
4. DNSA-Reagent (1 L): 10 g 3,5-dinitrosalicylic acid, 2 mL phenol, 0.5 g Na_2SO_3, 200 g K-Na-tartrat, 10 g NaOH. Store at 4°C (protected from light).
5. McIlvaine buffer: 0.2 M Na_2HPO_4 (A), 0.1 M citric acid (B). pH 6.5 was adjusted by the addition of (B) to (A) at 65°C.

2.6. Analysis of Cellulose Breakdown Products by HPLC

1. SepPack cartridge 18 (Waters, Milford, Mass.).
2. HPX-42A carbohydrate column (300×7.8 mm; BioRad, Munich, Germany).
3. Differential refractometer.

2.7. Analysis of Cellulase Reaction Products by Thin-Layer Chromatography

1. Used substrates might be carboxymethyl cellulose (1%, Sigma, Heidelberg, Germany), lichenan (1%, from *Cetraria islandica*, Sigma, Heidelberg, Germany), and cellooligosaccharides (1%, Sigma, Heidelberg, Germany).
2. Cellulase extract in 50 mM K_2HPO_4.
3. Silica 60 TLC plate (Merck KGaA, Darmstadt, Germany).
4. 1-propanol, nitromethane, H_2O (5:3:2, vol/vol/vol).
5. Ethanol/concentrated sulphuric acid (9:1, vol/vol). Prepare fresh.
6. Ethylacetate, acetic acid, H_2O (2:1:1, vol/vol/vol).
7. Phosphoric acid.
8. Stock solution: 1 g diphenylamine, 1 mL anilin, 100 mL acetone.
9. 1-propanol, ethylacetate, H_2O (6:1:3, vol/vol/vol).

2.8. 16S/18S rDNA Phylogenetic Analysis

1. Reaction tubes, pipettes, thermocycler.
2. Archaea 20 F: TTC CGG TTG ATC CGC CRG.
   927R: TCC GGC GTT GAM TCA ATT.
3. Bacteria 616 V: AGA GTT TGA TYM TGG CTC AG (38).
   1492R: CGG YTA CCT TGT TAC GAC (39).
4. Eucarya (18S) E4: AGG AAT TGA CGG AGG GCA C
   E1688: GGA CAT CTA AGG GAT.
   CAC A.
5. 10× Taq reaction buffer, 25 mM $MgCl_2$, 10 mM dNTPs, Taq polymerase (e.g. Fermentas, St. Leon-Rot), sterile water.
6. Gel extraction and Purification Kit.
7. TA cloning Kit (e.g. Invitrogen, Karlsruhe).

8. Sequencing Reaction Kit (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems, Darmstadt) and a capillary sequencer.

2.9. Enrichment

1. Mineral Salt Medium (MSM).
   (a) Solution 1 (1 L, 10×): 70 g $Na_2HPO_4 \times 2H_2O$, 20 g KH_2PO_4.
   (b) Solution 2 (1 L, 10×): 10 g $(NH_4)_2SO_4$, 2 g $MgCl_2 \times 6H_2O$, 1 g $Ca(NO_3)_2 \times 4H_2O$.
   (c) Trace elements (2,000×, 1 L): 5 g EDTA, 3 g Fe(III) $SO_4 \times 7H_2O$, 30 mg $MnCl_2 \times 4H_2O$, 50 mg $CoCl_2 \times 6H_2O$, 20 mg $NiCl_2 \times 2H_2O$, 10 mg $CuCl_2 \times 2H_2O$, 30 mg $Na_2MoO_4 \times 2H_2O$, 50 mg $ZnSO_4 \times 7H_2O$, 20 mg H_3BO_4, pH 4.0.
   (d) Vitamins (1,000×, 100 mL): 1 mg biotin, 10 mg nicotinic acid, 10 mg thiamin-HCl (vit. B1), 1 mg *p*-aminobenzoic acid, 10 mg Ca-D(+) pantothenic acid, 10 mg vit. B6 hydrochloride, 10 mg vit. B12, 10 mg riboflavin, 1 mg folic acid.
2. Sole carbon source (CMC, Avicel, cellulolosic filter paper or plant material).

3. Methods

3.1. Sampling

The elephant faeces were chosen because of the elephant's diet, all other guts of herbivores are also suitable for finding cellulases. Also there is great potential in termites, *Teredo navalis* and bark beetles (all of them consume wood). Compost is suitable for finding thermophilic cellulases because its core temperature is up to 70°C.

The sampling method should always be adapted to the habitat. Some probes are to be treated anaerobically, or must be protected against light.

1. For sampling of elephant faeces, gloves were worn and the faeces (see Note 1) were collected in a few 50 mL sterile falcon tubes for direct use and storage.
2. The falcon tubes were charged with glycerine before storage at –20°C.

3.2. Construction of a Metagenomic Large Insert Library

Fosmids are developed to clone large genomic DNA fragments up to 40 kb, in spite of these large inserts fosmids show high stability (40). The fosmid used for the construction of metagenome libraries (pCC1FOS) contains both the *E. coli*, F-factor single-copy origin of replication and the inducible high-copy *ori*V. They are typically grown at single copy to ensure insert stability and successful

cloning of encoded and expressed toxic protein and unstable DNA sequences; then the fosmids can be induced (41) up to 50 copies per cell immediately before DNA purification. This step increases DNA yield, while maintaining the stability of the plasmid. Kits like the Copy Control™ Fosmid Library Production Kit (e.g. pCC1FOS) are commercially available and combine all advantages to stable insert large DNA fragments into the vector with little loss of time. In the following the corresponding instructions of the manufacturer are presented:

1. *Preparation of DNA*: High molecular weight (meta) genomic DNA is isolated using a genomic DNA isolation kit and diluted in TE buffer or H_2O at a concentration of 0.5 μg/μL. If faeces were selected for establishing a metagenomic library, the use of a stool kit is recommended (see Subheading 2.2). Every habitat/organism is different and the isolation should be adapted to each.
2. *Shearing*: DNA fragments in the range of 20–40 kb, for highly random generation of DNA fragments in contrast to more biased libraries that result from partial restriction endonuclease digestion, are obtained by multiple pipetting the DNA solution using a 1,000-μL pipette tip (see Note 2).
3. *End-Repair of the DNA fragments*: The end-repair reaction described below generates blunt-ended, 5′-phosphorylated DNA fragments and can be scaled up or down depending on the amount of available DNA (see Note 3).

Sterile water	× μL
10× end-repair buffer	8 μL
2.5 mM dNTPs	8 μL
10 mM ATP	8 μL
Up to 20 μg sheared DNA	× μL
End-repair enzyme mix	4 μL
Total reaction volume	80 μL

   The reaction is incubated at room temperature for 45 min; afterwards gel loading buffer is added and incubated at 70°C for 10 min to inactivate the End-Repair-Enzyme Mix.
4. *Size Selection*: To receive DNA fragments ≥25 kb (smaller DNA fragments may result in unwanted chimeric clones), the total End-Repair mix is loaded to a 1%, 20 cm LMP agarose gel at 30–35 V overnight (see Note 4). For the size selection, 100 ng (1 μL) of Fosmid Control DNA into each of the outside lanes and the end-repaired insert DNA in between are loaded to the gel. After electrophoretic separation, the outer lanes containing

the DNA size markers and a small portion of the next lane that contains the random sheared end-repaired genomic DNA are detached, stained with ethidium bromide, and visualized with UV light. This step is for marking the height of the wished DNA size in the gel slice. The gel with the end-repaired DNA should be sliced 2- to 4-mm below the position of the Fosmid Control DNA (see Note 5).

5. *Recovery of DNA*: The LMP agarose is melted by incubating the tube at 70°C for 10–15 min; the tube needs to be quickly transferred to 45°C after melting. GELase Enzyme Preparation (45°C) to 1× final concentration and 1 U of GELase Enzyme Preparation for each 100 μL of melted agarose (see Note 6) are added. Incubation is possible from 45 min to overnight at 45°C. The enzyme is inactivated at 70°C for 10 min. 500 μL aliquots are chilled on ice for 5 min, and then centrifuged for 20 min at maximum speed to pellet any insoluble oligosaccharides. The upper 90–95% of the supernatant, which contain the DNA, is removed to a sterile 1.5-mL tube. The DNA precipitation is done via the well-known ethanol precipitation.
6. *Ligation*: The ligation reaction is mixed in a 10:1 M ratio of CopyControl pCC1FOS vector to insert DNA and incubated for 2 h at RT followed by 4 h to overnight incubation at 16°C. The following reagents are combined in the order listed.

Sterile water	× mL
10× fast-link ligation buffer	1 μL
10 mM ATP	1 μL
CopyControl pCC1FOS vector (0.5 mg/mL)1 μL insert DNA (0.25 μg of 40 kb DNA) x μL	
Fast-link DNA ligase	1 μL
Total reaction volume	10 μL

7. *Packaging reaction*: 10 μL of the ligation reaction are added to one-half of the provided MaxPlax Lambda Packaging extract (25 μL) in a reaction tube being kept on ice; the other half of the packaging extract has to be returned to a –70°C freezer immediately. The packaging reaction is incubated at 30°C. After 2 h, the remaining 25 μL of Lambda Packaging Extract is added and the reaction is incubated for additional 2 h at 30°C. Following the incubation, the Phage-Dilution buffer are added to 1 mL final volume plus 25 μL of chloroform and mixed gently.
8. *Titration of the packaged CopyControl fosmid library*: Prior to transducing the complete packaging reaction, it is recommended

to determine the phage particle titer via dilution series (e.g. CopyControl Fosmid clones). 10 µL of each packaging reaction (diluted an undiluted) is added to 100 µL of the previously prepared EPI300-T1^R host cells (LB containing 10 mM $MgSO_4$) followed by incubation at 37°C for 1 h. The infected cells are spreaded on LB plates supplemented with 12.5 µg/mL chloramphenicol and incubated overnight at 37°C to select for the CopyControl Fosmid clones. Colonies are counted and the phage particles titer is calculated.

9. *Transduction and plating the CopyControl fosmid library*: Based on the titer of phage particles, they have to be diluted in Phage Dilution Buffer to obtain the desired number of clones and clone density on the plate (see Note 7). The transduction into EPI300-T1^R host cells is performed as described above in several parallel reactions. Appropriate aliquots are plated on LB plates supplemented with 12.5 µg/mL chloramphenicol and incubated overnight.

   Fosmid clones obtained are grown in microtiter plates (96 wells) containing 200 µL LB medium supplemented with 12.5 µg/mL chloramphenicol and subsequently stored at –70°C in the presence of 100 µL glycerine (86%) per well.

10. *Induction to higher copy numbers*: The fosmid clones of a library can be induced to high copy numbers for high yields of DNA for sequencing, fingerprinting, or other downstream applications. Induction to higher copy numbers is also recommended for direct function-based screening assays of the clone library, e.g. on plates. The appropriate amount of LB medium supplemented with 12.5 µg/mL and 1‰ of induction solution is inoculated with 10% of an overnight culture and incubated for 5 h (see Note 8).

3.3. Identification of Cellulase-Positive Clones by Screening on Congo Red Plates (see Note 9)

Cellulase-positive clones are usually screened for by using a colorimetric assay on plates containing a cellulosic substrate. The interaction of the direct dye congo red with intact β-D-glucans provides the basis for a rapid and sensitive screening test for cellulolytic bacteria possessing β-D-glucan-hydrolase activities (42).

1. The *E. coli* clones are stamped or streaked from the microtiter plates on LB agar with 0.2% CMC (carboxymethylcellulose) as substrate and incubated overnight at 37°C followed by an incubation of 2–7 days at RT.
2. Colonies are washed off the agar plates with ddH_2O to permit the homogeneous penetration of the staining dye into the medium.
3. Agar plates are stained with congo red solution for 30 min.

4. The solution is poured off and the agar plates are de-stained up to 3 times for 30 min with 1 M NaCl. Cellulase expressing clones are detected by the formation of a yellow halo against a red background.

3.4. Preparation of Crude Cell Extracts of Clones with Cellulolytic Activity

1. For the preparation of crude cell extracts of cellulase positive clones, 200 mL LB cultures containing an appropriate antibiotic are grown at 30°C to an OD of 1.0–1.5.
2. Cells are harvested and resuspended in 50 mM Tris–HCl pH 8.0 prior to cell disruption through sonication (Sonicator UP 200S, Hielscher, Germany) at 50% amplitude and cycle 0.5 for 5 min.
3. After centrifuging at 16,000 × *g* and 4°C for 30 min, the crude cell extract can be stored at 4°C for several days.

3.5. Enzyme Assays for Cellulase Activities

3.5.1. DNSA-Assay (see Note 10)

Cellulase activity is routinely assayed by measuring the amount of reducing sugar released from CMC using 3,5-dinitrosalicylic acid reagent (Fig. 2, also see Subheading 2.5). Hydrolysis of cellulose produced glucose oligomers and monomers. By this process, the number of reducing ends increases. These reducing groups react with 3,5-dinitrosalicylic acid forming brown 3-amino-5-nitrosalicylic acid at 100°C. The amount of 3-amino-5-nitrosalicylic acid formed is equimolar to the number of reducing ends. Therefore, the amount of reducing sugars can be quantified at 546 nm.

Units of enzyme activity (U) are expressed as micromoles of reducing sugar released per minute per milligram protein. Enzyme activities are formulated by regressing absorbance on concentration following the Beer's law. That is the relationship between known concentrations and absorbance is linear except at very low or high concentration of the product, in this case reducing sugar. One unit is equal to 1 µmol of reduced sugar per minute. The enzymatic activity volume was calculated according to the following formula:

$$\mathrm{U/mL} = (\Delta E \,/\, \mathrm{min} \times V) \,/\, (\varepsilon \times d \times v),$$

$\Delta E/\mathrm{min}$ = Extinction

V = Volume of the test reaction mix.

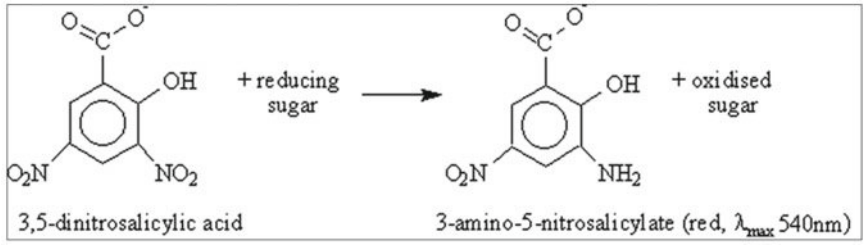


Fig. 2. DNSA assay reaction for the measurement of cellulolytic activity based on the release of reducing sugar ends.

d = Thickness of the cuvette [cm].

ε = Ascendant of straight calibration line

v = Sample volume

The specific enzymatic activity [U/mg protein] is defined as the amount of enzyme that liberates 1 μmol of substrate per minute and is calculated as follows.

$$\text{pecificactivity [U/mg protein]} = \text{Enzymatic activity volume [U/mL]} / \text{protein concentration [mg/mL]}$$

The standard assay mixture contains 2 μg of the enzyme or crude cell extract and 1% CMC in a final volume of 0.5 mL with 150 μL McIllvaine buffer (see Subheading 2.5). This mixture is incubated at an appropriate temperature (usually 37°C) for 15 min.

1. After this incubation, 750 μL DNSA reagent is added. The samples are boiled at 100°C for 15 min.
2. After cooling down on ice the samples are centrifuged at 16,000 × *g* for 2 min to precipitate falling proteins.
3. The samples are transferred to cuvettes and absorbance was measured at 546 nm.
4. The pH range of the enzyme is usually determined by measuring standard assay activity between pH 4 and pH 10.5 using 50 mM of appropriate buffers. Acetate buffer is used for pH 4 to pH 6.0; citrate/phosphate buffer (McIllvaine buffer) is used for pH 6 to pH 7.5; Tris–HCl was used for pH 7.5 to pH 9.0; and *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) was used for pH 9.7 to pH 10.5.
5. For the analysis of the temperature range of the enzyme, activity of the standard assay mixture is assayed at temperatures between 20°C and 95°C.
6. To analyse substrate specificity, CMC can be replaced in the standard assay mixture by lichenan, barley β-glucan, laminarin, oat spelt xylan, or avicel (see Note 11).
   Reaction mix:

Sample	100 μL
CMC in ddH_2O (2%)	250 μL
McIllvaine buffer, pH 6.5	150 μL
DNSA reagent	750 μL

3.5.2. Analysis of Cellulose Breakdown Products by HPLC (see Note 9)

For the investigation of the reaction products of hydrolysis of carbohydrates, HPLC analysis is an appropriate method.

1. First, enzyme preparation and substrate are co-incubated for 2 h at optimal temperature and pH of McIllvaine buffer. As well as for TLC analysis and DNSA assays, different substrates and reaction conditions can be investigated.

2. For stopping the reactions, assays were incubated at 100°C for 10 min.
3. The assay mixtures were centrifuged and proteins of the supernatant were removed with a SepPack cartridge 18. There are a lot of different HPLC columns and elution buffers that can be used for the analysis of carbohydrate hydrolysis reaction products. One possibility is the analysis of the samples with an HPX-42A carbohydrate column.
4. Elution was carried out with H_2O at 85°C; the flow rate was 0.6 mL/min.
5. Detection was performed with a differential refractometer.

3.5.3. Analysis of Cellulase Reaction Products by Thin-Layer Chromatography (TLC) (see Note 9)

To determine whether a cellulase has an endo- or exo- mode of action, TLC analyses are an adequate tool. These analyses can also give a good overview of the substrate range hydrolyzed by the enzyme.

1. As substrates different carbohydrates can be used, e.g. cellooligosaccharides, lichenan, and CMC. These substrates are co-incubated with cellulase extract in 50 mM K_2HPO_4 at adequate pH and temperature conditions.
2. To determine which reaction products occur first, aliquots from different incubation times can be spotted on a silica 60 TLC plate.
3. The cellooligosaccharide reaction products are developed and eparated in 1-propanol, nitromethane, H_2O (5:3:2, vol/vol/vol) for 2 h.
4. After separation, sugars were visualized by spraying the plates with a freshly prepared mixture of ethanol/concentrated sulphuric acid (9:1, vol/vol).
5. The lichenan reaction products are developed in ethylacetate, acetic acid, H_2O (2:1:1, vol/vol/vol) for 3 h. After separation, sugars were visualized by spraying the plates with a freshly prepared mixture of 1 mL phosphoric acid and 10 mL stock solution (1 g diphenylamine, 1 mL anilin, 100 mL acetone).
6. The CMC reaction products are separated and developed in 1-propanol, ethylacetate, H_2O (6:1:3, vol/vol/vol) for 2×3 h, and the sugars visualized using the same mixture as for visualization of lichenan products.

3.6. Retransformation of Putative Positive Clones

To ensure that the observed catalytic activity of clones is not due to contaminations, the isolation and retransformation of the vector and a subsequent activity assay is recommended. Only then clones should be stored and used for further work.

3.7. Enrichment of Highly Cellulolytic Microbial Communities (Optional)

In our experience, the number of clones that encode cellulases in environmental libraries is rather low. Therefore, it is sometimes useful to slightly enrich on a suitable substrate to increase the frequency of cellulolytic organisms and therefore cellulolytic enzymes. Therefore, usually mineral salt medium (see Subheading 2.9) is used.

3.7.1. Enrichment Cultures (see Note 12)

1. The cultures are run under the desired conditions of pH, temperature, oxygen supply, etc.
2. For the enrichment of cellulolytic organisms, cellulosic substrates like carboxymethylcellulose (CMC), crystalline cellulose like avicel, cellulosic filter paper, or plant material like wood or silage can be used as sole carbon source. Once microbial communities are established, they can be used for library construction. Library construction from the enriched consortia is similar to the library construction of non-enriched microbial communities.
3. It is recommended to analyse the microbial community by 16S profiling in order to verify the diversity (see Subheading 3.7.2). Please note that due to the enrichment steps the diversity is probably significantly reduced.

3.7.2. Verification of Biodiversity in Enriched Cultures

The 16S rRNA analysis is suitable for evaluation of enrichment cultures and other samples.

1. Perform standard PCR using small aliquots of the enrichment cultures of other samples using appropriate primers (see in Subheadings 2.8, step 2–4).
2. PCR fragments will then be cloned into pDrive (Qiagen, Hilden, Germany) or any comparable AT-cloning vector.
3. The resulting plasmids will be purified, e.g. with the plasmid isolation kit from Qiagen (Hilden, Germany) and sequenced. Sequences will be corrected according to the sequencing fluorescence curve.
4. Sequences can be analysed by using the ARB database (http://www.arb-home.de) or the ribosomal database (http://rdp.cme.msu.edu).
5. Complementarily, a 454-based deep sequencing of the amplified PCR fragments will result in a much better coverage and estimate on the biodiversity.

4. Notes

1. The elephant faeces were collected fresh at the zoo in Hamburg, when possible the faeces were from the current day.
2. In most cases, it is not necessary to shear the DNA by pipetting. Standard DNA preparation steps already shear the DNA in

sufficient manner. Before proceeding to the end repair, this should be checked by gel analysis.

3. For the construction of a metagenome library, it is commendable to use 20 μg of DNA, not less. For constructing a genome library, less DNA is sufficient. If only 10 μg of DNA is available, it is possible to work with a reaction batch reduced by 50%.
4. In some cases, 0.75% LMP agarose should be enough, depending on the various gel chamber manufacturers. Check the instruction book.
5. Exposing DNA to UV light should be avoided. Even exposure for short periods of time can decrease the efficiency of cloning by two orders of magnitude or more.
6. Even 0.2 U GELase Enzyme Preparation per 100 μL melted agarose worked out.
7. A scale up is also possible if the titer of the phage particles is too low. It worked out with 20 μL per 100 μL of cells.
8. Aeration is very critical in this step, so the vessel should be as big as possible.
9. Screening for cellulase-active clones on congo red indicator plates is easy, only the time period for growth of bacteria and expression of cellulolytic activity might be variable. Washing off bacterial cells is critical, when cellulolytic activity is rather low.
10. The same occurs for the DNSA-assay, where gloves should be worn and when samples are boiled, the lid should be stabilized to protect from spraying phenol (in DNSA solution, see Subheading 2.3). When ionic liquids are added to the assay mixtures, it is necessary to completely agitate IL and aquatic phase, otherwise results are falsified.
11. Inhibition or enhancement of cellulase activity can be determined for a range of different metal chloride salts, solvents, detergents, and EDTA using in general 1 mM concentrations.
12. The most “critical” step in this procedure and for the discovery of a pool of enzymes which is adequate for the detection of one or more cellulases with interesting properties might be the choice of sample respectively the quality of the enrichment culture and of the metagenomic library. We recommend investigating habitats with a high potential of the occurrence of cellulolytic bacteria, like intestinal tracts of herbivores or rotting trees. If an enrichment step is desired or inevitable, it is reasonable to enrich over a rather short-time period to keep diversity as broad as possible.

References

1. Beguin, P. and Aubert, J.P. (1994) The biological degradation of cellulose. FEMS *Microbiol. Rev.* 13, 25–58.
2. Lynd, L. R. and Zhang, Y. (2002) Quantitative determination of cellulase concentration as distinct from cell concentration in studies of microbial cellulose utilization: analytical framework and methodological approach. *Biotechnol. Bioeng.* **77**, 467–475.
3. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci.* 92, 7090–7094.
4. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**, 548–557.
5. Kumar, R., Singh, S., and Singh, O. V. (2008) Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* **35**, 377–391.
6. Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., et al. (1998) Mechanisms of cellulases and xylanases. *Biochem. Soc. Trans.* **26**, 156–160.
7. Hilden, L. and Johansson, G. (2004) Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnol. Lett.* **26**, 1683–1693.
8. Ando, S., Ishida, H., Kosugi, Y., and Ishikawa, K. (2002) Hyperthermostable endoglucanase from *Pyrococcus horikoshii*. *Appl. Environ. Microbiol.* **68**, 430–433.
9. Zhang, Y. H. and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol. Bioeng.* **88**, 797–824.
10. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., et al. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331 (Pt 3)**, 775–781.
11. Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., et al. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *J. Biol. Chem.* **279**, 34785–34793.
12. Coutinho, J. B., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J., and R. C. Miller, J. (1993) The nature of the cellulose-binding domain effects the activities of a bacterial endoglucanase on different forms of cellulose. *FEMS Microbiol. Lett.* **113**, 211–217.
13. Fontes, C. M., Clarke, J. H., Hazlewood, G. P., Fernandes, T. H., Gilbert, H. J., and Ferreira, L. M. (1997) Possible roles for a non-modular, thermostable and proteinase-resistant cellulase from the mesophilic aerobic soil bacterium *Cellvibrio mixtus*. *Appl. Microbiol. Biotechnol.* **48**, 473–479.
14. Klyosov, A. A. (1990). Trends in biochemistry and enzymology of cellulose degradation. *Biochemistry.* 29(47), 10577–10585.
15. Schwarz, W. H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* **56**, 634–649.
16. Shoham, Y., R. Lamed and E. A. Bayer (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol.* 7(7), 275–281.
17. Sanchez-Torres, J., Perez, P., and Santamaria, R. I. (1996) A cellulase gene from a new alkalophilic *Bacillus* sp. (strain N186-1). Its cloning, nucleotide sequence and expression in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **46**, 149–155.
18. Cazemier, A. E., Verdoes, J. C., Op den Camp, H. J., Hackstein, J. H., and van Ooyen, A. J. (1999) A beta-1,4-endoglucanase-encoding gene from *Cellulomonas pachnodae*. *Appl. Microbiol. Biotechnol.* **52**, 232–239.
19. Solingen, P., Meijer, D., Kleij, W., Barnett, C., Bolle, R., Power, S., et al. (2001) Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake. *Extremophiles* **5**, 333.
20. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–249.
21. Streit, W. R. and Schmitz, R. A. (2004) Metagenomics - the key to the uncultured microbes. *Curr. Opin. Microbiol.* **7**, 492–498.
22. Daniel, R. (2004) The soil metagenome - a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* **15**, 199–204.
23. Schmeisser, C., Steele, H., and Streit, W. R. (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl. Microbiol. Biotechnol.* **75**, 955–962.
24. Schmidt, T. M., DeLong, E. F., and Pace, N. R. (1991) Analysis of a marine picoplankton community by 16 S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378.

25. Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., et al. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ. Microbiol.* **7**, 1996–2010.
26. Ferrer, M., Golyshina, O. V., Plou, F. J., Timmis, K. N., and Golyshin, P. N. (2005) A novel alpha-glucosidase from the acidophilic archaeon *Ferroplasma acidiphilum* strain Y with high transglycosylation activity and an unusual catalytic nucleophile. *Biochem. J.* **391**, 269–276.
27. Beloqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O. V., Zumarraga, M., et al. (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *J. Biol. Chem.* **281**, 22933–22942.
28. Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., and Streit, W. R. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* **69**, 6235–6242.
29. Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O., and Shanmugam, K. T. (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl. Microbiol. Biotechnol.* **43**, 667–674.
30. Feng, Y., Duan, C. J., Pang, H., Mo, X. C., Wu, C. F., Yu, Y., et al. (2007) Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl. Microbiol. Biotechnol.* **75**, 319–328.
31. Grant, S., Sorokin, D. Y., Grant, W. D., Jones, B. E., and Heaphy, S. (2004) A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. *Extremophiles* **8**, 421–429.
32. Rees, H. C., Grant, S., Jones, B., Grant, W. D., and Heaphy, S. (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **7**, 415–421.
33. Voget, S., Steele, H. L., and Streit, W. R. (2006) Characterization of a metagenome-derived halotolerant cellulase. *J. Biotechnol.* **126**, 26–36.
34. Pottkämper, J., Barthen, P., Ilmberger, N., Schwaneberg, U., Schenk, A., Schulte, M., et al. (2009) Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chemistry* **11**, 957–965.
35. Guo, H., Feng, Y., Mo, X., Duan, C., Tang, J., and Feng, J. (2008) [Cloning and expression of a beta-glucosidase gene umcel3G from metagenome of buffalo rumen and characterization of the translated product]. *Sheng Wu Gong Cheng Xue Bao* **24**, 232–238.
36. Pang, H., Zhang, P., Duan, C. J., Mo, X. C., Tang, J. L., and Feng, J. X. (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr. Microbiol.* **58**, 404–408.
37. Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565.
38. Brosius, J., Ullrich, A., Raker, M. A., Gray, A., Dull, T. J., Gutell, R. R., et al. (1981) Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. *Plasmid* **6**, 112–118.
39. Kane, M. D., Poulsen, L. K., and Stahl, D. A. (1993) Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16 S rRNA sequences. *Appl. Environ. Microbiol.* **59**, 682–686.
40. Wild, J., Hradecna, Z., Posfai, G., and Szybalski, W. (1996) A broad-host-range *in vivo* pop-out and amplification system for generating large quantities of 50- to 100-kb genomic fragments for direct DNA sequencing. *Gene* **179**, 181–188.
41. Sektas, M. and Szybalski, W. (1998) Tightly controlled two-stage expression vectors employing the Flp/FRT-mediated inversion of cloned genes. *Mol. Biotechnol.* **9**, 17–24.
42. Teather, R. M. and Wood, P. J. (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl. Environ. Microbiol.* **43**, 777–780.

Chapter 2

Reversal of NAD(P)H Cofactor Dependence by Protein Engineering

Sabine Bastian and Frances H. Arnold

Abstract

There is increasing interest in utilization of engineered microorganisms for the production of renewable chemicals and next-generation biofuels. However, imbalances between the cofactor consumption of the engineered production pathway and the reducing equivalents provided by the cell have been shown to limit yields. This imbalance can be overcome by adjusting the cofactor dependencies of the pathway enzymes to match the available cofactors in the cell. We show how cofactor preference can be reversed by structure-guided directed evolution of the target enzyme.

Key words: Cofactor switch, Enzyme engineering, Homology modeling, Nicotine amide dinucleotide phosphate, Nicotine amide dinucleotide, Directed evolution

1. Introduction

Recent advances in biotechnology have made it possible to construct microorganisms harboring engineered metabolic pathways that convert renewable sugars into a wide array of chemicals (1, 2). However, the sugar-to-product yields of these engineered strains can be limited by imbalances between the specific (nicotine amide dinucleotide phosphate (NADPH) or nicotine amide dinucleotide (NADH)) cofactor requirements of the pathway enzymes and the form in which the reducing equivalents are provided by the cells. Reduced yields can pose a major obstacle to commercialization and economic viability of a microbial process for production of chemicals and fuels.

One solution to balancing NAD(P)H cofactor utilization in *Escherichia coli* is to overexpress a transhydrogenase, such as *E. coli* PntAB (3), which catalyzes the transfer of a hydride ion between $NADP^+$ and NADH resulting in the simultaneous regeneration of

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_2,

NADPH and NAD^+. However, applicability of this approach to yeast or other industrial microorganisms is questionable, since the transhydrogenase may not always shift the hydride ion in the preferred direction (4). A more general solution is to adjust the cofactor specificity of pathway enzymes to match the available reducing equivalents. Such an approach has improved ethanol yields of xylose fermentations (5, 6) and recently led to anaerobic production of isobutanol from glucose at theoretical yield (7).

Early strategies for engineering an enzyme's cofactor specificity in the absence of structural information involved disrupting or constructing salt bridges to the NADPH 2′ phosphate group. While this is sometimes sufficient to switch cofactor specificity, it is often accompanied by loss of catalytic activity (8, 9). Engineering an enzyme whose catalytic efficiency with its new cofactor is comparable to that of the native enzyme with its native cofactor requires more intricate remodeling of the cofactor-binding site. An accumulated wealth of structural information (10), software for homology modeling when structural information is not available, and established methods of directed enzyme evolution (11) supply the tools to successfully reverse cofactor dependence.

The following protocol was tailored for switching the cofactor preference of wild-type ketol-acid reductoisomerase (IlvC) from *E. coli* from NADPH to NADH (7). The overall strategy was to identify target residues based on their potential to affect cofactor recognition and use site-saturation mutagenesis and screening to discover beneficial mutations at those sites. Individual beneficial mutations were recombined to generate the cofactor-switched enzyme. This protocol is appropriate for a recombinant enzyme expressed in *E. coli* and can be adapted with appropriate adjustments in heterologous expression conditions, buffers and storage, substrates, etc.

2. Materials

2.1. Cell Culture and Lysis

1. Luria-Bertani (LB) broth supplemented with 100 μg/mL of ampicillin (Note 1).
2. Aqueous solution of 100 mg/mL ampicillin (=1,000× stock solution).
3. Aqueous solution of 0.5 M isopropyl β-D-1-thiogalactopyranoside (IPTG).
4. Electro-competent *E. coli* strain BL21(DE3) (Lucigen Corp., Middleton, WI, USA), stored at −80°C.
5. Lysis buffer for high-throughput lysis: 250 mM potassium phosphate buffer, pH 7.0, 750 mg/L lysozyme, 0.5 U/mg DNaseI.

6. Lysis buffer for purification: Buffer A (see Subheading 2.2).
7. Lysis buffer for assays in crude lysate generated by sonication: 250 mM potassium phosphate, pH 7.0.
8. Humidified plate shaker (e.g., Infors (Switzerland)).

2.2. Purification

1. Buffer A: 20 mM Tris pH 7.4, 20 mM imidazole, 100 mM NaCl, and 10 mM $MgCl_2$.
2. Buffer B: 20 mM Tris pH 7.4, 300 mM imidazole, 100 mM NaCl, and 10 mM $MgCl_2$.
3. 1 or 5 mL Histrap High Performance (HP) columns precharged with nickel (GE Healthcare, Waukesha, WI, USA).
4. Sonicator (Misonix, Inc., Farmingdale, NY, USA).
5. FPLC system such as an AKTA purifier (GE Healthcare, Waukesha, WI, USA).

2.3. Activity Assay

1. Substrate precursor ethyl 2-acetoxy-2-methylacetoacetate (EAMAA) (Sigma, St. Louis, MO), stored at room temperature.
2. 2 M NaOH solution, stored at room temperature.
3. Cofactors NADPH and NADH (Codexis, Inc., Redwood City, CA, USA), 10 mM, dissolved in water, stored in 1-mL aliquots at –20°C.
4. 2 M stock solution of $MgCl_2$, dissolved in water, stored at room temperature.
5. 100 mM stock solution of dithiothreitol (DTT), dissolved in water, stored in 1-mL aliquots at –20°C.
6. Bio-Rad Protein Assay (Bio-Rad, Hercules, CA, USA), stored at 4°C.
7. UV–vis spectrophotometer (Varian, Inc., Palo Alto, CA, USA).
8. Transparent flat-bottom assay plates (Evergreen Scientific, Los Angeles, USA).
9. Liquid handling robot (e.g., Beckman Coulter, Inc., Brea, CA, USA).
10. Plate reader (e.g., TECAN Group Ltd., Switzerland).

2.4. Construction of Enzyme Mutant Libraries by Site-Saturation Mutagenesis

1. PyMOL software (Molecular Graphics System, Version 1.3, Schrödinger, LLC., San Diego, USA). For questions and assistance using PyMOL, refer to http://www.pymolwiki.org/.
2. SWISS-Model: http://swissmodel.expasy.org/ (University of Basel, Switzerland).
3. Thermomixer (e.g., Eppendorf, Germany).
4. Thermocycler (e.g., Eppendorf, Germany).

5. Colony picking robot (e.g., Genetix, San Jose, CA, USA).
6. Design primers with Clone Manager or similar software, or use online tools such as http://arep.med.harvard.edu/labgc/adnan/projects/Utilities/revcomp.html.
7. Dissolve lyophilized primers in PCR-grade water to obtain 100-mM stock solutions. For the working solutions, prepare a 1:10 dilution in PCR-grade water. Primer stocks and working solutions are stored at −20°C.
8. NEB Phusion™ polymerase (New England Biolabs, Ipswich, MA, USA), stored at −20°C.
9. Aqueous stock solutions of deoxynucleotide triphosphates (dNTPs) (10 mM each), stored at −20°C.
10. *Dpn*I (Roche, Indianapolis, IN, USA), stored at −20°C.
11. Restriction enzymes *Nde*I and *Xho*I (New England Biolabs, Ipswich, MA, USA), stored at −20°C.
12. Expression vector pET22b(+) (EMD Chemicals Group, Darmstadt, Germany), stored at −20°C.
13. T4 ligase (New England Biolabs, Ipswich, MA, USA), stored at −20°C.
14. DNA sequencing may be performed by Laragen (Los Angeles, CA, USA) or other companies providing sequencing service.
15. Freeze "n" Squeeze tubes (Bio-Rad, Hercules, CA, USA), stored at room temperature.
16. Pellet paint kit (Merck, Germany), stored at −20°C.
17. Zymo DNA clean & concentrator™-25 kit (Zymo Research Corporation, Irvine, CA, USA), stored at room temperature.

3. Methods

The methods described herein are in order of the workflow. For purification purposes, an N- or a C-terminal his-tag is required (Note 2).

3.1. Structure Alignment and Choice of Target Residues

1. Assess cofactor–side chain interactions with PyMOL.
2. Target mutation sites can be selected following inspection of the cofactor-binding site in PyMOL, if the structure of choice has the native cofactor co-crystallized. Otherwise prepare a structural alignment of the X-ray crystal structures of homologs that contain the cofactor.
3. If the parent enzyme's structure is not available, computational models based on the structures of homologs and the amino acid sequence of the target enzyme (e.g., with SWISS-Model) will be very helpful in choosing the target amino acid residues (12).

4. When examining the cofactor-binding site for residues that when mutated could disrupt the salt bridges to the 2′ phosphate of NADPH, one should consider not only the positively charged residues such as arginines and lysines, but also residues that can impact the general positioning of the cofactor orientation.

3.2. Preparation of Site-Saturation Libraries

1. Flanking primers may bind up to 50 bp upstream and downstream of the gene (see Note 3).
2. Design mutagenesis primers for targeted positions with NNK (N stands for A, C, T, G; K stands for T, G; reverse complement MNN) in lieu of the triplet encoding the targeted amino acid residue. Adding 12–15 base pairs upstream and downstream of the NNK site will result in fragments with sufficiently long overlapping sequences for successful assembly PCR (Fig. 1) (see Note 4).
3. Use flanking primer and mutagenesis primer pairs to generate fragments with Phusion polymerase: PCR cycler program for Phusion: 98°C for 30 s initial denaturation; 98°C for 10 s, 55°C for 10 s, 72°C for 30 s/kb (25 cycles); 10 min at 72°C, 4°C hold step (not required) (see Note 5).
4. PCR mix for generation of fragments: The numbers in parentheses are the concentrations of the stock solutions: 10 μL HF buffer (5×), 1 μL dNTPs (10 μM each), 1 μL forward primer (10 μM), 1 μL reverse primer (10 μM), 0.5 μL plasmid template (100 ng/μL), 0.5 μL Phusion polymerase, and 36 μL PCR-grade water (see Note 6).
5. Check for correctly sized PCR products on Tris-acetic acid-EDTA (TAE) agarose gel.
6. *Dpn*I digest template DNA for 1 h at 37°C: Add 1 μL of *Dpn*I to 50 μL of PCR mixture and mix gently by snapping it with your fingers rather than pipetting up and down. No desalting

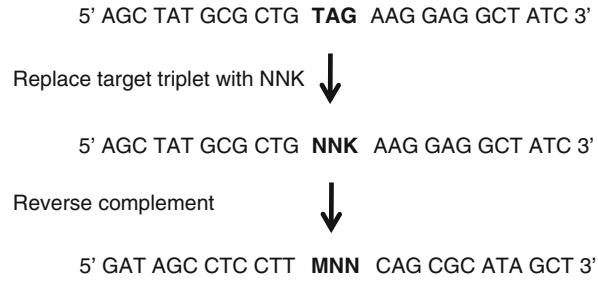


Fig. 1. Primer design for site-saturation libraries: replace target triplet TAG (in *bold*) with NNK to obtain forward primer. Then, reverse complement NNK forward primer to obtain reverse primer. N stands for A, T, G, and C; K stands for T and G; M is the reverse complement of K.

or any other kind of PCR purification is required prior to addition of the enzyme.

7. Separate fragments on preparative TAE agarose gel, excise on UV table, transfer gel fragments into Freeze "n" Squeeze tubes, and then freeze for at least 10 min at –20°C (see Note 7).
8. Spin for 10 min at room temperature at 10,000 × *g*. For larger gel fragments, longer spin times may be necessary.
9. Precipitate DNA contained in the eluate of the Freeze "n" Squeeze tubes using the pellet paint kit (see steps 10–16) (see Note 8).
10. For sample volumes smaller than 500 µL, add 2 µL of co-precipitant and vortex. For larger sample volumes, split samples in half and add pellet paint co-precipitant.
11. Add 1/10 volume of 3 M sodium acetate, pH 5, provided in the kit, and vortex.
12. Add 2.5× volume of 100% ethanol and vortex.
13. Incubate on ice for 2 min and centrifuge for 5 min at 16,000 × *g* and 4°C.
14. Remove supernatant with pipette; add 200 µL 70% ethanol, vortex, and spin again as described above.
15. Remove supernatant with pipette; add 200 µL 100% ethanol, vortex, and spin as described above.
16. Remove supernatant with pipette, dry DNA in thermomixer at 55°C for 10 min or until pellet is dry by visual inspection, and resuspend pellets in 10 µL of PCR-grade water.
17. Use fragments as templates for assembly PCR with flanking primers, the numbers in parentheses are the concentrations of the stock solutions: 10 µL HF buffer (5×), 1 µL dNTPs (10 mM each), 1 µL forward primer (10 µM), 1 µL reverse primer (10 µM), 0.5 µL fragment a and fragment b each (~50 ng/µL each), 0.5 µL Phusion polymerase, and 35.5 µL PCR-grade water.
18. Program for assembly PCR with Phusion: 98°C for 30 s initial denaturation; 98°C for 10 s, 55°C for 10 s, 72°C for 30 s/kb (25 cycles); 10 min at 72°C, 4°C hold step (not required).
19. Check for successful assembly PCR on TAE agarose gel.
20. Zymo purification according to modified manufacturer's instructions and elute with 20 µL PCR-grade water (see Note 9).
21. Restriction digest insert and vector with *Nde*I overnight at 37°C: Use the entire 20 µL of the purified insert, and add 5 µL NEB buffer 4, 23 µL PCR-grade water, and 1 µL *Nde*I; mix gently and spin briefly. On the next morning, add 1 µL *Xho*I, mix gently, spin briefly, and incubate for another 2 h at 37°C.

Use 10 µL of vector (100 ng/µL stock solution) and treat the same as insert, but adjust volume of PCR-grade water so that the total volume is 50 µL (see Note 10).

22. Gel extract as described above (steps 7–16, this Section).
23. Ligation: Use T4 ligase from NEB. Prepare the following mixture: molar ratio of 3–5:1 for insert:vector, 3 µL ligation buffer (10×), 2 µL ligase, add PCR-grade water to a total reaction volume of 30 µL. Incubate at 16°C overnight.
24. Zymo purify ligation mixture as described above (step 20), but elute in 10 µL PCR-grade water.
25. Use appropriate volumes of desalted ligation mixture for electroporation of *E. coli* BL21(DE3) cells.
26. Plate on LB_{amp} plates to obtain single colonies and incubate at 37°C overnight.

The entire process described in this section is summarized in Fig. 2.

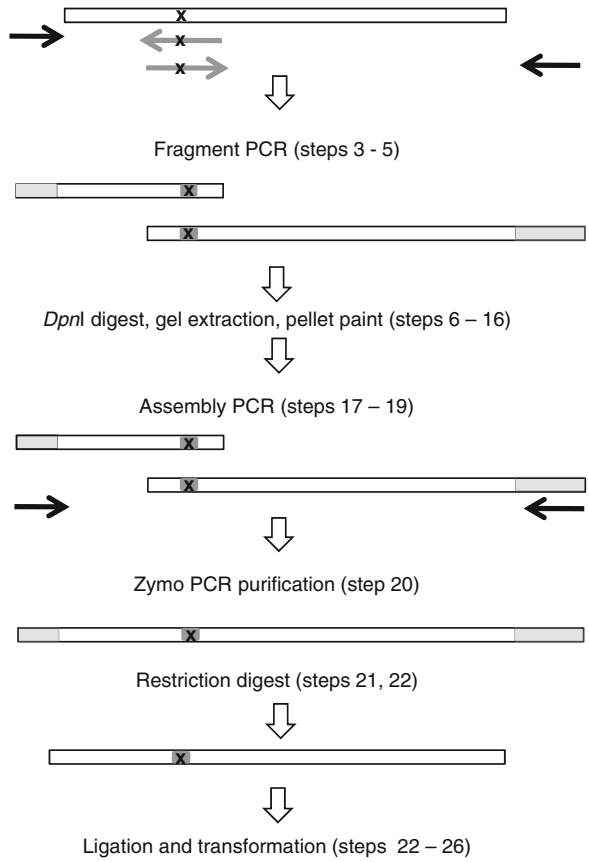


Fig. 2. Schematic for preparation of site-saturation mutagenesis libraries, based on SOE PCR (13). *Black arrows*: flanking primers; *gray arrows*: mutagenesis primers; *black x*: targeted site.

3.3. Protein Expression for Library Screening

1. A picking robot may be used to transfer single *E. coli* colonies into shallow 96-well plates filled with 300 µL LB_{amp}. As an alternative, single colonies can be picked manually using sterile toothpicks.
2. The 96-well plates should contain several controls: parent or wild-type clones, BL21(DE3) colonies carrying pET22b(+) only to control for background reactions in cell lysates, and a well filled with medium only, which serves as sterility control (see Note 11).
3. Cover plates with either air pore tapes or plastic lids and let grow overnight at 37°C with shaking at 210 rpm at 80% humidity (see Note 12).
4. The next morning, fill 96-well deep-well plates with 600 µL of LB_{amp} per well and inoculate with 75 µL of the overnight cultures.
5. Let cultures grow at 37°C and 210 rpm for 4 h in a humidified plate shaker with 80% humidity. Then, reduce the temperature of the incubator to 25°C, let cultures cool for 1 h and induce with IPTG at a final concentration of 0.5 mM (see Note 13).
6. After induction, growth and expression continue for up to 24 h at 25°C and 210 rpm.
7. Centrifuge the cells at 5,300 × *g* and 4°C for 5 min, pour out supernatant, tap plates onto paper towels, and store upside down at −20°C for a minimum of 2 h (see Note 14).

3.4. Screening of Libraries

1. Prepare the substrate (*R/S*)-2-acetolactate fresh for each assay (see Note 15): Mix 50 µL ethylethyl-2-acetoxy-2-methylacetoacetate (EAMAA) with 990 µL of water. Then, gradually add 260 µL of 2 M NaOH in 10-µL increments. Each addition is followed by a 15-second vortexing step. After the entire amount of 2 M NaOH is added, mix the solution on a shaker at 130 rpm for 20 min at room temperature. Store substrate on ice until further use (14).
2. Thaw 96-well deep well plates containing the libraries at room temperature for 20 min upside down with open lids on paper towels (see Note 16).
3. Add lysis buffer (see Subheading 2.1, step 5 and see Note 17) using a multichannel pipette and vortex vigorously until pellets are resuspended. Then, agitate plates at 37°C and 130 rpm for 1 h.
4. Dilute lysate appropriately using 250 mM potassium phosphate buffer, pH 7.0 and centrifuge for 10 min at 5,000 × *g* and 4°C (see Note 18).
5. Transfer appropriate amount of lysate (see Notes 19 and 20) into transparent flat-bottom assay plates using a liquid handling robot.

6. Start reaction by addition of assay buffer (250 mM potassium phosphate pH 7, 10 mM $MgCl_2$, 1 mM DTT, 4 mM acetolactate, and 200 mM NADH or NADPH).
7. Monitor the depletion of cofactor at 340 nm in a plate reader over 1 min at 25°C.

3.5. Rescreening and Recombination Library

1. Calculate the ratio of NADH/NADPH consumed during the screening assay.
2. Choose variants that exhibit favorable activity with the new cofactor, but are only moderately active or inactive with the native cofactor.
3. Streak selected variants on LB_{amp} plates and pick at least in triplicate for rescreening in 96-well plates. This is necessary to ascertain that monoclonal variants are being rescreened.
4. Rescreen as described above and sequence improved variants.
5. Design primers for recombination of beneficial mutations: (1) design one complementary primer pair for each targeted site, (2) ensure that the primers encode all beneficial mutations found at the targeted site, (3) include the wild-type sequence, (4) use the IUPAC nomenclature to find the correct degenerate codon triplet.
6. If the IUPAC degeneration covers the desired mutations, but also adds unwanted amino acid changes (which enlarge the library size and increase the screening effort), design primers individually for each beneficial mutation and prepare a mixture of equimolar concentrations. Include a primer pair encoding the wild-type sequence (see Note 21).
7. Generate the recombination library using the protocol described above (see Subheading 3.2) with the difference that the recombination library contains several targeted sites in the gene rather than just one as in the example above.
8. Create all fragments individually (Fig. 3) and treat as described above.
9. Use up to five fragments for one assembly PCR, and then continue as described above or refer to Note 22 for solutions to encountered difficulties.
10. Proceed with screening and rescreening of recombination library and sequence the best hits (see Subheadings 3.3 and 3.4).

3.6. Protein Expression on 25-mL Scale

1. Inoculate 5 mL Luria-Bertani (LB) broth supplemented with 100 μg/mL of ampicillin with an *E. coli* BL21(DE3) transformant carrying the pET22b(+) expression plasmid. Grow overnight at 37°C with shaking at 250 rpm.

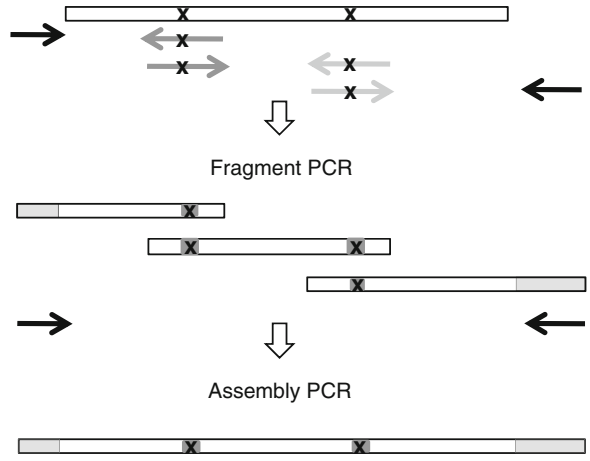


Fig. 3. Schematic for preparation of the fragments and the assembly of the fragments of a recombination library. *Black arrows*: flanking primers; *dark gray* and *light gray arrows*: mutagenesis primers; *black x*: targeted sites. For the entire process, refer to Fig. 2.

2. Use overnight culture to inoculate a 25-mL LB_{amp} expression culture supplemented with 100 µg/mL of ampicillin to an initial optical density (OD_{600}) of 0.1.
3. Grow expression culture at 250 rpm and 37°C until an OD_{600} of 1 is reached.
4. Reduce temperature to 25°C and let culture cool for 30 min at 250 rpm.
5. Induce expression by the addition of IPTG to a final concentration of 0.5 mM.
6. After incubation for 24 h at 250 rpm and 25°C, centrifuge cells at 5,300 × *g* and 4°C for 10 min, discard supernatant and freeze pelleted cells at −20°C.

3.7. Protein Purification

1. For additional information on purification, refer to Note 23.
2. Thaw *E. coli* cell pellets at room temperature for 20 min.
3. Resuspend cell pellets at a ratio of 0.25 g wet weight/mL in Buffer A.
4. Lyse the resuspended cells by sonication for 1 min with a 50% duty cycle.
5. Centrifuge at 11,000 × *g* and 4°C for 15 min and then filter through a 0.45-µm filter.
6. Perform the following purification steps at 4°C.
7. Use 1 mL/min flow rate for 1-mL histrap columns and 5 mL/min for 5-mL histrap columns.
8. Check binding capacity of columns.

9. Purification method: (1) 4-column volume (cv) equilibration step with buffer A, (2) injection of sample, (3) washout unbound sample step with buffer A for 2 cv, (4) 5-cv wash step with 10% elution buffer B + 90% buffer A, (5) elution at 40% buffer B + 60% buffer A, (6) clean after elution at 100% B, (7) re-equilibration with buffer A.
10. IlvC and its variants were stored at 4°C in 40% B as eluted. A desalting step may be added, if the enzyme does not store well in high imidazole concentrations.

3.8. Characterization of Enzyme Variants

1. As a preliminary assessment of enzyme variants, the activity ratios (in U/mg) in the presence of the new and the native cofactor can be determined in *E. coli* lysates. Interesting variants should be purified using a highly reproducible method (as described in Subheading 3.7) for ease of comparison.
2. IlvC activity was assayed by monitoring the linear decrease in NAD(P)H concentration at 340 nm over 1 min at 25°C on a UV–vis spectrophotometer.
3. Assay buffers: 250 mM potassium phosphate pH 7, 1 mM DTT, 10 mM (R/S)-2-acetolactate, and 10 mM $MgCl_2$. Prepare one batch of assay buffer containing 200 μM NADPH and a second batch containing 200 μM NADH final concentration.
4. Place lysate or purified enzyme (sample volume 100 μL) into a cuvette (see Note 24), and initiate the reaction by addition of 900 μL of the assay buffer containing either NADPH or NADH, thereby diluting the assay buffer by 10%.
5. Use Bio-Rad Protein assay or similar assays by other suppliers to determine the concentration of total protein in lysates and in purified samples.
6. Desired properties: After first round of NNK libraries, the variants should have gained activity on the new cofactor while maintaining or losing activity on the native one. After the recombination round, the preferred variants should have native-like activity with the new cofactor while exhibiting no or very low activity with the native cofactor.

4. Notes

1. Unless mentioned otherwise, all solutions should be prepared in water that has a resistivity of 18.2 MΩ-cm and a total organic content of less than five parts per billion. The standard referred to as "PCR-grade water" for molecular biology applications was autoclaved.

2. This protocol can be adapted for other proteins that are expressed in *E. coli*.
3. We found using flanking primers, which bind ~50 bp upstream and downstream of the insert, to be beneficial. These 50-bp fragments will be visible on an agarose gel after restriction digest and will be an indicator of digest completion.
4. NNK primers:
   (a) For DNA degeneracies, see IUPAC nomenclature http://www.chem.qmul.ac.uk/iupac/.
   (b) NNK (32 codons, 1 stop codon, 20 amino acids, 94 colonies to screen for threefold oversampling) is preferred over NNN (64 codons, 3 stops, 20 amino acids, 190 colonies to screen for threefold oversampling).
5. We found Phusion polymerase worked very well for generating fragments and assembly PCR products. Also, it was beneficial for the incorporation of the mutations to reduce the annealing temperature to 55°C in cases where the recommended annealing temperature was higher.
6. General advice for assembly PCRs: Always perform a positive control PCR with the flanking primers using the plasmid as template. The length of this PCR product will serve as comparison to identify correctly assembled products on the agarose gel.
7. Freeze "n" Squeeze tubes with DNA containing gel fragments may be frozen longer than 10 min, even overnight, but then spinning at lower speeds than 10,000 × *g* is recommended to avoid breaking the tubes in the centrifuge.
8. Pellet paint recommendations:
   (a) The manual recommends precipitating for 2 min at room temperature. For optimal results, precipitate on ice or in the freezer (−20°C). Also, the centrifugation steps are better performed at 4°C instead of the suggested room temperature spin step.
   (b) The pink pellets are fairly loose. It is best to discard the supernatant carefully with a pipette rather than pouring it out.
   (c) Pellets containing ethanol have a dark pink appearance. As soon as they are dry, the color is very light pink.
9. Zymo PCR purification kit modifications and recommendations:
   (a) For optimal results, add a spin step (13,000 × *g*, 4°C, 1 min) to remove residual ethanol after washing twice with ethanol and before eluting the DNA in PCR-grade water. Be sure to empty collecting tube before spinning to dry.

(b) For elution of a total volume of 20 μL, place 10 μL on top of the spin columns, let sit for at least 1 min, centrifuge, add the remaining 10 μL, let sit again for 1 min, and finally elute by centrifugation. This increases the yield.

10. Even though fast digest enzymes from NEB were used, we obtained better results with longer incubation. We digested with *Nde*I overnight and then the next morning added *Xho*I for 2 h instead of the recommended 10–30 min. This was not detrimental to the inserts.
11. High-throughput expression for screening: it is crucial to place parent and negative control on each plate, since plate to plate expression and activity variations may be observed.
12. Cultivation conditions for high-throughput expression: when growing small culture volumes at 37°C, prevent uneven evaporation of cultures across the plates by increasing the shaker's humidity to 80%.
13. Induction with 0.5 mM IPTG worked very well for the expression of wild-type IlvC and its variants. However, the IPTG concentration should be adjusted in cases where inclusion bodies are observed.
14. Freezing the 96-well plates containing the cell pellets is essential for a successful lysis step. The plates should be frozen for a minimum of 2 h, preferably overnight.
15. In the case of acetolactate, it was important to make the substrate fresh as required. The substrate cannot be stored in the freezer or fridge. This may not be true for other substrates.
16. Thawing the cells at room temperature is important to assist the resuspension step. If the cell pellets do not resuspend despite vigorous vortexing, it is best to resuspend individual wells by pipetting up and down with a P200 or larger.
17. It is crucial to add DNaseI to the lysis buffer of the high-throughput screen. Without the addition of DNaseI, the lysate will turn out too viscous for accurate transfer with either a liquid handling robot or a multichannel pipette.
18. Dilution of the lysate should be done prior to centrifugation. The amount of buffer added depends on how active an enzyme is. This should be tested prior to the actual screening in a screening validation experiment using an entire 96-well plate containing parent with three wells carrying the vector only to account for background activity.
19. When establishing a high-throughput assay, we found that transferring larger lysate volumes (>20 μL) reduces the pipetting error.

20. Lysate transfer with the liquid handling robot usually generates bubbles, which will disturb activity measurements in a plate reader. A simple, but efficient, way to remove them is to centrifuge the assay plates briefly (e.g., 1 min at 4,000 × *g*).
21. When using an equimolar solution of mixed primers in the recombination library, it is useful to sequence randomly chosen clones to control for potential nucleotide biases.
22. Solutions for difficulties encountered with recombination libraries:
    (a) More than five fragments to assemble: Assemble four or fewer fragments, Zymo purify the resulting products, and then use these products as templates for a final assembly PCR step.
    (b) Two targeted sites are so close to each other that the resulting fragment would be <100 bp, but the sites cannot be covered with one primer: Recombine all sites except one of the two, then use the full-length assembly product as template to introduce the last site following the site-saturation mutagenesis protocol.
    (c) Less than five fragments, but one fragment does not assemble: Assemble stepwise. If necessary, elongate the primer to generate a longer overlap sequence. This problem can occur with very short fragments (<100 bp) at the 5′ or the 3′ end.
23. General notes on purification:
    (a) In general, all purification steps, from sonication to elution, should be performed at 4°C to minimize potential protease activities and to ascertain protein stability.
    (b) The histrap column manuals contain information on storage, flow rate, pressure limits, and binding capacity.
    (c) For optimal purification results, it is beneficial to determine the amount of his-tagged protein in the crude extract and then choose a column size (1 or 5 mL) that will allow for loading at or even slightly exceed the capacity limit.
    (d) When working in Tris buffer and at 4°C, it is important to readjust the pH after the buffers have cooled down. The pH of Tris is temperature-sensitive.
    (e) Determine the point of elution of your protein via a linear gradient. Then, establish a step gradient such as the method described for IlvC.
24. We determined the Michaelis–Menten constants for the cofactors in a spectrometer rather than using a plate reader. The plate reader did not have the required sensitivity to accurately measure K_M values.

Acknowledgments

The authors would like to thank Dr. Christopher Snow for assistance with homology modeling. This work was sponsored by the U.S. Army Research Laboratory and was accomplished under cooperative Agreement number W911NF-09-2-002.

References

1. Rude, M. A., and Schirmer, A. (2009) New microbial fuels: a biotech perspective, *Current Opinion in Microbiology 12*, 274–281.
2. Yan, Y., and Liao, J. (2009) Engineering metabolic systems for production of advanced fuels, *Journal of Industrial Microbiology & Biotechnology 36*, 471–479.
3. Weckbecker, A., and Hummel, W. (2004) Improved synthesis of chiral alcohols with *Escherichia coli* cells co-expressing pyridine nucleotide transhydrogenase, NADP(+)-dependent alcohol dehydrogenase and NAD(+)-dependent formate dehydrogenase, *Biotechnology Letters 26*, 1739–1744.
4. Nissen, T. L., Anderlund, M., Nielsen, J., Villadsen, J., and Kielland-Brandt, M. C. (2001) Expression of a cytoplasmic transhydrogenase in *Saccharomyces cerevisiae* results in formation of 2-oxoglutarate due to depletion of the NADPH pool, *Yeast 18*, 19–32.
5. Bengtsson, O., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. F. (2009) Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*, *Biotechnology for Biofuels 2*.
6. Jeppsson, M., Bengtsson, O., Franke, K., Lee, H., Hahn-Hagerdal, R., and Gorwa-Grauslund, M. F. (2006) The expression of a *Pichia stipitis* xylose reductase mutant with higher Km for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*, *Biotechnology and Bioengineering 93*, 665–673.
7. Bastian, S., Liu, X.; Meyerowitz, J.; Snow, C. D.; Chen, M. M.Y.; Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic isobutanol production at theoretical yield in *Escherichia coli*, *Metabolic Engineering 13*, 345–352.
8. Clermont, S., Corbier, C., Mely, Y., Gerard, D., Wonacott, A., and Branlant, G. (1993) Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase - role of the acidic residue in the fingerprint region of the nucleotide-binding fold, *Biochemistry 32*, 10178–10184.
9. Scrutton, N. S., Berry, A., and Perham, R. N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering *Nature 343*, 38–43.
10. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity, *Protein Science 18*, 2125–2138.
11. Romero, P. A., and Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution, *Nature Reviews Molecular Cell Biology 10*, 866–876.
12. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M.-y., Pieper, U., Sali, A. (2006) Comparative Protein Structure Modeling Using Modeller, *Current Protocols in Bioinformatics 15*, 5.6.1–5.6.30.
13. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymology 154*, 367–382.
14. Krampitz, L. O. (1957) Preparation and determination of acetoin, diacetyl, and acetolactate, *Methods in Enzymology 3*, 277–283.

Chapter 3

Quantifying Plasmid Copy Number to Investigate Plasmid Dosage Effects Associated with Directed Protein Evolution

Samuel Million-Weaver, David L. Alexander, Jennifer M. Allen, and Manel Camps

Abstract

Our laboratory specializes in directed protein evolution, *i.e.*, evolution of proteins under defined selective pressures in the laboratory. Our target genes are encoded in ColE1 plasmids to facilitate the generation of libraries *in vivo*. We have observed that when random mutations are not restricted to the coding sequence of the target genes, directed evolution results in a strong positive selection of plasmid origin of replication (*ori*) mutations. Surprisingly, this is true even during evolution of new biochemical activities, when the activity that is being selected was not originally present. The selected plasmid *ori* mutations are diverse and produce a range of plasmid copy numbers, suggesting a complex interplay between *ori* and coding mutations rather than a simple enhancement of level of expression of the target gene. Thus, plasmid dosage may contribute significantly to evolution by fine-tuning levels of activity. Here, we present examples illustrating these observations as well as our methods for efficient quantification of plasmid copy number.

Key words: Recombinant gene expression, ColE1 plasmid, Plasmid copy number, Green fluorescent protein, ALKBH2, Transformation, Mutagenesis, R-loop, RNA I, RNA II, Directed evolution, Methyl methane sulfonate, *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine

1. Introduction

ColE1-like plasmids share mechanisms for control of replication. ColE1-like replication initiation is orchestrated by a ~600-bp sequence known as plasmid origin of replication or *ori* (reviewed in refs. (1–3)). This sequence is transcribed, generating a preprimer that forms a stable DNA–RNA hybrid (R-loop) at its 3′ end. The preprimer RNA is then processed to a primer by RNAse H and extended by DNA polymerase I (Pol I), initiating leading-strand synthesis and facilitating the recruitment of the Pol III primosome, the replication complex responsible for completing plasmid replication (4).

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_3, © Springer Science+Business Media, LLC 2012

ColE1-like plasmids are maintained at medium- and high-copy numbers, which make them the most popular vectors for recombinant protein expression in *E. coli* and useful as shuttle vectors for other organisms. Plasmid copy number is controlled by a negative feedback mechanism mediated by transcription of an antisense RNA from an alternative promoter (P1). The resulting 108 bp antisense transcript (RNA I) hybridizes with the 5′ end of the preprimer RNA (RNA II) as it is being transcribed, locking the preprimer in a conformation that is incompatible with the formation of the R-loop at its 3′ end and thus preventing replication initiation (1, 2). Both preprimer RNA II and antisense RNA I form three stem-loops that are critical for this regulatory mechanism (SL1, 2, 3 and SL1′, 2′, 3′ respectively), as hybridization depends on the formation of a "kissing complex" between unpaired bases at their respective complementary loops (1, 2). As preprimer transcription proceeds further, the formation of a fourth loop (SL4) makes the preprimer refractory to RNA I inhibition (5). Mutations in SL1, 2, 3, and 4 are frequently found in *ori* mutants exhibiting increased plasmid copy number (reviewed in ref. (3)).

Our laboratory specializes in the evolution of proteins in the laboratory, an approach known as *directed evolution*. Our random mutant libraries are generated *in vivo* using a mutator strain of *E. coli* expressing a low-fidelity form of Pol I (6, 7). Error-prone plasmid replication in this strain generates mutations throughout a Pol I-dependent plasmid sequence encoding the gene of interest. Plasmids isolated from these cultures constitute a mutant library, where mutations are randomly distributed across our gene of interest as well as sequences regulating transcription and plasmid replication. Our libraries can be subjected to functional selections in order to direct the evolution of specific changes in biochemical activity of the plasmid-encoded gene (6, 7). We previously established that since Pol I is gradually replaced by the Pol III primosome (7, 8), the mutation frequency in our libraries decreases with increasing distance from the plasmid DNA/RNA switch. In the absence of a functional selection, the plasmid *ori* shows the lowest mutation load of the plasmid because it is the most distal sequence relative to the DNA/RNA switch (7).

One of our targets for directed evolution is the human oxidative demethylase ALKBH2 (reviewed in ref. (9, 10)). This enzyme removes two highly cytotoxic DNA *N*-methyl adducts (N3-methylC and N1-methylA), which are made in abundance by the agent methyl methane sulfonate (MMS). *E. coli* cells that are deficient in AlkB (the *E. coli* homologue of ALKBH2) are hypersensitive to MMS, and this hypersensitivity can be partially complemented by ALKBH2 (7). By contrast, ALKBH2 expression confers no measurable protection against *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG), a stronger methylating agent that (unlike MMS) generates abundant cytotoxic oxygen adducts.

We performed two functional selections on human ALKBH2 mutant libraries containing an average of 1.5 mutations/kb. The two selections were protection from MMS toxicity (through repair of ALKBH2 canonical lesions), and protection from MNNG toxicity (through repair of alternative cytotoxic substrates). The procedures for library generation and selection for resistance to methylation toxicity using LB agar gradients have been described in detail in (7). Our sequencing data for 6 MMS-selected clones and for 6 MNNG-selected clones are presented in Table 1. This table shows that the plasmid *ori* shows strong signs of positive selection in both cases. In the case of the MMS selection, we estimated a 23-fold increase in *ori* mutation frequency relative to an unselected library; by contrast, the coding sequence shows only a 2.2-fold enrichment. The modest magnitude of enrichment seen in the coding sequence is likely due to a strong purifying selection (reviewed in ref. (11)). Unexpectedly, the plasmid *ori* showed an almost identical level of positive selection in the MNNG selection (22-fold enrichment, compared to 23-fold for MMS). This suggests that *ori* mutations may not only contribute to optimizing existing activities but also facilitate the evolution of new biochemical activities. As a control, we also looked for evidence of positive selection within ~300 bp downstream of the RNA/DNA switch, a nonconserved intervening sequence (Table 1). In this case, we detected only a 1.8-fold enrichment in mutations. In this case, the weak positive selection is likely due to a subset of mutations immediately adjacent to the RNA/DNA switch that may either influence primer extension by Pol I or modulate local supercoiling at the *ori* (Fig. 1a).

In agreement with the observed positive selection of sequences involved in regulation of plasmid replication, all selected ALKBH2 mutants analyzed (a total of 9, shown in Fig. 2) exhibited alterations in plasmid copy number. Except for one case, these alterations were not dramatic. Indeed, only 9 out of 25 *ori* mutations map to sites where mutations are known to produce strong increases in copy number (SL1, 2, 3, 4 or P1); of these, four occurred in SL4, with generally weaker effects (Fig. 1b). Thus, both the mutation and phenotypic profiles of our ALKBH2 mutants suggest that alterations in plasmid dosage involve a complex interplay between *ori* and coding mutations rather than a simple enhancement of level of expression of the target gene. We propose that *ori* mutations likely contribute significantly to evolution by fine-tuning levels of activity.

Interestingly, plasmid copy number determined by gel band quantification does not appear to correlate with transformation efficiency (Fig. 3). This indicates that transformation efficiency is not necessarily a good indicator of ColE1 plasmid copy number. Thus, it seems that altering plasmid replication initiation often has pleiotropic effects on plasmid establishment.

Table 1
Sequence of ALKBH2 clones identified following methylating agent selection

Clone_selection	Nonsynonymous coding	*Ori*			Downstream of *ori*	
1_MMS	G50D G64D	T113C	G189A	C409T		
2_MMS	H144R	T202A	T589C		G646T	C754T
3_MMS	H144Y	C238T			C643T	
4_MMS	A53T P55L	T170C	G325A	T479A	G755A	
5_MMS	P27L P207S	G512A				
6_MMS	H51L T121N	C328T	G592A		T840C	
1_MNNG	A78V L138I R193S D201G P207L S237C	G338A				
2_MNNG	P180L	G53A	G95T	G373A		
3_MNNG	A97T	G53A	C198T	G477A	T687A	
4_MNNG	H144Y N231D	G180A				
5_MNNG	R58W A219D P229L K255E	G44A	C116T	G180A	T666A, C667T	G850A
6_MNNG	R58W A118T	G258A			A696G	
Total mutations	26 (22 new)	25			9	
Expected without selection	10	1.1			5.1	
Enrichment through positive selection (fold)	2.2[a]	23[b]			1.8[c]	
MMS	10	13			4	
Expected without selection	5	0.55			2.55	
Enrichment through positive selection (fold)	2.0	23			1.6	
MNNG	16 (12 new)	12			5	
Expected without selection	5	0.55			2.55	
Enrichment through positive selection (fold)	3.2 (2.4 new)	22			2.0	

[a]Total bases: $795 \times 12 = 9{,}546$; expected frequency of bp substitutions (without selection) ~1.5/kb × 9.55 kb = 14. Expected frequency of amino acid substitutions (~30% synonymous (23)) ~10. Mutations present before randomization are underlined

[b]Total bases: $613 \times 12 = 7{,}356$; expected frequency of base-pair substitutions (without selection) ~0.15/kb × 7.36 kb = 1.1

[c]Total bases: $280 \times 12 = 3{,}360$; expected frequency of base-pair substitutions (without selection) ~1.5/kb × 3.4 kb = 5.1. For MMS and MNNG selections, comprising 6 and 6 clones respectively, each of the expected values calculated for the total needs to be divided by half

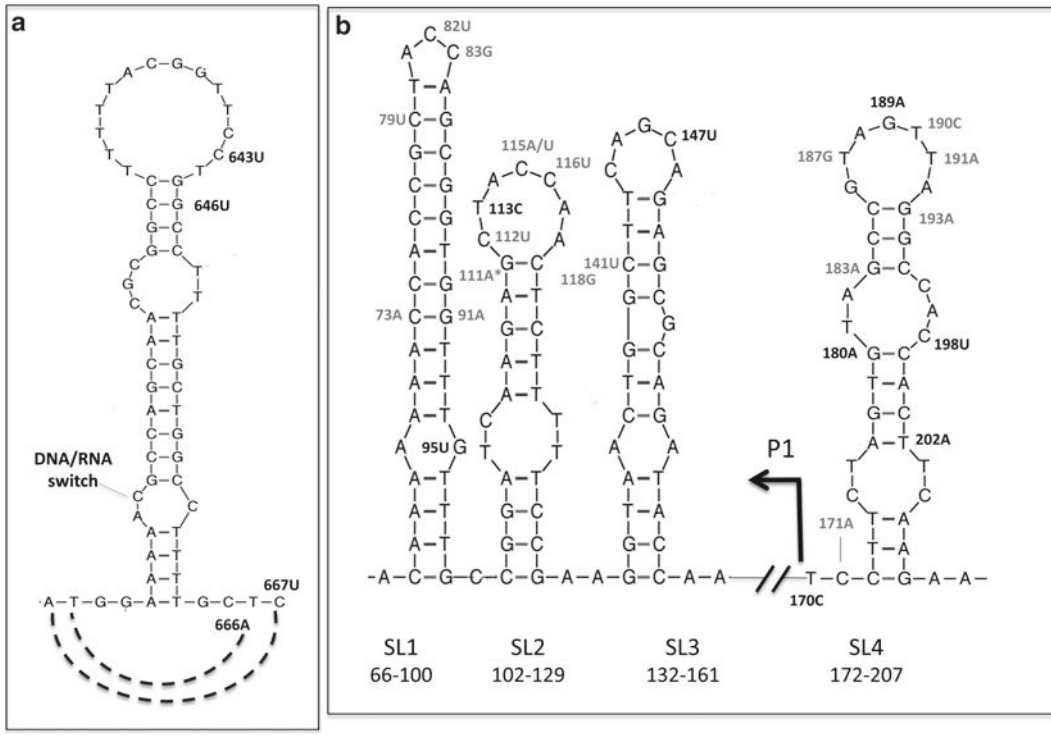


Fig. 1. Location of plasmid dosage-modulating mutations within known plasmid copy number regulatory sites. Methylating agent-selected ALKBH2 mutations located in sites controlling plasmid replication initiation are shown in *black*. For comparison, previously described mutations increasing ColE1 plasmid copy number are shown in *grey* (their references in the literature can be found in (3)). Graphic representations were generated using the mfold program (22), which predicts secondary structures of single-stranded nucleic acids based on a thermodynamic model. (**a**) Area surrounding the DNA/RNA switch, where R-loop formation, primer processing and Pol I extension occur. (**b**) Antisense RNA I regulatory elements (SL 1, 2, 3 and promoter P1) and preprimer SL4.

We also looked to see whether green fluorescent protein (GFP) (12) could be used as a reporter for plasmid copy number, as this would greatly facilitate monitoring how different treatments, culture conditions, or genetic backgrounds impact plasmid copy number. To that end we used a commercially available plasmid, pGFPuv (Clontech) encoding "cycle 3 GFP," a mutant GFP with improved expression in *E. coli* (13). We generated nine pGFPuv *ori* mutants exhibiting a range of plasmid copy numbers and compared the amount of plasmid recovered with level of GFP fluorescence. We found a good but not perfect correlation between these two variables ($r^2 = 0.67$, Fig. 4a). As a control, we confirmed that GFP fluorescence is representative of GFP protein levels in the cell, although again the correlation was not very tight ($r^2 = 0.60$, Fig. 4b).

In sum, this work suggests that plasmid dosage likely contributes significantly to evolution by fine-tuning levels of the activity under selection. This mechanism could be particularly relevant

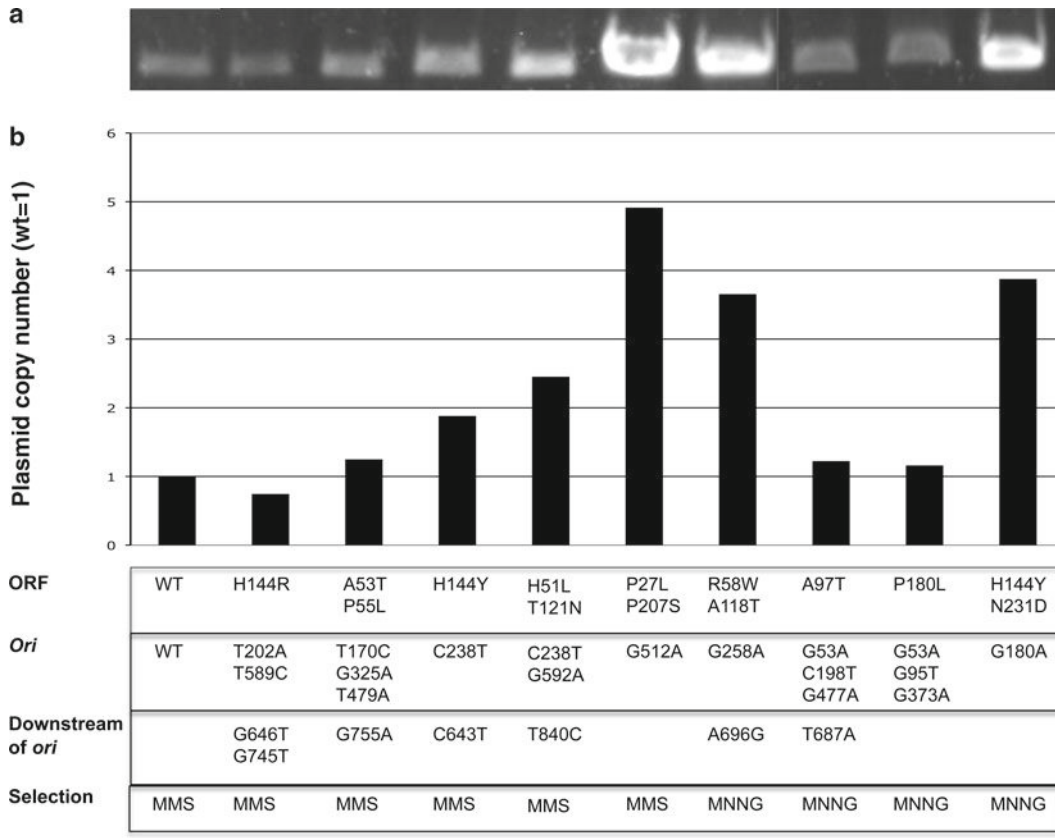


Fig. 2. Copy number of selected ALKBH2 mutants. Plasmids bearing different ALKBH2 mutants were extracted from BL21 cells following transformation. Plasmid DNA was isolated from equal numbers of cells based upon the cell density of the culture, then linearized by digestion with a single-cut restriction enzyme (*Pci I*) and separated on an agarose gel for imaging and quantification. The amino acid substitutions found in the coding sequence of ALKBH2, base pair substitutions found in *ori* and adjacent downstream sequence, and the methylating agent used for selection are listed at the bottom. (**a**) Representative images of quantification gels. (**b**) Quantification of corresponding gel band intensities normalized to WT.

for the evolution of metabolic pathways, as ColE1 regulation of plasmid replication is designed to integrate complex metabolic input (reviewed in ref. (3)). Below we describe the methods we used to efficiently quantify plasmid copy number, which are generally applicable to any instance where plasmid dosage affects functional protein expression. We isolated plasmid DNA from bacterial cultures normalized by optical density (dry weight equivalents) and measured average plasmid DNA content by visualization on agarose gels. Alternative plasmid copy number quantification methods include [^{3}H]thymidine labeling and quantification of released radiation from closed circular DNA (14, 15); Southern blotting, either as a relative measure of copy number (16), or using a [^{32}P] labeled probe for quantification (17); and real-time quantitative PCR with primers targeting both the plasmid of interest and a gene within the bacterial chromosome (18, 19). Note that these

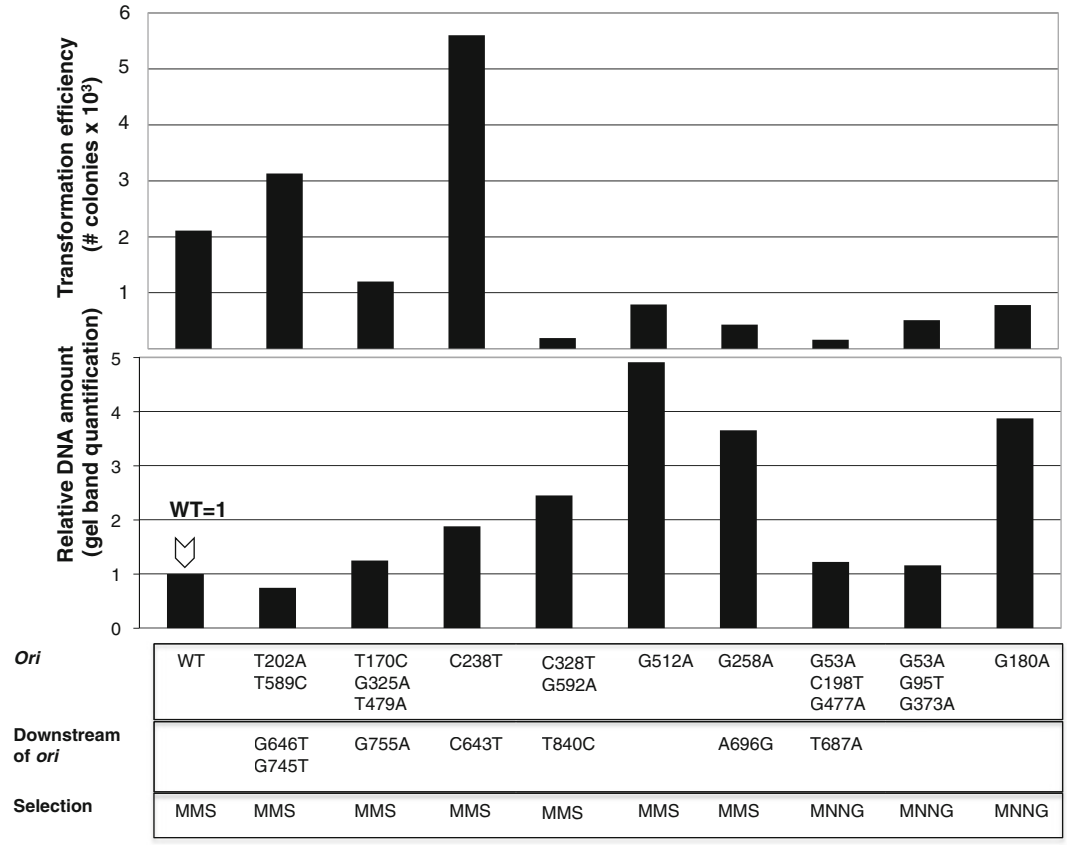


Fig. 3. Transformation efficiency of selected ALKBH2 mutants. One hundred nanograms of plasmid DNA were transformed into BL21 cells. Cells were then plated on agar plates containing carbenicillin, the selective antibiotic for the ALKBH2 plasmids. Equal numbers of cells where subject to transformation (5×10^7 cells) and the total number of successful transformations determined as colonies growing on carbenicillin plates are shown (*top panel*). For comparison, the *bottom panel* shows plasmid copy number levels as shown in Fig. 2b.

methods are a measure of the average plasmid copy number present in the cell culture. The distribution of copy numbers (which is likely linked to plasmid stability (3)) is not addressed here.

2. Materials

2.1. Transformation

1. Competent cells.
   (a) BL21 *ompT gal* [*dcm*] [*lon*] *hsdSB* (*rB- mB-*) (*F-*) (ALKBH2 experiments).
   (b) CJ278 *CM4722* (*A*(*gal-bio*) *thi-J relAl spoTl poA*+ $^{+}$ (F^{+})) *ΔPol A with unknown suppressor* transformed with pHSG-Pol I (pGFPuv experiments).
2. ColE1 vector.

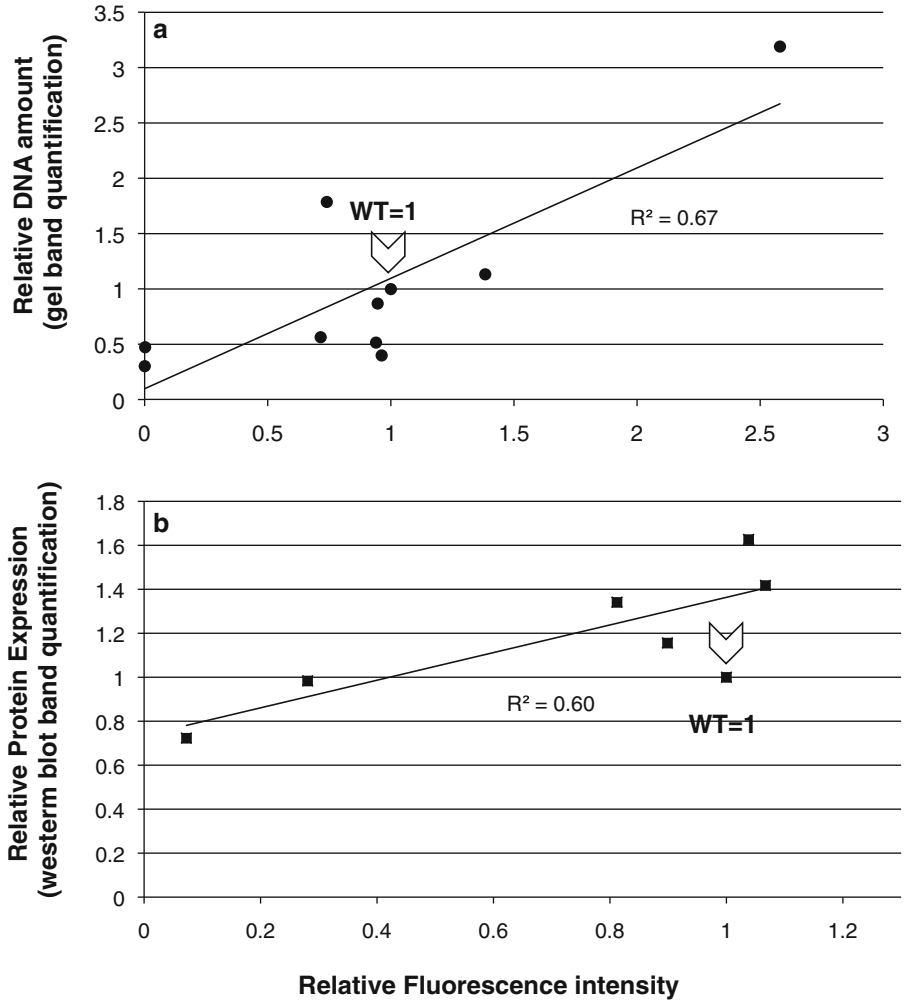


Fig. 4. GFP as a reporter for plasmid copy number. A set of pGFPuv mutants exhibiting a range of plasmid copy numbers was chosen and characterized by standard gel-based plasmid quantification (linearizing the vector with the restriction enzyme *HindIII*), measurement of GFP fluorescence, and Western blot analysis of GFP protein. Values are expressed relative to those of the wild-type *ori*. (**a**) Correlation between plasmid gel quantification and GFP fluorescence. For this experiment cells had to be grown at 30°C because some of the pGFPuv mutants in our panel were not viable at 37°C, possibly because of temperature-dependent additional increases in copy number leading to runaway plasmid replication (see Note 3). (**b**) Correlation between GFP fluorescence signal and GFP protein expression as detected by Western blot.

3. Luria-Bertani (LB) medium: 10 g/L tryptone, 5.0 g/L yeast extract, and 5.0 g/L NaCl.
4. LB agar: add 1.5% agar to LB medium for agar plates.
5. Fisher 100×15 mm disposable Petri dishes.
6. Carbenicillin solution, 100 mg/mL, store at –20°C.
7. Carbenicillin (100 μg/mL) LB agar plates.
8. Eppendorf Electroporator 2510.
9. 2 mm gap electroporation cuvettes.

2.2. Washing Plates and Plasmid Recovery

1. Clear, Flat-Bottomed 96-well plates.
2. Molecular Devices Spectra Max M2e S Plate Reader.
3. Plasmid miniprep kit.

2.3. Digestion and Quantification of Recovered Plasmids

1. Restriction enzymes and corresponding buffers.
2. Agarose.
3. TAE Buffer: 40 mM Tris-Acetate, 1 mM EDTA.
4. 1 kb DNA ladder.
5. Gel loading buffer: 0.25% Bromophenol Blue, 0.25% Xylene Cyanol, 50% Glycerol in diH_2O.
6. SYBR Gold, Store at –20°C in opaque microcentrifuge tubes protected from visible light.
7. Large opaque container, such as a Tupperware with painted sides or covered in aluminum foil.
8. UVP Biospectrum 300 outfitted with dual UV light box, Cohu 6400 CCD camera, and SYBR (517–570 nm) and GFP emission filters (570–620 nm) and VisionworksLS version 6.8 software (UVP, LLC).

2.4. Quantification of GFP Fluorescence

1. Molecular Devices SpectraMax M2e Fluorometric and Spectrophotemtric plate reader. Dual monochromaters Absorbance 200–1,000 nm and excitation 250–850 nm (Molecular Devices).
2. Black Greiner 96-well microplates with clear bottoms (Fisher Scientific).
3. Phosphate Buffered Saline (PBS): 0.2 g/L KCl, 0.2 g/L KH_2PO_4, 8 g/L NaCl, 2.16 g/L $Na_2HPO_4.4H_2O$.

2.5. Sequencing of Plasmids of Interest

1. NanoDrop Spectrophotometer for DNA quantification (Thermo Scientific).
2. MacVector version 10.5 for sequence analysis (Symantic).

3. Methods

Below we show protocols for quantifying plasmid copy number. Plasmids are recovered following transformation, from cells washed from the selecting plate to avoid possible generation and amplification of additional regulatory mutations through growth in liquid culture (see Notes 1–3). Recovered plasmids are linearized with a single-cutting restriction enzyme, separated on an agarose gel, and imaged as a means to quantitate plasmid copy number.

3.1. Transformation of Plasmids of Interest

3.1.1. Transformation of ColE1 Plasmid by Heat-Shock (Used for Our pGFPuv Experiments; see Note 1–4)

1. Prepare a 4-mL overnight culture in LB media of the cell line of interest. Include selective antibiotic in the media for the desired cell line.
2. Expand this overnight culture into a sterile 2-L Erlenmeyer flask containing 400 mL of LB media.
3. Incubate the flask at 37°C with shaking until the cells reach exponential growth phase, as measured by turbidity (cultures with an OD_{600} of 0.6).
4. Chill the flask containing cells on wet ice for 15 min.
5. Transfer the liquid cultures to plastic centrifuge bottles, and centrifuge for 15 min at 4°C and 4,000 rpm.
6. Resuspend the pelleted cells in approximately 10 mL of chilled calcium chloride solution (60 mM $CaCl_2$, 15% Glycerol, 10 mM HEPES, pH7).
7. Transfer the cells into a 50-mL conical tube, and fill to the 50 mL line with chilled calcium chloride solution.
8. Centrifuge the cells at 4°C, 4,000 rpm for 15 min.
9. Pour off the supernatant.
10. Resuspend the cells in calcium chloride, fill the tube to the 50 mL mark with chilled calcium chloride solution, and centrifuge at 4°C, 4,000 rpm for 15 min.
11. Repeat the wash step.
12. After the third and final wash, decant or pipette off the supernatant from the pelleted cells.
13. Resuspend the cell pellet in approximately 1:1 volume of chilled calcium chloride solution.
14. Keep cells on wet ice and use immediately, or aliquot competent cells into 1 mL microcentrifuge tubes on dry ice for storage at –80°C (see Note 5).
15. Pipette 100 μL chemically competent cells into a 5-mL culture tube.
16. Add 500 ng plasmid to the tube.
17. Incubate on ice for 10 min.
18. Heat-shock in a water bath at 42°C for 2 min.
19. Resuspend cells into 1 mL of LB medium in a 12×75-mm or larger capped culture tube, and recover cells for an hour with shaking at 37°C (see Note 6).
20. Plate transformed cells by spreading onto Petri dishes containing LB agar and 0.1 mg/mL carbenicillin (see Note 7).
21. Let the cells grow overnight at 37/30°C.

3.1.2. Transformation of ColE1 Plasmid by Electroporation (Used for Our ALKBH2 Mutants; see Note 1–4)

1. Expand desired cell line then pellet 400 mL culture as described above (Subheading 3.1.1, step 1–5).
2. Resuspend the pelleted cells in approximately 10 mL of chilled 10% glycerol solution.
3. Transfer the cells into a 50-mL conical vial, and fill the vial to the 50 mL mark with chilled 10% glycerol.
4. Centrifuge at 4°C, 4,000 rpm for 15 min.
5. Pour off the supernatant.
6. Wash the cells 3 more times with chilled 10% glycerol.
7. Resuspend the pelleted cells from the final wash in 1:1 volume of chilled 10% glycerol.
8. Keep cells on wet ice and use immediately, or aliquot competent cells into 0.6 mL microcentrifuge tubes on dry ice for storage at −80°C (see Note 5).
9. Add 40 μL electrocompetent cells to a 0.2 cm gap electroporation cuvette.
10. Add 100 ng plasmid to the cuvette in volume <1 μL.
11. Electroporate at 1,800 V, with a time-constant between 4 and 6 ms.
12. Transfer cells to 1 mL LB medium and recover for 1 h with shaking at 37°C (see Note 6).
13. Plate transformed cells by spreading onto Petri dishes containing LB agar and 0.1 mg/mL carbenicillin (see Note 7).
14. Let the cells grow overnight at 37/30°C.

3.2. Washing Plates and Recovery of Plasmids

1. Wash plates using a sterile plate spreader into 1.5 mL eppendorf tubes (see Note 8).
2. Measure the optical density at 600 nm (OD_{600}) of each plate wash using a plate reader (see Note 9).
3. Normalize each plate by volume based upon its optical density (see Note 10).
4. Centrifuge normalized plate washes for 1 min at 11,000 × g.
5. Pipette off supernatant.
6. Harvest plasmids of interest from pelleted cells using a plasmid miniprep kit.

3.3. Digestion and Quantification of Recovered Plasmids

1. Digest 5 μL of eluted plasmid from each miniprep (see Note 10), using an appropriate restriction enzyme and buffer in a 20-μL volume reaction, following the manufacturer's specifications (see Note 11).
2. Prepare a 0.8% agarose gel using TAE buffer.

3. Load the gel with an appropriate DNA ladder.
4. Mix each digest with an appropriate amount of gel loading buffer. Load the entire digest into each well of the gel.
5. Run the gel at 120 V, 400 mA for 60–90 min.
6. Prepare 50 mL of a 1:10,000 dilution of SYBR Gold in TAE (see Note 12).
7. Place the gel in an opaque container. Pour the SYBR Gold solution over the gel, allow the gel to stain overnight, shaking at 4°C.
8. Visualize the gel with using a UVP BioSpectrum 300 LM26E transilluminator equipped with a SYBR filter (515–570 nm). Set the shutter speed to 544 ms. Adjust the camera aperture and image contrast so that none of the fluorescent bands saturate the image (see Note 13).
9. Use VisionWorksLS version 6.8 to calculate the Optical Density of the gel band representing each plasmid assayed (see Note 14).

3.4. Quantification of GFP Fluorescence

1. Colonies of transformed cells are harvested from plates and normalized by volume and OD_{600} (see Notes 7–9). These plate washes then represent a population of cells. The fluorescence quantification represents an average of this population.
2. The OD_{600} normalized cultures are washed into PBS at room temperature (see Note 15). Cells are diluted in PBS (threefold and ninefold).
3. 200 μL of neat and diluted cells are loaded into black Greiner 96-microwell plates with clear bottoms.
4. GFP fluorescence is quantified using a Molecular Devices Spectromax M2e plate reader at excitation 395 nm and emission 510 nm.
5. GFP fluorescence readings are compared to total GFP immunoreactive protein as determined by Western blot (see Note 16).

3.5. Sequencing of Plasmids of Interest

Purified plasmids are first quantified. We prefer using a Nanodrop spectrophotometer for the small sample requirement and automated analysis. In accordance with the manufactures instructions, a 2-μL droplet of the plasmid-containing solution is placed on the pedestal and the DNA concentration determined based upon a fixed length light path and absorbance at 260 nm determined relative to a solvent blank. Aliquots of plasmid DNA containing 500 ng of DNA in 5–7 μL of water are sent for single primer Dye termination sequencing (Sequentech). Sequences are assembled and analyzed using the program MacVector version 10.5 (see Note 1).

4. Notes

1. Verifying the purity of the plasmid sample to be tested; plasmid *ori* mutations (particularly those at the area of RNAI/RNAII overlap) alter the compatibility properties of plasmids (20), allowing the stable maintenance of more than one plasmid species within a clone. Thus, *ori* mutants isolated *in vivo* from single colonies often contain mixed sequences. Transformation of mixed plasmid species can confound experimental results because a mutant initially present in a small fraction of the sample can become predominant if it has higher transformation efficiency and/or provides higher fitness than the initially most abundant species. Thus, plasmid sequences need to be checked for the presence of mixed sequence, which is seen as a double peak on the chromatogram.
2. Culture conditions are critical. Rich media increases plasmid copy number, possibly by allowing a higher metabolic burden on the cell (18). Saturation conditions also favor increased copy number, possibly due to increased R-loop formation resulting from alterations in supercoiling associated with titration of R-loop-suppressing factors and/or shifts in the transcriptional profile of the cell (for a review see ref. (3)).
3. The temperature at which the culture is grown also has a strong impact upon the plasmid copy number. Growth at higher temperatures tends to increase plasmid copy number due to a destabilizing effect on RNAI/RNA II hybrid formation and other temperature-sensitive alterations in the folding of the preprimer RNA (5, 21), and to direct and indirect effects on R-loop formation.
4. Electroporation tends to produce higher transformation efficiencies, but the efficiency of transformation can vary substantially between individual electroporations. Transformation using chemically competent cells produces more reproducible transformation efficiencies, facilitating comparison between different clones.
5. Competent cells should be stored at –80°C, and will keep for several months. Competent cells should be thawed slowly on ice before use, and refrozen immediately after use. Competent cells left out at room temperature for over an hour should be discarded.
6. We usually grow cells at 37°C. Failure of certain clones to grow at 37°C may be indicative of runaway plasmid replication. In that case the experiment is performed at 30°C.
7. Following transformation, adjust the number of cells plated or plate different cell dilutions to produce semi-lawns, *i.e.*, a large

number of separately growing colonies (200–800 for a 100 × 15-mm Petri dish). This ensures adequate clonal representation while avoiding additional physiological variables (such as early nutrient and O_2 deprivation) occurring in a lawn.

8. The plate washes may be too dense to obtain an accurate OD_{600} reading directly. The range across which a plate reader can accurately report OD_{600} of a culture is 0.1–1.0. Dilute each wash 1:10 and 1:100 within a 96-well plate to obtain accurate readings. Use OD_{600} results from the dilute cultures to calculate the actual OD_{600} of each wash.
9. For example: 0.25 mL of cells with an O_{D600} of 4.0 is equivalent in dry weight to 1 mL of cells with an O_{D600} of 1.0.
10. Because DNA analyzed on the agarose gel is obtained from a fixed number of cells, variations in copy number will result in differences in band intensity on the gel. For optimal detection, adjust volume of digested DNA based upon intensity of bands obtained on the gel: if too faint, repeat with an increased volume of DNA from each prep; if overloaded, repeat with a decreased volume of DNA.
11. When selecting restriction enzymes to linearize plasmid DNA, if possible choose enzymes that cut the plasmid of interest only once and that do not have star activity. Be sure all plasmid is linearized as incomplete digestion will skew analysis.
12. SYBR Gold is used to visualize the bands on the gel because it provides a greater dynamic range of signal intensity than ethidium bromide and thus gives a more quantitative readout of plasmid copy number. This reagent is not considered a carcinogen, greatly simplifying safety and disposal. Keep solutions containing SYBR gold and gels stained with SYBR gold protected from visible light to avoid photodegradation.
13. VisionsworksLS has a digital imaging utility that highlights areas of the gel-documentation image that are light-saturated. Adjust the camera's parameters so that none of the bands are saturated in order to best compare the pixel area density of each band.
14. Gel band intensities are quantified using VisionWorksLS version 6.8's area density utility. Regions encompassing each band are defined and the mean pixel area density determined for each band. The area densities obtained for each clone are divided by the area density of the band corresponding to the wild-type *ori* on the gel. This yields a measure of plasmid copy number expressed as "fold wild-type" for each clone.
15. For fluorescence determination, cultures need to be washed and resuspended in PBS because LB has an interfering autofluorescence at excitation 395 nm and emission 510 nm, the excitation and emission optima for GFPuv.

16. Protein extracts for Western blot analysis are prepared from washed transformation plates as described in Notes 6–8. Western analysis was performed as described previously (24) blocking with 5% dry milk in PBS. Santa Cruz Biotechnology rabbit polyclonal antibodies (#SC8334) were used to probe for GFP protein levels, and detected using a Millipore goat antirabbit IgG Horse Radish Peroxidase conjugated secondary antibody (#12-384); protein bands are visualized using Pierce Supersignal WestPico reagents and protocols (#34077). Films are then imaged using the UVP BioSpectrum 300 LM26E Transilluminator with white transillumination. Band intensity is quantified using Visionworks version 6.8 (see Note 12).

Acknowledgments

The authors would like to thank Dr. Barbara Sedgwick for the gift of the BS141 (AB1157 F') and BS143 (AB1157 *alkB* F') strains, Dr. Catherine Joyce for the gift of the CJ278 (*polA*Δ) strain, Dr. Lawrence Loeb for mentorship in the initial stages of this work, and Jacob Marquette for his help with the generation of pGFPuv libraries. This work was supported by K08 award CA116429-04 to Manel Camps.

References

1. Polisky, B. (1988) ColE1 replication control circuitry: sense from antisense. *Cell*, **55**, 929–932.
2. Cesareni, G., Helmer-Citterich, M. and Castagnoli, L. (1991) Control of ColE1 plasmid replication by antisense RNA. *Trends Genet*, **7**, 230–235.
3. Camps, M. (2010) Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Pat DNA Gene Seq*, **4**, 58–73.
4. Masai, H., Nomura, N., Kubota, Y. and Arai, K. (1990) Roles of phi X174 type primosome- and G4 type primase-dependent primings in initiation of lagging and leading strand syntheses of DNA replication. *J Biol Chem*, **265**, 15124–15133.
5. Polisky, B., Zhang, X.Y. and Fitzwater, T. (1990) Mutations affecting primer RNA interaction with the replication repressor RNA I in plasmid CoIE1: potential RNA folding pathway mutants. *EMBO J*, **9**, 295–304.
6. Camps, M., Naukkarinen, J., Johnson, B.P. and Loeb, L.A. (2003) Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc Natl Acad Sci USA*, **100**, 9727–9732.
7. Troll, C.J., Alexander, D.L., Allen, J.M. and Marquette, J.T. (2010) Mutagenesis and functional selection protocols for directed evolution of proteins in *E. coli*. *The Journal of Visualized Experiments*, Mar **16**(49). pii: 2505. doi: 10.3791/2505.
8. Allen, J.M., Simcha, D.M., Ericson, N.G., Alexander, D.L., Marquette, J.T., Van Biber, B.P., Troll, C.J., Karchin, R., Bielas, J.H., Loeb, L.A. *et al.* (2011) Mutational footprints of ColE1 plasmid replication by error-prone DNA polymerase I identify primosome loading and sites of Okazaki primer processing. *Nucleic Acids Research*, **39**, 7020–7033.
9. Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Bratlie, M.S., Pena-Diaz, J., Otterlei, M., Slupphaug, G. and Krokan, H.E. (2004) Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)*, **3**, 1389–1407.
10. Sedgwick, B., Robins, P. and Lindahl, T. (2006) Direct removal of alkylation damage from

DNA by AlkB and related DNA dioxygenases. *Methods Enzymol*, **408**, 108–120.

11. Camps, M., Herman, A., Loh, E. and Loeb, L.A. (2007) Genetic constraints on protein evolution. *Crit Rev Biochem Mol Biol*, **42**, 313–326.
12. Phillips, G.N., Jr. (2006) The three-dimensional structure of green fluorescent protein and its implications for function and design. *Methods Biochem Anal*, **47**, 67–82.
13. Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol*, **14**, 315–319.
14. Hashimoto-Gotoh, T. and Inselburg, J. (1979) ColE1 plasmid incompatibility: localization and analysis of mutations affecting incompatibility. *J Bacteriol*, **139**, 608–619.
15. Schmidt, L. and Inselburg, J. (1982) ColE1 copy number mutants. *J Bacteriol*, **151**, 845–854.
16. Fitzwater, T., Zhang, X.Y., Elble, R. and Polisky, B. (1988) Conditional high copy number ColE1 mutants: resistance to RNA1 inhibition in vivo and in vitro. *EMBO J*, **7**, 3289–3297.
17. Merlin, S. and Polisky, B. (1993) Analysis of establishment phase replication of the plasmid ColE1. *J Mol Biol*, **230**, 137–150.
18. Wang, Z., Xiang, L., Shao, J., Wegrzyn, A. and Wegrzyn, G. (2006) Effects of the presence of ColE1 plasmid DNA in Escherichia coli on the host cell metabolism. *Microb Cell Fact*, **5**, 34.
19. Carapuca, E., Azzoni, A.R., Prazeres, D.M., Monteiro, G.A. and Mergulhao, F.J. (2007) Time-course determination of plasmid content in eukaryotic and prokaryotic cells using real-time PCR. *Mol Biotechnol*, **37**, 120–126.
20. Kim, D., Rhee, Y., Rhodes, D., Sharma, V., Sorenson, O., Greener, A. and Smider, V. (2005) Directed evolution and identification of control regions of ColE1 plasmid replication origins using only nucleotide deletions. *J Mol Biol*, **351**, 763–775.
21. Castagnoli, L., Lacatena, R.M. and Cesareni, G. (1985) Analysis of dominant copy number mutants of the plasmid pMB1. *Nucleic Acids Res*, **13**, 5353–5367.
22. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406–3415.
23. Wong, T.S., Roccatano, D., Zacharias, M. and Schwaneberg, U. (2006) A statistical analysis of random mutagenesis methods used for directed protein evolution. *J Mol Biol*, **355**, 858–871.
24. Ausubel F M (1995) Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology (Wiley, New York).

Chapter 4

High Isoprenoid Flux *Escherichia coli* as a Host for Carotenoids Production

Wonchul Suh

Abstract

A noncarotenogenic microbe *E. coli* was engineered for high production of carotenoids. To increase the isoprenoid flux, the chromosomal native promoters of the rate-controlling steps (*dxs, idi* and *ispDispF*) in the isoprenoid pathway were replaced with a strong bacteriophage T5 promoter (P_{T5}) by using the λ-Red recombinase system in combination with the Flp/FRT site-specific recombination system for marker excision and P1 transduction for gene trait stacking. The resulting high isoprenoid flux *E. coli* can be used as a starting strain to produce various carotenoids by introducing heterologous carotenoid genes. In this study, the high isoprenoid flux *E. coli* was transformed with a plasmid carrying the β-carotene biosynthetic genes from *Pantoea stewartii* for β-carotene production.

Key words: High isoprenoid flux *Escherichia coli*, Chromosomal promoter replacement, Metabolic engineering, Isoprenoid pathway, Carotenoids, β-carotene

1. Introduction

Carotenoids are pigments that are ubiquitous throughout nature and synthesized by plants and some microorganisms. Industrial uses of carotenoids include pharmaceuticals, food supplements, animal feed additives, and colorants in cosmetics. *E. coli* has drawn attention as a convenient host for various natural and novel carotenoids production because of easy expression of many carotenoid gene clusters, availability of genetic tools, and common usage as a production host for large-scale bioprocesses (1, 2). *E. coli* naturally cannot produce carotenoids, but has a common upper isoprenoid pathway for carotenoid biosynthesis, namely the 1-deoxyxylulose-5-phosphate (DXP) pathway (3), which produces a common precursor, isopentenyl pyrophosphate (IPP), for carotenoids. *E. coli* can produce carotenoids by introducing the carotenoid biosynthetic

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_4,

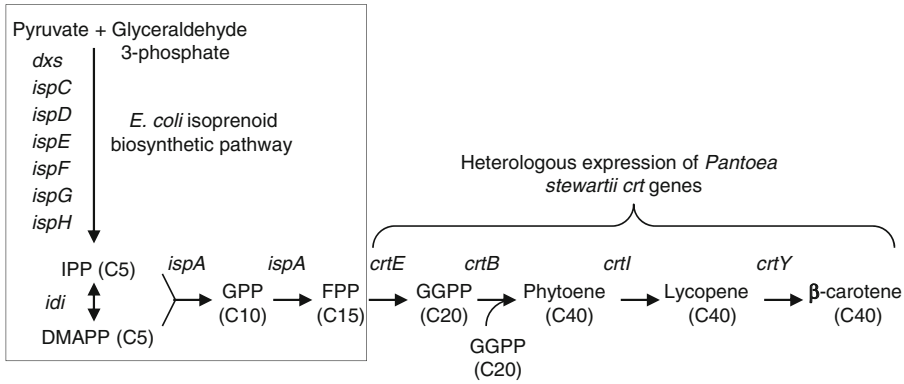


Fig. 1. *E. coli* 1-deoxyxylulose-5-phosphate (DXP) pathway, an upper pathway for carotenoid biosynthesis. *IPP* isopentenyl diphosphate; *DMAPP* dimethylallyl diphosphate; *GPP* geranyl diphosphate; *FPP* farnesyl pyrophosphate; *GGPP* geranylgeranyl diphosphate.

gene clusters isolated from carotenogenic microorganisms (4). The DXP pathway for the isoprenoid biosynthesis in *E. coli* transformed with β-carotene biosynthetic genes from *Pantoea stewartii* is shown in Fig. 1.

It has been known that carotenoids produced in recombinant *E. coli* is limited by the availability of isoprenoid precursors (5). Engineering the supply of isoprenoid precursors for increased production of carotenoids is necessary. To increase metabolic flux of the isoprenoid pathway, in general, a multicopy vector has been used to express several rate-limiting genes (*dxs* and *idi*) of the pathway (6). This method of metabolic engineering has several drawbacks such as segregational instability, vector burden, and difficulty in controlling gene expression (7, 8). To avoid the undesirable effects of using a multicopy vector, we previously used a chromosomal promoter replacement approach using homologous recombination *via* the λ-Red recombination system in combination with the Flp/FRT site-specific recombination system for marker excision and P1 transduction for gene trait stacking (9). The method utilizes the λ-Red homologous recombination system (10) to introduce an integration cassette including a kanamycin resistance gene (Kan^R) as a selection marker and bacteriophage T5 promoter (P_{T5}) in front of target isoprenoid genes on the chromosome of a recombination proficient host cell (*E. coli* MC1061), and subsequently utilizes a phage P1 transducing system to transfer the integration cassette into a production host cell (*E. coli* MG1655) in a combinatorial way (9). After selection of the transductant, a helper plasmid pCP20 (10), expressing the flippase (Flp) site-specific recombinase, is introduced into the cells to excise a Kan^R selectable marker bounded by FRT site-specific recombinase sites. The method facilitates assessment of various combinations of chromosomal modifications for biosynthetic pathway optimization. A reporter plasmid pPCB15 carrying β-carotene synthesis genes *Pantoea*

stewartii crtEXYIB gene cluster was used to monitor increased metabolic flux through the isoprenoid pathway based on the yellow color characteristic of β-carotene accumulation in cell.

2. Materials

2.1. Strains and Plasmids

1. *E. coli* K-12 strain MC1061 (*F– araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL(Strr) hsdR2(rK–mK+) mcrA mcrB1*) (11) was used as a recombination proficient host cell for chromosomal promoter replacement mediated by λ-Red homologous recombination.
2. *E. coli* K-12 strain MG1655 (*λ–, F–, rph–1*) (12) was used as a host strain for stacking promoter modifications.
3. The λ-Red recombinase system plasmids, plasmid pKD46 (ampicillin^R), pKD4 (kanamycin^R), and pCP20 (ampicillin^R) were obtained from Dr. Wanner (10). Plasmid pKD46 contains the λ-Red recombinase system, which comprised three genes *exo, bet, and gam,* expressed under the control of an arabinose-inducible promoter. The flippase (Flp) site-specific recombinase is contained on pCP20, which has a temperature-sensitive origin of replication and can be cured from the host cells by culturing cells at 43°C. Plasmid pKD4 was used as a template for PCR amplification of the *FRT*-flanked kanamycin resistance marker.
4. Plasmid pPCB15 (chloramphenicol^R) (9) contains β-carotene biosynthetic genes crtEXYIB from *Pantoea stewartii*.
5. Plasmid pSUH5 (kanamycin^R) (9) that is derived from pKD4 was used as a template for PCR amplification of the fused kanamycin selectable marker (Kan^R)-bacteriophage T5 promoter (P_{T5}).

2.2. Culture Media and Antibiotics

1. LB Agar (L): 10 g Bacto-tryptone, 5 g Yeast Extract, 10 g NaCl, and 15 g Agar, and LB Broth (L): 10 g Bacto-tryptone, 5 g Yeast Extract, and 10 g NaCl.
2. SOB medium (L): 20 g Bacto-tryptone, 5 g Yeast Extract, 0.5 g NaCl, 2.5 mL of 1 M KCl, adjust pH to 7.0 with 10 N NaOH, autoclave to sterilize, and add 10 mL of 1 M $MgCl_2$ before use.
3. SOC medium (L): 20 g Bacto-tryptone, 5 g Yeast Extract, 0.5 g NaCl, 2.5 mL of 1 M KCl, 10 mL of 1 M $MgCl_2$, 10 mL of 1 M $MgSO_4$, 20 mL of 1 M glucose, and sterilize by autoclaving.
4. Ampicillin (100 mg/mL water); chloramphenicol (25 mg/mL 100% ethanol); kanamycin (25 mg/mL water). Dissolve

antibiotic compounds in water or ethanol, sterilize by filtration, and store in aliquots at −20°C.

5. LB agar plates containing ampicillin, chloramphenicol, and/or kanamycin were prepared by first allowing autoclaved media to cool at 55°C and then adding antibiotic to a final concentration of 100, 25, and 25 μg/mL, respectively.

2.3. PCR and λ-Red Recombination

1. Primers for PCR amplification of a cassette Kan^R-P_{T5} for the promoter replacement of *dxs* gene; 5′-kan(dxs) (5′-TGGAAGCGCTAGCGGACTACATCATCCAGCGTAATAAATAACGTCTTGAGCGATTGTGTAG-3′) (20 μM) which contains a homology arm (underlined, 41 bp) (see Note 1) chosen to match sequences in the upstream region of *dxs* gene and a priming sequence (20 bp) and 3′-T5(dxs) (5′-GGAGTCGACCAGTGCCAGGGTCGGGTATTTGGCAATATCAAAACTCATAGTTAATTTCTCCTCTTTAATG-3′) (20 μM) that contains a homology arm (underlined, 48 bp) chosen to match sequences in the downstream region of the *dxs* start codon and a priming sequence (22 bp). Primers for testing correct integration of Kan^R-P_{T5} in the front of *dxs* gene; T2 (5′-CGCCAGTCGCTGAAACAACTGGCTGA-3′) (20 μM), K1 (5′-CAGTCATAGCCGAATAGCCT-3′)(20μM),T5-1(5′-TAACCTATAAAAATAGGCGTATCACGAGGCCC-3′) (20 μM), and B2 (5′-TGGCAACAGTCGTAGCTCCTGGGTGG-3′) (20 μM).
2. Primers for PCR amplification of a cassette Kan^R-P_{T5} for the promoter replacement of *idi* gene; 5′-kan(idi) (5′-TCTGATGCGCAAGCTGAAGAAAAATGAGCATGGAGAATAATATGACGTCTTGAGCGATTGTGTAG-3′) (20 μM) which contains a homology arm (underlined, 45 bp) chosen to match a sequence in the upstream region of the *idi* gene and a priming sequence (20 bp) and 3′-T5(idi) (5′-TGGGAACTCCCTGTGCATTCAATAAAATGACGTGTTCCGTTTGCATAGTTAATTTCTCCTCTTTAATG-3′) (20 μM) which contains a homology arm (underlined, 46 bp) chosen to match sequences in the downstream region of the *idi* start codon and a priming sequence (22 bp). Primers for testing correct integration of Kan^R-P_{T5} in the front of *idi* gene; primer stock solution of T3 (5′-TCATGCTGACCTGGTGAAGGAATCC-3′) (20 μM) and B3 (5′-CAGCCAACTGGAGAACGCGAGATGT-3′) (20 μM).
3. Primers for PCR amplification of a cassette Kan^R-P_{T5} for the promoter replacement of *ispDispF* gene; 5′-kan(ispDF) (5′-GACGCGTCGAAGCGCGCACAGTCTGCGGGGCAAAACAATCGATAACGTCTTGAGCGATTGTGTAG-3′)

(20 μM) which contains a homology arm (underlined, 45 bp) chosen to match a sequence in the upstream region of the *isp-DispF* genes and a priming sequence (20 bp) and 3′-T5(ispDF) (5′-CGGCCGCCGGAACCACGGCGCAAACATCC|AAATGAGTGGTTGCCATAGTTAATTTCTCCTCTTTAATG-3′) (20 μM) that contains a homology arm (underlined, 46 bp) chosen to match sequences in the downstream region of the *ispDispF* start codon and a priming sequence (22 bp). Primers for testing correct integration of Kan^R-P_{T5} in the front of *idi* gene; primer stock solution of T8 (5′-GCCGAAATTGACGATCTCAATGGCG-3′) (20 μM) and B8 (5′-CCAGCAGCGCATGCACCGAGTGTTC-3′) (20 μM).

4. PCR amplification was performed by using one of the following PCR master mixes containing DNA Polymerase, nucleotides, and optimized PCR reaction buffer including $MgCl_2$; Phusion High-Fidelity PCR Master Mix (New England Biolabs, lpswich, MA), MasterAmp™ High Fidelity PCR PreMix (Epicentre, Madison, WI), or Platinum® PCR SuperMix (Invitrogen Co., Carlsbad, CA).
5. 5× stock solution of Tris-borate-EDTA (TBE) buffer for agarose gel electrophoresis (L): 54 g Tris base (FW = 121.14), 27.5 g boric acid (FW = 61.83), 20 mL of 0.5 M EDTA (pH8.0). Store at room temperature.
6. Precast agarose minigel "1% SeaKem Gold Agarose, 1× TBE buffer + ethidium bromide" (Cambrex Bio Science Rockland, Inc., Rockland, ME).
7. E-Gel® 1Kb Plus DNA Marker and BlueJuice™ Gel Loading Buffer (10×) (Invitrogen Co., Carlsbad, CA).
8. QIAquick Gel Extraction Kit™ and Qiaquick PCR Purification Kit (QIAGEN Inc. Valencia, CA).
9. Gene Pulser Electroporator and Gene Pulser 0.1 cm gap sterile electroporation cuvettes (Bio-Rad Inc., Hercules, CA).
10. 133 mM L-arabinose used for induction of λ-Red recombinase.

2.4. P1 Transduction

1. Stock solution and reagent for P1 transduction: $CaCl_2$ (1 M stock); $MgSO_4$ (1 M stock); NaCitrate pH 5.5 (1 M stock); chloroform.
2. Media for P1 transduction: LB broth; LB broth + chloramphenicol (25 μg/mL); LB broth + 100 mM $MgSO_4$ + 5 mM $CaCl_2$; LB broth + 10 mM NaCitrate; LB agar plate supplemented with 10 mM sodium citrate and kanamycin (25 μg/mL).
3. Bacteriophage ~10^7 $P1_{vir}$ lysate (13).

3. Methods

3.1. PCR to Generate a Promoter Replacement Cassette

1. Linear DNA fragment (1,641 bp) containing a cassette Kan^R-P_{T5} fusion was synthesized by PCR from plasmid pSUH5 with primer pairs, 5′-kan(dxs)/3′-T5(dxs), 5′-kan(idi)/3′-T5(idi), and 5′-kan(ispDF)/3′-T5(ispDF), separately, in which ~45–50 bp respective homology patches of the target integration sites were introduced at the ends of 5′-primer and 3′-primer. PCR reaction mixture: 25 μL MasterAmp™ High Fidelity PCR PreMix (2×), 1 μL pSUH5 (0.1 μg/μL), 1 μL 5′-primer (20 μM), 1 μL 3′-primer (20 μM), 12 μL sterilized dH_2O; PCR reaction conditions: 96°C×4 min, (95°C×30 s, 55°C×1 min, 72°C×3 min) 30 cycles, 72°C×5 min. After completing the PCR reactions, PCR products were purified using the QIAquick Gel Extraction Kit™.
2. Added 6 μL of loading buffer (10×) to 50 μL of each PCR sample, loaded PCR samples and standard 1 Kb Plus DNA Marker on a 1% SeaKem Gold Agarose gel, and electrophoresed at 95V for 40 min.
3. Visualized DNA bands using UV lightbox or gel imaging system and excised the amplified DNA fragment from the agarose gel with a clean, sharp scalpel (see Note 2).
4. Weighted gel slice in an eppendorf tube, and after adding 3 volumes of buffer QG to 1 volume of gel, incubated the tube at 50°C for 10 min (or until the gel slice has completely dissolved).
5. In order to bind DNA, applied the sample to the QIAquick gel extraction column and centrifuged for 1 min at maximum speed.
6. Applied 750 μL of buffer PE to the column for washing. After centrifuging for 1 min at maximum speed to remove the wash buffer, eluted PCR DNA products with 40 μL of distilled water (dH_2O) by standing sample for 1 min, and then by centrifuging for 1 min. Concentration of PCR DNA sample was about 0.2–0.5 μg/μL.

3.2. Chromosomal Promoter Replacement by Using λ-Red Recombination System

1. Chromosomal promoter replacement by using λ-Red recombination system is illustrated in Fig. 2. First, *E. coli* MC1061 strain carrying both a λ-Red recombinase expression plasmid pKD46 (amp^R) and a β-carotene biosynthesis expression plasmid pPCB15 (cam^R) was constructed by cotransformation with two plasmids. Transformants were selected on 100 μg/mL ampicillin and 25 μg/mL chloramphenicol LB plates at 30°C.
2. *E. coli* MC1061 cells carrying pKD46 and pPCB15 were grown in SOB medium with 100 μg/mL ampicillin, 25 μg/mL

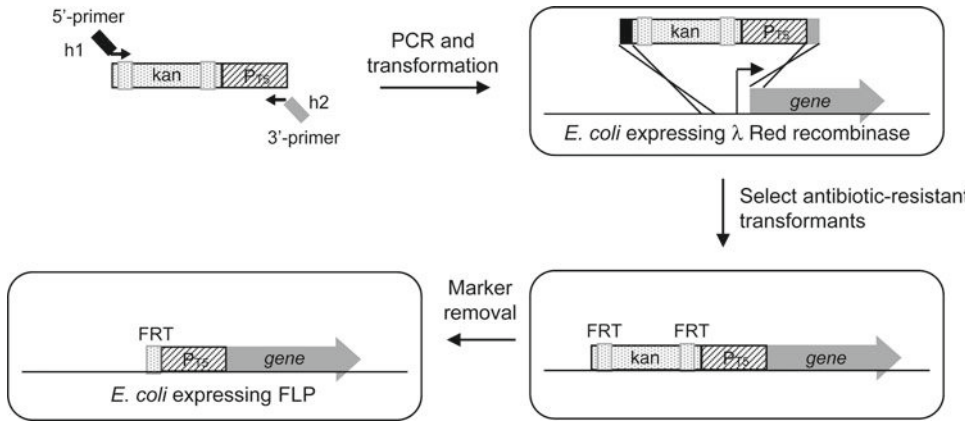


Fig. 2. Method for chromosomal promoter replacement by PCR-based homologous recombination mediated by λ red recombinase, and the removal of the antibiotic selection marker. (1) Linear DNA fragments of the kanamycin marker-P_{T5} promoter (kan-P_{T5}) was generated by PCR from pSUH5 with 5′- and 3′-primer pairs having 40–50 nt homology arms (h1 and h2); (2) *E. coli* MC1061 strain carrying both a λ red recombinase expression plasmid pKD46 (*amp*^{R}) and a reporter plasmid pPCB15 (*cam*^{R}) containing β-carotene biosynthesis genes was transformed with kan-P_{T5} PCR fragment by electroporation; (3) Double antibiotic-resistant recombinants were selected on LB plates containing both 25 μg/mL kanamycin and 25 μg/mL chloramphenicol; (4) The kanamycin gene flanked by FRT sites was removed by transformation with pCP20 (*amp*^{R}) containing Flp recombinase gene.

chloramphenicol, and 1 mM L-arabinose at 30°C to an OD_{600} of 0.5, followed by chilling on ice for 20 min.

3. Bacterial cells were centrifuged at 4,500 × *g* using a Sorvall® RT7 PLUS (Kendro Laboratory Products, Newton, CT) for 10 min at 4°C.
4. After decanting the supernatant, the cell pellet was resuspended in ice-cold water and centrifuged again. This was repeated twice and the cell pellet was resuspended in 1/100 volume of ice-cold 10% glycerol. The competent cells were stored in aliquots at –80°C.
5. Kan^R-P_{T5} PCR products (~1–2 μg) were mixed with 50 μL of the competent cells and pipetted into a precooled electroporation cuvette (0.1 cm) on ice.
6. Electroporation was performed by using a Bio-Rad Gene Pulser set at 1.8 kV, 25 μF with the pulse controller set at 200 Ω. SOC medium (1 mL) was added after electroporation. The cells were transferred into a culture tube and incubated with shaking (200 rpm) at 37°C for 1 h.
7. Approximately one-half of cells were spread on LB plates containing both 25 μg/mL kanamycin and 25 μg/mL chloramphenicol. The method yielded routinely ~100–200 double antibiotic-resistant recombinants per transformation with 1 μg PCR fragments.

8. After incubating the plate at 37°C overnight (see Note 3), successful recombinants having correctly integrated the kanamycin marker and P_{T5} promoter were easily identified based on the increased color intensity imparted by β-carotene accumulation in the cell.
9. The chromosomal integration of a cassette of Kan-P_{T5} in the front of *dxs*, *idi*, and *ispDispF* genes, separately, on *E. coli* chromosome was confirmed by colony PCR analysis. Figure 3 illustrates the PCR fragment analysis for kan-P_{T5}-dxs, kan-P_{T5}-idi, and kan-P_{T5}-ispDispF constructs and removal of the antibiotic selection marker.
10. Total volume for a colony PCR reaction in a PCR tube is 25 μL that contains 24 μL Platinum® PCR SuperMix, 0.5 μL of 5′-primer (20 μM), 0.5 μL of 3′-primer (20 μM). Different 5′- and 3′-primer pairs T2/K1, T5-1/B2, and T2/B2 were used to test the correct integration of Kan-P_{T5} in front of the *dxs* gene. Likewise, T3/K1, T5-1/B3, and T3/B3 were used to test for Kan-P_{T5}-*idi*, and T8/K1, T5-1/B8, and T8/B8 were used to test for Kan-P_{T5}-*ispDispF*. A small amount of colony was added using a fine pipette tip.

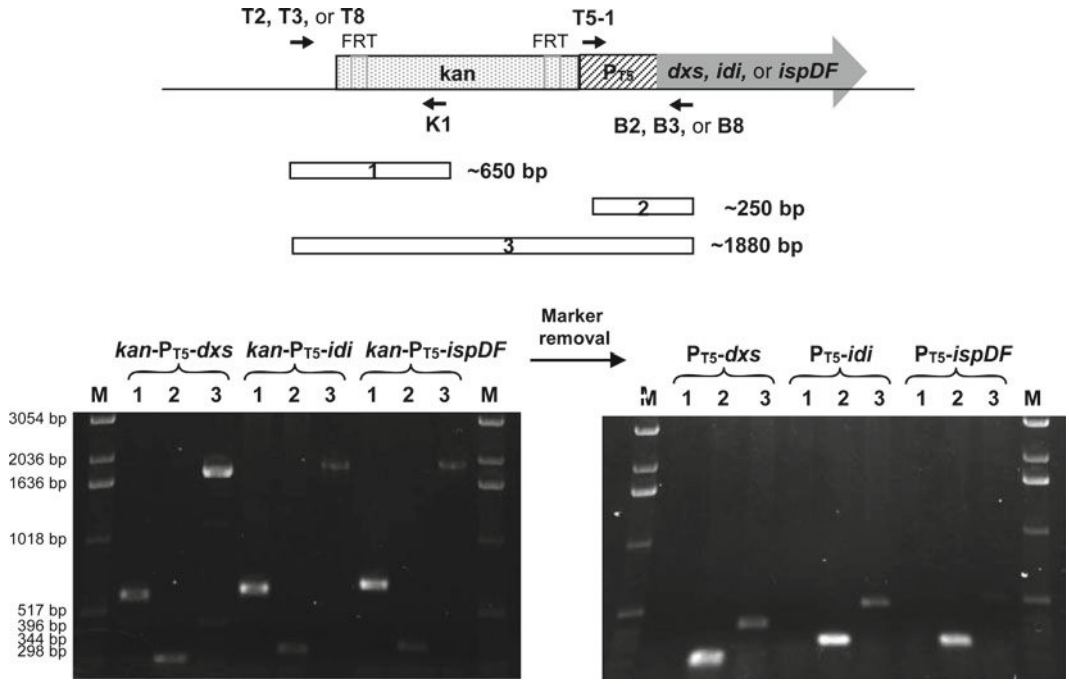


Fig. 3. PCR fragment analysis for kan-P_{T5}-dxs, kan-P_{T5}-idi, and kan-P_{T5}-ispDispF constructs and removal of the antibiotic selection marker. The integrated kanamycin marker and P_{T5} promoter in front of the *dxs, idi,* or *ispDispF* gene was verified by colony PCR analysis with different sets of test primers. The *bars* represent the expected sizes of PCR fragments amplified using corresponding primer sets (shown by *arrows*).

11. PCR amplification was carried out under the reaction conditions: 96°C × 4 min, (95°C × 30 s, 55°C × 1 min, 72°C × 2 min) 30 cycles, 72°C × 5 min.
12. Pipeted 5 μL PCR product to a 0.5-mL tube, add 0.5 μL of BlueJuice™ gel loading dye (10×) to each of these tubes, load 5.5 μL from the tube into a well in 1% SeaKem gold agarose gel and E-Gel® 1 Kb Plus DNA Marker on the each edge of the gel, and run a gel electrophoresis in TBE buffer at 95 V for 30 min. After the gel electrophoresis is completed, place it on a UV light box and take a photo.
13. Colony PCR analysis with test primers that were chosen to amplify regions located either in the vicinity of the integration region or in the kanamycin and phage P_{T5} promoter region revealed the expected sizes as shown in Fig. 3. The colony PCR results indicated the correct integration of the *PT5* promoter fragment in front of the chromosomal *dxs, idi,* and *ispDispF* genes, separately. Resulting strains are *E. coli* MC1061 kan-P_{T5}-dxs, *E. coli* MC1061 kan-P_{T5}-idi, and *E. coli* MC1061 kan-P_{T5}-ispDispF.

3.3. P1 Transduction to Stack Promoter Modifications

1. Serial P1 transductions were used to stack the P_{T5}-promoter modifications from donor strains *E. coli* MC1061 carrying kan-P_{T5}-dxs, kan-P_{T5}-idi, or kan-P_{T5}-ispDispF to a recipient strain MG1655 (see Note 4). Alternatively, *E. coli* MC1061 kan-P_{T5}-dxs, after marker removal, can be used as a starting host to stack first kan-P_{T5}-idi and then kan-P_{T5}-ispDispF by using either P1 transduction or the λ-Red recombination system (see Note 5).
2. P1 lysates of the donor strains *E. coli* MC1061 kan-P_{T5}-dxs, *E. coli* MC1061 kan-P_{T5}-idi, and *E. coli* MC1061 kan-P_{T5}-ispDispF were prepared separately by infecting a growing culture of bacteria with the P1 phage and allowing the cells to lyse. For P1 infection, each strain was inoculated in 4 mL LB medium with 25 μg/mL kanamycin, grown at 37°C overnight, and then subcultured with 1:100 dilution of an overnight culture in 10 mL LB medium containing 5 mM $CaCl_2$ in a 125-mL flask (see Note 6).
3. After 20–30 min of growth at 37°C, ~10^7 $P1_{vir}$ phages were added. The cell-phage mixture was aerated for 2–3 h at 37°C until lysed.
4. Several drops of chloroform were added, and the mixture vortexed for 30 s and incubated for an additional 30 min at room temperature.
5. The mixture was then centrifuged for 10 min at 4,500 × *g*, and the supernatant transferred into a new tube to which several drops of chloroform were added. The lysates were stored at 4°C.

6. In parallel, the recipient strain *E. coli* MG1655 was transformed with a β-carotene biosynthesis expression plasmid pPCB15 (cam^R). Transformants were selected on 25 μg/mL chloramphenicol LB plates at 30°C.
7. P1 lysate made from the donor strain *E. coli* MC1061 kan-P_{T5}-dxs was transduced into the recipient strain, *E. coli* MG1655 containing a β-carotene biosynthesis expression plasmid pPCB15 (cam^R). *E. coli* MG1655/pPCB15 recipient cells were grown to mid-log phase ($1–2 \times 10^8$ cells/mL) in 4 mL LB medium with 25 μg/mL chloramphenicol at 37°C.
8. Cells were spun down for 10 min at $4{,}500 \times g$ and resuspended in 2 mL of 10 mM $MgSO_4$ and 5 mM $CaCl_2$.
9. Recipient cells (100 μL) in a 1.5-mL eppendorf tube were mixed with 0, 10, 50, or 100 μL of P1 lysate stock (10^6 pfu/μL) made from the *E. coli* kan-P_{T5}-dxs strain and incubated at 30°C for 30 min. A control mixture of 100 μL P1 lysate made from *E. coli* kan-P_{T5}-dxs with 100 μL LB + 10 mM $MgSO_4$ + 5 mM $CaCl_2$ was also prepared in a 1.5-mL eppendorf tube.
10. The recipient cell-lysate mixture and control mixture were spun down at 6,500 rpm for 30 s, resuspended in 100 μL of LB medium with 10 mM of sodium citrate, and incubated at 37°C for 1 h.
11. Cells were plated on LB plates containing both 25 μg/mL kanamycin and 25 μg/mL of chloramphenicol in order to select for antibiotic-resistant transductants and incubated at 37°C for 1 or 2 days.
12. Single transductant that was pigmented with the characteristic yellow β-carotene color were selected, streaked and incubated overnight at 37°C. The resulting strain was *E. coli* MG1655 kan-P_{T5}-dxs.
13. To eliminate the kanamycin selectable marker from the chromosome, an FLP recombinase expression plasmid pCP20 (amp^R), which has a temperature-sensitive replication of origin, was transformed into *E. coli* MG1655 kan-P_{T5}-dxs by electroporation.
14. Transformant cells were spread onto LB agar containing 100 μg/mL ampicillin and 25 μg/mL chloramphenicol LB plates, and grown at 30°C for 1 day.
15. Colonies were picked and streaked on 25 μg/mL chloramphenicol LB plates without ampicillin antibiotics and incubated at 43°C overnight. Plasmid pCP20 has a temperature-sensitive origin of replication and was cured from the host cells by culturing cells at 43°C.
16. The colonies were tested for ampicillin and kanamycin sensitivity to test loss of pCP20 and the kanamycin selectable marker by

streaking colonies on 100 μg/mL ampicillin LB plate and 25 μg/mL kanamycin LB plate, respectively.

17. Elimination of the kanamycin selectable marker from the chromosome of *E. coli* MG1655 kan-P_{T5}-dxs was confirmed by PCR analysis (Fig. 3). Colony PCR was carried out under the reaction conditions: 96°C × 4 min, (95°C × 30 s, 55°C × 1 min, 72°C × 2 min) 30 cycles, 72°C × 5 min. The selected colonies were resuspended in 25 μL of PCR reaction mixture that contained 24 μL Platinum® PCR SuperMix using a sterilized tooth-pick, and 1 μL of different combinations of specific 5′-primer (0.5 μL of 20 μM)/3′-primer (0.5 μL of 20 μM) pairs, T2/K1, T5-1/B2, and T2/B2. The PCR results indicated elimination of the kanamycin selectable marker from the *E. coli* chromosome. The presence of the P_{T5} promoter fragment upstream of the *dxs* coding sequence was confirmed based on the production of a PCR product of the expected size (229 bp). The resulting strain was *E. coli* MG1655 P_{T5}-dxs/pPCB15.
18. To stack kan-P_{T5}-idi upon *E. coli* MG1655 P_{T5}-dxs/pPCB15, P1 transduction with P1 lysates of the donor strains *E. coli* MC1061 kan-P_{T5}-idi and kanamycin marker removal were carried out as described above, yielding *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi/pPCB15.
19. Repeatedly, kan-P_{T5}-ispDispF was stacked upon *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi/pPCB15 by P1 transduction with P1 lysates of the donor strains *E. coli* MC1061 kan-P_{T5}-ispDispF and kanamycin marker removal as described above, yielding the high isoprenoid flux strain *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi P_{T5}-ispDispF/pPCB15.

3.4. Measurement of Carotenoid Production

1. The quantitative analysis of β-carotene production was achieved by measuring the spectra of acetone-extracted β-carotene's characteristic λ_{max} peaks at 453 nm (14) (see Note 7). *E. coli* MG1655 P_{T5}-dxs/pPCB15*E, E. coli* MG1655 P_{T5}-dxs P_{T5}-idi/pPCB15, *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi P_{T5}-ispDispF/pPCB15, and *E. coli* MG1655/pPCB15 control strains were grown in 5 mL LB containing chloramphenicol (25 μg/mL) at 37°C for 24 h.
2. Cell density of each strain was determined by measuring the absorption spectrum at 600 nm by using an Ultrospec 3000 spectrophotomer (Amersham Biosciences, Piscataway, NJ).
3. Cells were harvested by centrifugation at 4,500 × *g* for 10 min.
4. β-carotene pigments were extracted from the cell pellet by resuspending in 300 μL of acetone and then homogenizing cells with glass beads in a Bead-Beater (Biospec Products, Bartlesville, OK) for 30 s.

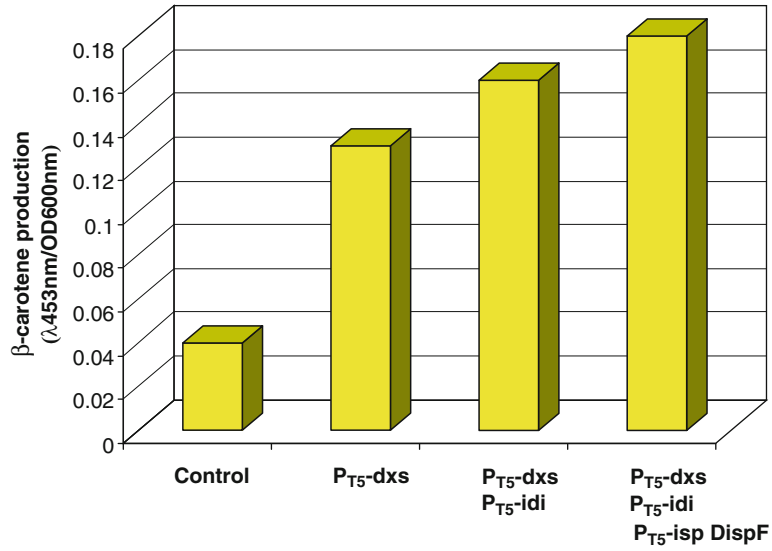


Fig. 4. β-carotene production ($\lambda_{453nm}/OD_{600nm}$) of *E. coli* MG1655/pPCB15 (control), *E. coli* MG1655 P_{T5}-dxs/pPCB15*E, E. coli* MG1655 P_{T5}-dxs P_{T5}-idi/pPCB15, and *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi P_{T5}-ispDispF/pPCB15 strains.

5. After centrifuging the sample at 16,000 × *g* for 1 min, the absorption spectrum of the acetone layer containing β-carotene was measured at λ 450 nm using an Ultrospec 3000 spectrophotometer (Amersham Biosciences, Piscataway, NJ).
6. The value of λ 453 nm of each sample was normalized to the cell density (OD_{600nm}). Figure 4 illustrates the β-carotene production ($\lambda_{453nm}/OD_{600nm}$) of the engineered and control strains. The β-carotene production of the high isoprenoid flux strain *E. coli* MG1655 P_{T5}-dxs P_{T5}-idi P_{T5}-ispDispF/pPCB15 increased approximately 4.5-fold, compared to the *E. coli* MG1655/pPCB15 control strain.

4. Notes

1. The λ-Red recombination system facilitates efficient recombination between linear recombination elements and the chromosome over relatively small regions of homology (36–50 nt).
2. To avoid DNA damage with UV light, always use a long wavelength UV (360 nm) light-box during excision of the gel slice. If only a short-wavelength UV light-box is available, minimize the UV exposure to a few seconds.
3. The λ Red recombinase expression plasmid pKD46 has a temperature-sensitive replication origin and can be eliminated at 37°C.

4. P1 transduction can be done either by the conventional P1 transduction that can move only one genetic trait at a time from one host to another host or in a parallel combinatorial fashion using pooled mixtures of bacteriophage P1 (9).

   We previously developed the combinatorial P1 transduction for transferring multiple genetic traits into an *E. coli* host in a parallel fashion using pooled mixtures of P1 lysates followed by easy yellow color screening for selecting higher production of β-carotene than parental strain. For example, to identify next rate-limiting enzyme beside the *dxs* gene in the isoprenoid pathway, the recipient MG1655 P_{T5}-*dxs* (pPCB15) was infected with the pooled P1 lysate from MC1061 strains containing different P_{T5}-isoprenoid genes (kan-P_{T5}-ispC, kan-P_{T5}-ispDF, kan-P_{T5}-ispE, kan-P_{T5}-ispG, kan-P_{T5}-ispH, kan-P_{T5}-idi, or kan-P_{T5}-ispB) and then kanamycin-resistant transductants were visually screened for increased yellow color. Among the transductants, the colonies carrying *kan-*P_{T5}*-idi* exhibited the deepest yellow color, suggesting that the isopentenyl pyrophosphate isomerase encoded by *idi* gene is a rate-controlling enzyme in the isoprenoid flux. The repeated P1 transduction with the pooled P1 lysate enables us to identify sequential bottlenecks in isoprenoid flux.

5. This approach can be done by repeating a cycle of (1) a λ-Red recombinase-mediated promoter replacement by transformation with a PCR-generated promoter replacement cassette, (2) a marker removal by a temporal transformation with a Flp expressing plasmid (pCP20), (3) transformation with λ-Red recombinase expressing plasmid (pKD46) and a reporter plasmid (pPCB15), and (4) preparation of λ-Red recombinase expressing competent cells.

6. Calcium facilitates the physical interaction between the phage and the bacterial cell.

7. The β-carotene production also was qualitatively determined by streaking the strains on LB agar plates containing chloramphenicol (25 μg/mL) and incubating at 37°C for 2–3 days. The high isoprenoid flux strain exhibited deeper yellow color than the control strain, indicating higher production of β-carotene.

Acknowledgments

The author thanks Professor Barry L. Wanner for providing λ-Red strains and plasmids used in this work and Wendy Suh for critical reading of the manuscript.

References

1. Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, P., and Böger, P. (1999) The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. *Trends Biotechnol.* **17**, 233–237.
2. Schmidt-Dannert, C., Lee, P. C., and Mijts, B. N. (2006) Creating carotenoid diversity in E. coli cells using combinatorial and directed evolution strategies. *Phytochemistry Reviews* **5**, 67–74.
3. Rohdich, F., Zepeck, F., Adam, P., Hecht, S., Kaiser, J., Laupitz, R., Gräwert, T., Amslinger, S., Eisenreich, W., Bacher, A., and Arigoni, D. (2003) The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: Studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proc. Natl. Acad. Sci. USA.* **100**, 1586–1591.
4. Ruther, A., Misawa, N., Boger, P., and Sandmann, G. (1997) Production of zeaxanthin in *Escherichia coli* transformation with different carotenogenic plasmids. *Appl. Microbiol. Biotechnol.* **48**, 162–167.
5. Matthews, P. D. and Wurtzel, E. T. (2000) Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl. Microbiol. Biotechnol.* **53**, 396–400.
6. Albrecht, M., Misawa, N., and Sandmann, G. (1999) Metabolic engineering of the terpenoid biosynthetic pathway of *Escherichia coli* for production of the carotenoids β-carotene and zeaxanthin. *Biotechnol. Letters* **21**, 791–795.
7. Noack, D., Roth, M., Geuther, R., Müller, G., Undisz, K., Hoffmeier., C., and Gáspár, S. (1981) Maintenance and genetic stability of vector plasmids pBR322 and pBR325 in *Escherichia coli* K12 strains grown in a chemostat. *Mol. Gen. Genet.* **184**, 121–124.
8. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990) Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng.* **35**, 668–681.
9. Yuan, L. Z., Rouvière, P. E., LaRossa, R. A., and Suh, W. (2006) Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* **8**, 79–90.
10. Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA.* **97**, 6640–6645.
11. Casadaban, M. J. and Cohen, S. N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**, 179–207.
12. Bachmann, B. J. (1996) Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. *In* "*Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology" (F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, Eds.), 2nd ed., pp. 2460–2488, ASM Press, Washington, DC.
13. Miller, J. H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Abe, T., Abboud, J.-L. M., Belio, F., Bosch, E., Garcia, J. I., Mayoral, J. A., Notario, R., Ortega, J., Roses, M. (1998) Empirical treatment of solvent-solute interactions: Medium effects on the electronic absorption spectrum of β-carotene. *J. Phy. Org. Chem.* **11**, 193–200.

Chapter 5

Mutagenic Inverted Repeats Assisted Genome Engineering (MIRAGE) in *Saccharomyces cerevisiae*: Deletion of *gal7*

Nikhil U. Nair and Huimin Zhao

Abstract

MIRAGE is a unique in vivo genome editing technique that exploits the inherent instability of inverted repeats (palindromes) in the *Saccharomyces cerevisiae* chromosome. As a technique able to quickly create deletions as well as precise point mutations, it is valuable in applications that require creation of designer strains of this yeast. In particular, it has various potential applications in metabolic engineering, systems biology, synthetic biology, and molecular genetics.

Key words: In vivo mutagenesis, Homologous recombination, Inverted repeats, Synthetic biology, Metabolic engineering

1. Introduction

The importance of *S. cerevisiae* as a model eukaryotic organism and a favorite industrial biocatalyst has necessitated the development of several strategies to modify its genome. In addition, recent advances in top–down genomics that demonstrate the feasibility of constructing and transplanting entire recombinant genomes have further necessitated genome engineering tools in this yeast (1, 2).

Although plasmid-based pop-in/pop-out methods work quite well in yeast (3), most current methods for precise chromosomal mutagenesis in yeast either require several laborious steps (2, 4–6) or suffer from high false positive rates (7). In order to address the limitations of the above methods, we recently developed a new method called Mutagenic Inverted Repeats Assisted Genome Engineering (MIRAGE). This method relies on the instability conferred by inverted repeats/palindromes in eukaryotic genomes (8).

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_5,

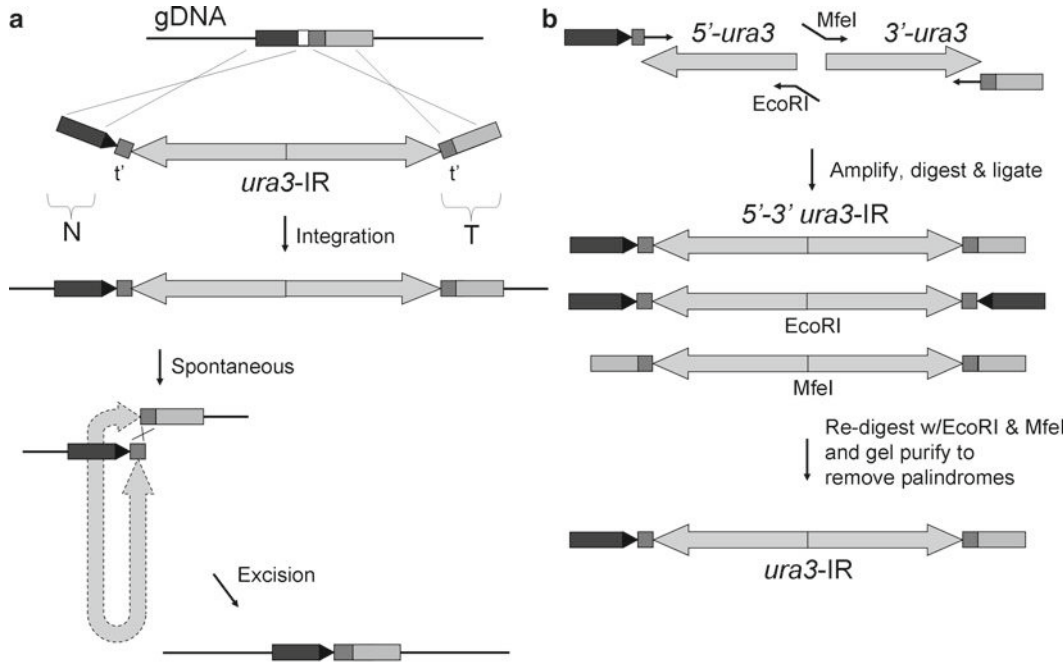


Fig. 1. Schematic of MIRAGE. (**a**) 45 bp flanking ends of neck (N) and tail (T) target the mutagenesis cassette into the yeast chromosome (gDNA), simultaneously replacing the undesired allele (*open square*) with the desired mutation (▶). The *ura3*-IR confers transformants with uracil prototrophy. Spontaneous and precise excision of the IR occurs between the two t' sequences, leaving behind only the desired mutation. Since the mechanism of MIRAGE excision is not known, the IR is drawn with hyphenated lines. Omitting the mutation in the "neck" primer (▶) will lead to deletion, rather than replacement. (**b**) Creation of the mutagenesis cassette requires preparation of each *ura3* fragment separately. After PCR amplification of the 5′- and 3′-*ura3* fragments and digestion with EcoRI and MfeI, the two halves are ligated together. Ligation of compatible EcoRI and MfeI ends abolishes both restriction sites. Redigestion with EcoRI and MfeI, followed by gel purification removes palindromic ligation products leaving behind the hybrid 5′-3′ *ura3*-IR cassette for MIRAGE.

Inverted repeats (IR) are able to form stable hairpin structures during chromosomal replication and pose as barriers to DNA polymerases. Resultantly, IR are hotspots for recombination, the frequency of which is dependent on the distance between the two copies of DNA that constitute the palindrome (9, 10). The hyperrecombinogenic nature of the IR is used to efficiently excise it from the chromosome without the need for any exogenous recombinases such as Flp or Cre. The utility and precision of this method is best demonstrated via an example, and here we use the deletion of *gal7* as an example to illustrate the experimental protocol of MIRAGE.

The concept and design of MIRAGE are shown in Fig. 1. Two copies of *ura3* (the 5′ fragment and 3′ fragment) are amplified using polymerase chain reaction (PCR), each with 45 bp of homology either upstream (N) or downstream (T) of *gal7*. The sequence to be introduced in the chromosome is included between N and t', which is just the first 25 bp of the T sequence, in the 5′ fragment (see Note 1). After amplification, the two fragments are ligated together and subsequently separated from exact palindromes

(3′-3′ or 5′-5′ products) to isolate hybrid 3′-5′ cassettes. These are then transformed into yeast and selected for Ura^+ phenotype. After verification of the integration event by PCR, the IR excision event is selected for by resistance to 5-fluoroorotic acid ($5FOA^R$). The excisants are finally verified by PCR and direct sequencing.

2. Materials

All buffers and media are made in distilled and deionized water with a resistivity of 18.2 mΩ·cm, and sterilized either by heat (autoclaving) or filtration through a 0.22-μm membrane. All reagents were stored at room temperature with the exception of 5-fluoroorotic acid (5FOA), and salmon sperm single-stranded DNA (ssDNA) stock solutions, which were maintained at −20°C.

2.1. DNA Manipulation

1. EDTA (ethylenediaminetetraacetic acid) stock solution (0.5 M, pH 8.0): A 500-mL stock of EDTA solution is prepared by vigorously mixing 93.05 g EDTA disodium salt (MW = 372.2 g/mol) in water and simultaneously adjusting pH to 8.0 using sodium hydroxide (NaOH) pellets. EDTA will not completely dissolve at low pH and therefore addition of NaOH must be concurrent.
2. TAE (tris-acetate-EDTA) stock solution (50×): Stock solution of TAE is prepared by completely dissolving 242 g Tris-base (MW = 121.14 g/mol), 57.1 mL glacial acetic acid (MW = 60.05 g/mol), and 100 mL 0.5 M EDTA solution pH 8.0 in water up to a final volume of 1,000 mL. There is no need to adjust pH, which should be at ~8.5.
3. TAE working solution (1×): Dilute the stock solution 50×, i.e., 200 mL in 9.8 L water.
4. Agarose gel (1.0%): 1× TAE gel is prepared by adding 0.5 g of agarose to 50 mL 1× TAE solution and microwaving until all the powder has been dissolved. Then add 2.5 μL 10 mg/mL ethidium bromide solution. Pour 30–50 mL into an agrose gel rack with an appropriate comb.

2.2. Growth and Transformation

1. Yeast strain: The lab strain HZ848 (*MATα ade2-1 Δura3 his3-11,15 trp1-1 leu2-3,112 can1-100*) was used for all MIRAGE work. Since we used the *S. cerevisiae ura3* as the selectable marker, it is necessary that the strain has a complete deletion of the chromosomal *ura3*.
2. 2×YPAD medium: Dissolve 12 g yeast extract, 24 g peptone, 24 g dextrose, and 120 mg adenine hemisulfate in water and sterilize in an autoclave.

3. SC-U medium: Dissolve 3 g ammonium sulfate, 1 g yeast nitrogen base without amino acids and ammonium sulfate, 0.5 g complete synthetic medium minus uracil (CSM-Ura), 60 mg adenine hemisulfate, and 12 g dextrose in 600 mL water. Sterilize by filtration. For solid media, make a 2× solution of above by dissolving components in 300 mL water and mix with an autoclaved solution of agar noble (9 g in 300 mL water).
4. SC+5FOA medium: Dissolve 3 g ammonium sulfate, 1 g yeast nitrogen base without amino acids and ammonium sulfate, 0.5 g complete synthetic medium minus uracil, 60 mg adenine hemisulfate, 30 mg uracil, 180 mg 5FOA, and 12 g dextrose in 600 mL water. Sterilize by filtration. For solid media, make a 2× solution of above by dissolving components in 300 mL water and mix with an autoclaved solution of agar noble (9 g in 300 mL water). Cover in foil and keep away from light.
5. Lithium acetate solutions: To prepare a 50 mL 1.0 M solution, dissolve 3.3 g lithium acetate (MW = 65.99 g/mol) in water and sterilize by filtration. A 100-mM solution is then prepared by mixing 5 mL of the 1.0 M solution with 45 mL sterile water.
6. 50% PEG (polyethylene glycol): Mix 50 g PEG (MW = 3,350 g/mol) in 30 mL water. The solution will need to be heated to ~60°C with constant stirring to ensure quick dissolution of PEG. Once all the PEG is dissolved, adjust volume to 100 mL and sterilize by filtration.
7. ssDNA (single strand carrier DNA): Dissolve 200 mg salmon sperm single strand carrier DNA in 100 mL TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0). Aliquot into 1.5-mL tubes and store at −20°C.

3. Methods

3.1. Cassette Preparation

1. PCR-amplify the two copies of *ura3* individually using pRS406 as template and primers listed in Table 1 (see Note 2). Products from each step are shown in Fig. 2.

   Set up reaction for each fragment as follows:

   40 μL 5× GoTaq reaction buffer.

   20 μL dNTP mix (2 mM each).

   12 μL 25 mM $MgCl_2$.

   10 μL forward primer F delGAL7 ura3-IR LN 25 or F/R ura3-IR MfeI 3 (10 pmol/μL).

Table 1
Sequences of all the primers used to delete *gal7* using MIRAGE

Purpose	Name	Sequence (5′ to 3′)
Primers used for all *ura3*-IR-based mutagenesis	NN R/F ura3-IR EcoRI 3	atc GAATTC tttttttattctttttttgatttcggtttctttg
	NN F/R ura3-IR MfeI 3	atc CAATTG tttttttattctttttttgatttcggtttctttg
Primers used to delete *gal7*	NN F delGAL7 ura3-IR LN25	gagcatcaacatgataaaaaaaaacagttgaata tccctcaaaa *tttgaaacgagtttcccataccaa* cctgatgcggtattttctcc
	NN R delGAL7 ura3-IR LN	aggagcctgatggataccca *ttgagtatgg gaaactcgtttcaaa* cctgatgcggtattttctccttacg
Verification primers	NN R/F ura3t IRchk	agatgcgtaaggagaaaataccgcatcagg
	NN F del conf GAL7p	tccttttggaaagctatacttcggagcactg
	NN R del conf GAL7t	tacagtctttgtagataatgaatctgaccatc

Capitals indicate restriction sites, underlines indicate sequences that anneal to template, and italics indicate t' sequence

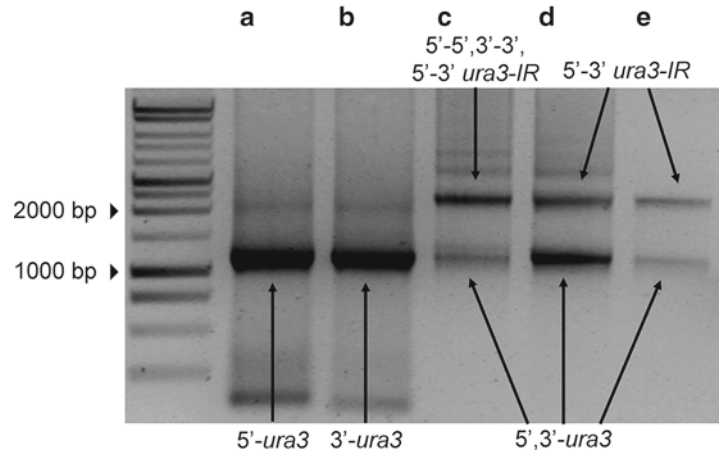


Fig. 2. Agarose gel from each step of the cassette preparation step. (**a**, **b**) The 5′- and 3′-*ura3* fragments are ~1.1 kb long. (**c**) After ligation, the three different types of *ura3*-IR fragments (5′-5′, 5′-3′, and 5′-3′) are ~2.2 kb. The ligation process is incomplete, even after overnight ligation at 16°C. Longer multimers at ~3.3 and ~4.4 kb can also be seen as minor products. (**d**) After redigestion with EcoRI and MfeI, only the 5′-3′ *ura3*-IR remains undigested. The monomeric *ura3* fragments are also enriched. (**e**) Gel purification of the 5′-3′ *ura3*-IR removes majority of the multimeric and monomeric fragments. Although monomeric fragments remain even after purification, they do not seem to cause any deleterious effects.

10 μL reverse primer R/F ura3-IR EcoRI 3 or R delGAL7 ura3-IR LN 25 (10 pmol/μL).

1 μL pRS406 template (20–50 ng).

1 μL Phusion DNA polymerase (1 U).

1 μL GoTaq DNA polymerase (1 U).

105 μL ddH_2O.

200 μL total volume.

Thermocycler program:

Fully denature DNA at 94°C for 2 min, followed by 45 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s. Follow with a final extension at 72°C for 5 min, and then cool to 4°C to stop reaction.

2. Purify the DNA product from each reaction separately using QIAquick PCR Purification Kit. Elute DNA from each column with two 51 μL ddH_2O elution to obtain each fragment in a final volume of 102 μL (see Note 3).
3. Double digest the 5′ fragment with EcoRI and DpnI, and the 3′ fragment with MfeI and DpnI for 3 h for 37°C as follows (see Notes 4 and 5):

   102 μL PCR purified DNA (3–4 μg) (see Note 6).

   12 μL NEBuffer 4 (10×).

   4 μL EcoRI or MfeI (80 U).

   2 μL DpnI (40 U).

   120 μL total volume.
4. Gel purify each fragment separately using a 1% agarose gel (120 V for 20 min) and QIAquick Gel Extraction Kit. Elute into 53 μL ddH_2O each.
5. Ligate both fragments together at 16°C for 12–20 h (see Note 7). The reaction mix is as follows:

   53 μL 5′ *ura3* fragment (EcoRI digested).

   53 μL 3′ *ura3* fragment (MfeI digested).

   12 μL T4 DNA ligase buffer (10×).

   2 μL T4 DNA ligase (800 U).

   120 μL total volume.
6. PCR purify the ligation mix using QIAquick PCR Purification Kit and again, elute twice with 51 μL ddH_2O to obtain a final volume of 102 μL.
7. Double digest the ligated fragments with EcoRI and MfeI to cleave perfect palindromes.

   102 μL PCR purified DNA (2–3 μg).

   12 μL NEBuffer 4 (10×).

3 μL EcoRI or MfeI (60 U).

3 μL MfeI (60 U).

120 μL total volume.

8. Gel purify the 2.2 kb fragment using a 1% agarose gel (100 V for 60 min) and QIAquick Gel Extraction Kit (see Note 8). Elute into 35 μL ddH_2O.
9. Quantify the DNA concentration using NanoDrop.

3.2. Transformation

1. Inoculate a single colony of HZ848 into 2×YPAD and culture overnight at 30°C, 250 rpm (see Notes 9–13).
2. Seed 15 mL of fresh 2×YPAD with 450 μL of saturated overnight culture and let cells grow for 2–4 doublings (4 h) at 30°C, 250 rpm.
3. Harvest cells by transferring culture to a 50-mL tube and centrifuging at 3,200×*g*.
4. Resuspend the cell pellet in 15 mL sterile, room temperature ddH_2O and centrifuge again. Discard supernatant.
5. Repeat wash step once more.
6. Resuspend the pellet in 1 mL 100 mM LiAc solution and aliquot cells equally into two 1.5-mL tubes. Centrifuge the cells for 15 s at 6,700×*g* and then discard supernatant.
7. Repeat wash once again. The pellets are now ready for transformation.
8. Place a 1.5-mL tube of frozen ssDNA solution with a floating waterbath rack directly into boiling water for 5–10 min. Transfer tube from boiling water directly onto ice.
9. Add the following to one of the cell pellets in a 1.5-mL centrifuge tube, in the given order:

   240 μL 50% PEG.

   36 μL 1.0 M LiAc solution.

   50 μL ssDNA solution.

   10–34 μL MIRAGE cassette (100–500 ng).

   0–24 μL ddH_2O.

   360 μL total volume.

   Use the second tube of cell pellet as negative control, i.e., add no MIRAGE cassette.
10. Vortex the tubes for 60 s to ensure complete mixing.
11. Place the tubes in a 42°C waterbath for 40 min.
12. Centrifuge the mixture at 3,200×*g* for 30 s and discard the supernatant.
13. Resuspend the pellet in 1 mL 2×YPAD each and transfer to culture tubes.

14. Shake tubes at 30°C, 250 rpm for 3–16 h. Longer recovery time ensures higher transformation efficiency (see Note 14).
15. Transfer 1-mL cultures to 1.5-mL centrifuge tubes and spin at 3,200 × *g* for 2 min.
16. Discard supernatant and resuspend cells in 1 mL SC-U.
17. Centrifuge the tubes again at 3,200 × *g* for 2 min.
18. Discard the supernatant and resuspend to a final volume of 300 μL in SC-U.
19. Spread evenly on three plates of SC-U each and incubate at 30°C for 3 days.

3.3. Verification of Integrants

1. Pick 4–6 colonies at random, restreak onto fresh SC-U plates, and also inoculate them into 2 mL SC-U liquid medium.
2. Incubate plates at 30°C for 2 days and culture in 30°C, 250 rpm shaker for 2 days.
3. Isolate genomic DNA (gDNA) from cultures using Wizard Genomic DNA Purification Kit.
4. Using gDNA as template and primers listed in Table 1 to carry out two diagnostic PCRs for each clone as follows:

   2.5 μL 5× GoTaq reaction buffer.

   1.25 μL dNTP mix (2 mM each).

   0.75 μL 25 mM $MgCl_2$.

   0.06 μL forward primer F del conf GAL7p or F/R ura3t IRchk (100 pmol/μL).

   0.06 μL reverse primer F/R ura3t IRchk or R del conf GAL7t (100 pmol/μL).

   0.5 μL gDNA template.

   0.06 μL GoTaq DNA polymerase (0.06 U).

   7.32 μL ddH_2O.

   12.5 μL total volume.

   Thermocycler program:

   Fully denature DNA at 94°C for 2 min, followed by 25 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Follow with a final extension at 72°C for 5 min, and then cool to 4°C to stop reaction.

   Expected band sizes: 350 bp, 275 bp (Fig. 3a).

3.4. Excision of ura3-IR

1. From the restreaks of clones with confirmed correct integration, scrape up some cells using a sterile toothpick and resuspend in 50 μL ddH_2O (see Note 15). Allow all the residual media to be depleted by incubating the cells at room temperature for 3–4 h.

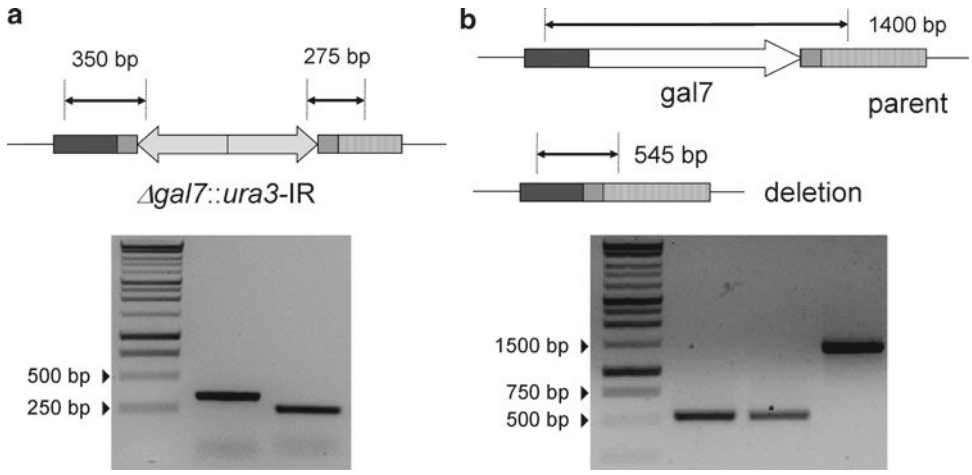


Fig. 3. (**a**) Expected PCR bands for the insertion step. A correctly inserted cassette will have bands at 350 bp (F del conf GAL7p and R ura3t IRchk) and 275 bp (R del conf GAL7t and R ura3t IRchk). (**b**) After excision (two independent clones shown), the PCR product amplified using F del conf GAL7p and R del conf GAL7t is significantly smaller (545 bp) than the parent (1,400 bp).

2. Spot the cells on the edge of an SC+5FOA plate and streak to ensure formation of distinct, single colonies (see Notes 16 and 17).
3. Incubate plates at 30°C for 3 days, away from light.

3.5. Verification of Excisants

1. Pick 1–2 colonies from each plate and inoculate 2 mL SC+5FOA cultures (see Note 18).
2. Incubate in a 30°C, 250 rpm shaker for 2 days.
3. Isolate genomic DNA (gDNA) from cultures using Wizard Genomic DNA Purification Kit.
4. Using gDNA as template and primers listed in Table 1 to carry a diagnostic PCR for each clone as follows:

   20 μL 5× HF reaction buffer.

   10 μL dNTP mix (2 mM each).

   0.5 μL forward primer F del conf GAL7p (100 pmol/μL).

   0.5 μL reverse primer R del conf GAL7t (100 pmol/μL).

   2.0 μL gDNA template.

   0.5 μL Phusion DNA polymerase (0.5 U).

   66.5 μL ddH_2O.

   100 μL total volume.

   Thermocycler program:

   Fully denature DNA at 98°C for 30 s, followed by 20 cycles of 98°C for 10 s, 55°C for 30 s, and 72°C for 30 s. Follow with a final extension at 72°C for 5 min, and then cool to 4°C to stop reaction.

   Expected band size: 545 bp (Fig. 3b).

5. Gel purify the 545 bp fragment using a 1% agarose gel (120 V for 20 min) and QIAquick Gel Extraction Kit. Elute into 35 μL ddH_2O.
6. Quantify the DNA concentration using NanoDrop.
7. Sequence PCR product using either of the PCR primers.

4. Notes

1. The t' sequence of 25 bp was optimized to ensure low frequency undesired integration events (double crossover at t' and T) compared to the desired event (double crossover at N and T). In addition, this length ensures precise excision of *ura3*-IR between t' and T.
2. Although the *S. cerevisiae ura3* gene was used as a selection marker here, a heterologous *ura3* or selectable/counterselectable marker can be used. The use of heterologous cassettes can circumvent the requirement for *ura3* deletion strain.
3. Multiple elutions from QIAquick Spin Columns ensure maximum possible recovery.
4. Any pair of restriction enzymes with mutually compatible cohesive ends can be used if ligation abolished both restriction sites. Other possibilities are AvrII/NheI, BspHI/NcoI, and BglII/BamHI.
5. The DpnI digestion is not required if the plasmid used as template for PCR does not contain any homology to the chromosome.
6. Due to the multiple DNA purification steps, it is necessary to ensure that the initial PCR yields >2 μg of each fragment.
7. Although ligation can be performed at room temperature for several hours, overnight ligation at 16°C is often more efficient.
8. The choice of selection markers is not limited by the maximum allowable length of the IR. Indeed, IR as long as 4.4 kb is as efficient for MIRAGE as the 2.2 kb cassette used here.
9. To ensure higher transformation efficiency, 2×YPAD is suggested, although YPAD can also be used.
10. Competent cells must be prepared fresh every time and cannot be stored for a long term.
11. Supplementation of adenine in media is not required for Ade^+ host strains.
12. For strains that grow poorly on dextrose or cannot at all, another carbon source such as galactose or ethanol can be

substituted. However, growth times must be adjusted according to doubling times.

13. Kanamycin (50 μg/mL) and tetracycline (10 μg/mL) may be added to any medium to minimize bacterial contamination without any deleterious effects to yeast.
14. Transformation efficiency can be increased by increasing recovery time. For example, if only 100 ng of the MIRAGE cassette is transformed, 16 h recovery is recommended. However, if 500 ng of DNA is transformed instead, recovery times as short as 3 h may suffice. Efficiencies range from 10 to 100 colonies/μg DNA transformed.
15. $5FOA^R$ colonies typically appear at a frequency of 10^{-5} to 10^{-6}.
16. Although the recommended working concentration of 5FOA is 1 g/L, 300 mg/L was sufficient for complete growth inhibition of Ura^+ strains.
17. Ensure the pH of SC+5FOA media is <3.5.
18. For point mutations (as opposed to large deletions), the differences between desired and undesired integration events are not directly evident by the diagnostic PCRs described here. This can be overcome either by sequencing the fragments or by simply picking more than one clone for excision.

References

1. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220.
2. Lartigue, C., Vashee, S., Algire, M.A., Chuang, R.Y., Benders, G.A., Ma, L., Noskov, V.N., Denisova, E.A., Gibson, D.G., Assad-Garcia, N. *et al.* (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science*, **325**, 1693–1696.
3. Scherer, S. and Davis, R.W. (1979) Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. *Proc Natl Acad Sci USA*, **76**, 4951–4955.
4. Gray, M., Piccirillo, S. and Honigberg, S.M. (2005) Two-step method for constructing unmarked insertions, deletions and allele substitutions in the yeast genome. *FEMS Microbiol Lett*, **248**, 31–36.
5. Storici, F., Durham, C.L., Gordenin, D.A. and Resnick, M.A. (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci USA*, **100**, 14994–14999.
6. Storici, F., Lewis, L.K. and Resnick, M.A. (2001) *In vivo* site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol*, **19**, 773–776.
7. Langle-Rouault, F. and Jacobs, E. (1995) A method for performing precise alterations in the yeast genome using a recycable selectable marker. *Nucleic Acids Res.* **23**, 3079–3081.
8. Nair, N.U. and Zhao, H. (2009) Mutagenic inverted repeat assisted genome engineering (MIRAGE). *Nucleic Acids Res.* **37**, e9.
9. Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E. and Resnick, M.A. (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol. Cell Biol.* **13**, 5315–5322.
10. Lobachev, K.S., Shor, B.M., Tran, H.T., Taylor, W., Keen, J.D., Resnick, M.A. and Gordenin, D.A. (1998) Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics*, **148**, 1507–1524.

Chapter 6

Creation of New Metabolic Pathways or Improvement of Existing Metabolic Enzymes by *In Vivo* Evolution in *Escherichia coli*

Isabelle Meynial-Salles and Philippe Soucaille

Abstract

A method for *in vivo* evolution of metabolic pathways in bacteria is described. This method is a powerful tool for synthetic biology type of metabolic design and can lead to the creation of new metabolic pathways or the improvement of existing metabolic enzymes. The proposed strategy also permits to relate the evolved phenotype to the genotype and to analyze evolution phenomenon at the genetic, biochemical, and metabolic levels.

Key words: *In vivo* evolution, New metabolic pathway, Metabolic pressure, Microorganisms

1. Introduction

It is now well known and studied that when a microorganism is growing under a strong pressure of selection (high temperature, strong substrate limitation, maladapted or nonnative carbon substrate etc.), variants are likely to emerge bearing changes in the DNA sequence that bring about an advantageous change in phenotype. If the selection is being applied to cell growing in liquid, the population with the original phenotype will be rapidly displaced by the variants (1, 2).

The genetic adaptation of a microorganism under a strong pressure of selection was used several years ago to change the substrate specificities of enzymes for the directed evolution of new functions. The first demonstration of this approach was done by using an *E. coli* strain with a *lacZ* deletion and by selecting spontaneous mutants that grow on lactose and other beta-galactoside sugars (3). In all the variants, the target of the evolution was the

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_6,

ebgA gene where point mutations in the structural gene alter the enzyme so that it hydrolyzes lactose, or lactulose or galactosylarabinose or lactobionic acid (4).

More recently, experiments have demonstrated that organisms are capable of evolving whole new metabolic pathways. The most common experimental process is to put a microorganism in a novel environment that contains a chemical that the organism has not been exposed to in the past. If that new chemical is the sole source of carbon or nitrogen the organism requires for survival, most of the time, the organism will die. However, if any of the organisms existing enzymes have the slightest ability to enhance reactions with the new resource, selection will strongly favor the evolution of the gene that produces that enzyme, and future mutations will further improve the ability of the enzyme to process the new chemical resource at high rate. This approach was successfully used in *E. coli* to evolve pathways for the use of 1.2 propanediol (5, 6), ethanol (7), glycerol, or lactate (8, 9) as carbon sources.

Here we presented an original and powerful technique which combines (1) the rational design of a microorganism to couple its growth rate/specific rate of substrate consumption to the flux in the metabolic pathway to create or evolve and (2) the ability of a microorganism to evolve *in vivo* under appropriate "metabolic pressure." The efficiency of this method was already shown to select for a new bifunctional enzyme for glycerol synthesis in *E. coli* (10), to create a novel pathway for methionine synthesis in *E. coli* (11) or to create a new respiratory chain in *E. coli* (12).

In this chapter, we will provide information on how to design and construct a recombinant *Escherichia coli* to create new metabolic pathways or new enzymes using *in vivo* evolution.

2. Materials

2.1. Deletion or Modification of the Expression of a Gene

1. Specific oligonucleotides were designed based on the gene sequence interested in.
2. High-fidelity DNA polymerase was used for all PCR amplification tests.
3. pKD3 or pKD4 and pkD46 plasmids were respectively used for resistance gene amplification and expression of the λ Red system to mediate homologous chromosomal recombination (13).
4. Flp recombinase from *S. cerevisiae* expressed from pCP20 plasmid was used to remove antibiotic resistance genes (14).
5. Chloramphenicol and kanamycin were respectively dissolved in ethanol and sterile water at 30 and 25 mg/mL as stock solutions and stored at −20°C.

6. L-arabinose was dissolved in deionized water and sterilized by filtration at 1 M as stock solution and a total of 1 mM is used for induction.
7. Luria Bertani (LB) medium: 10 g of tryptone, 5 g yeast extract, 10 g sodium chloride, supplemented with 30 μg/mL of chloramphenicol or 25 μg/mL of kanamycin when necessary, was used to culture strains during the knock-in process.

2.2. Overexpression of One or Several Genes Encoding a Complete Metabolic Pathway

1. pSC101 a stable, low copy number plasmid (and related plasmid) expressing the RepA replication initiator protein was routinely used for the overexpression of heterologous or nonheterologous metabolic pathways (15).
2. High-fidelity DNA polymerase was used for PCR amplification of heterologous or nonheterologous genes.
3. T4 ligase and appropriated restriction enzymes were used for ligation and plasmid construction validation.
4. Qiagen kits were used for plasmid, gel-purify fragments or purify PCR products.
5. Ampicilline, carbenicilline, and spectinomycine were respectively dissolved in sterile water at 100 and 50 mg/mL as stock solutions and stored at −20°C.

2.3. Metabolic Evolution of Microorganisms

1. Aerobic *E. coli* medium (1 L): 8 g KH_2PO_4, 2 Na_2HPO_4, 0.75 g $(NH_4)_2SO_4$, 8 $(NH_4)_2HPO_4$, 6.6 g Citric acid, 2.05 g $MgSO_4$, 40 mg $CaCl_2$, 40 mg $FeSO_4$, and 10 mL Trace Element Solution separately sterilized.
   (a) Trace Element Solution (1 L): 2 g $MnSO_4 \cdot H_2O$, 0.8 g $CoCl_2 \cdot 6H_2O$, 0.4 g $ZnSO_4 7H_2O$, 0.4 g $Na_2MoO_4 \cdot 2H_2O$, 0.2 g $CuCl_2 \cdot 2H_2O$, 0.1 g H_3BO_3, 10 mL concentrated HCl (37%).
   (b) Glucose (20 g/L) was autoclaved separately, while appropriated antibiotic and thiamine (15 mg/L) were filtrated for sterilization and added just before inoculation.
2. Anaerobic *E. coli* medium (1 L): 1 g K_2HPO_4, 1 g $(NH_4)_2SO_4$, 0.2 g NaCl, 0.2 g $NaHCO_3$, 0.2 g $MgSO_4$, 50 mg $FeSO_4$, 0.42 g sodium nitrate, 10 mg thiamine, and 10 mL Trace Metal Stock.
   (a) Trace Metal Stock (1 L): 4 g citric acid, 3 g $MnSO_4$, 1 g NaCl, 0.1 g $CoCl_2$, 0.10 g $ZnSO_4$, 10 mg $CuSO_4$, 10 mg H_3BO_3, 10 mg $NaMoO_4$.
   (b) Glucose (from 20 g/L up to 60 g/L), yeast extract 4 g/L, and appropriate antibiotics were separately sterilized and added just before inoculation.
3. Glass stirred bioreactors equipped with automatic pump (chemostat device, Sartorius) were used for continuous cultures under both anaerobic and aerobic conditions.

4. Growth was evaluated by measuring OD at 600 nm using a spectrophotometer.
5. High-Performance Liquid Chromatography (Agilent Technologies 1200) was used to determine product and substrate concentrations all along cultures.

2.4. Selection of Evolved Clones from the Evolved Population

1. Solid M9 minimal medium (1 L): 0.5 M $MgSO_4$, 0.1 M $CaCl_2$, 200 mL 5× M9 salts, and 15 g agar. Glucose (5 g/L) was separately autoclaved, while thiamine (15 mg/L) was filtrated for sterilization and added just before pouring the plates.
2. 5× M9 salts (1 L): 5.0 g NH_4Cl, 2.5 g NaCl, 64 g $Na_2HPO_47H_2O$, 15 g KH_2PO_4.
3. Appropriate antibiotics were added to the M9 minimum medium agar plate just before spreading on bacteria.

2.5. Genomic DNA Preparation

1. The QIAamp DNA mini kit was used for Genomic DNA extraction.
2. If necessary, genomic DNA is further concentrated by vacuum centrifugation (Speedvac).

3. Methods

The power of the technique described here has already been demonstrated in prokaryotes in particular in *E. coli*, but can be applied to eukaryotic cells like yeast or fungi.

Five steps are needed to select an evolved microorganism expressing a novel or improved phenotype and to determine the genetic basis of the metabolic evolution.

First, to favor directed *in vivo* evolution, a modified microorganism must be conceived to reroute its central metabolism and create an absolute link between specific substrate consumption rate or specific growth rate and the flux in the metabolic pathway/enzyme to improve or create. If the growth of the modified microorganism is hampered (a low specific growth rate is observed), a metabolic evolution is secondly carried out by serial batch subcultures maintaining exponential phase on an appropriate culture medium of known defined composition adapted for the microorganism and containing a simple carbon source. If the specific substrate consumption rate of the modified microorganism is hampered, a metabolic evolution is secondly carried out in continuous cultures under excess of substrate. During the evolution experiment, the specific growth rate/specific substrate consumption rate steadily increased over the generations, and the experiment is stopped once a stable specific growth rate/specific substrate

consumption rate is achieved and resulted in the evolved population. In a third step, clones are isolated as colonies from the evolved population for phenotypic characterization. The specific growth rate/specific substrate consumption rate of some randomly selected clones is analyzed in batch culture on the same defined medium used for evolution experiments. Among clones having the same specific growth rate/specific substrate consumption rate than the evolved population, one (generally the one having the highest specific growth rate/specific substrate consumption rate) is selected to determine the genetic basis of the metabolic evolution.

Fourth, genomic DNA of the selected evolved clone is extracted and genetic mutations are identified using comparative genome mutation mapping (CGS, Roche Nimblegen) or DNA sequencing (Illumina genome analyzer II). Indeed, the first method has already been demonstrated to be fast and efficient in detecting multiple mutations present in one bacterium relative to another (16, 17). Putative mutations detected from CGS are then validated by PCR and sequencing. In case of expression of a metabolic pathway from a plasmid, plasmid is also extracted and the entire metabolic pathway sequenced. Both methods permit to identify the genetic basis of the metabolic evolution phenomenon observed and to relate genotype to phenotype.

The final step is the analysis of the genetic mutations found. To evaluate the contribution of individual mutation identified that enabled microorganisms to increase their specific growth rate/specific substrate consumption rate on defined medium, each mutation is introduced one by one into the unevolved modified strain. The specific growth rate/specific substrate consumption rate recovery of the mutants newly constructed is analyzed on the same defined medium used for evolution experiments. When the specific growth rate/specific substrate consumption rate of reconstructed strains with almost mutations matches the specific growth rate/specific substrate consumption rate of the selected evolved clone, it indicates that these mutations are responsible for the metabolic evolution phenomenon observed. This last experiment permits to evaluate the contribution of each mutation to fitness, to relate the phenotype to genotype, and to identify the newly metabolic pathway created or modified.

Finally, a functional characterization of each mutation may be done to verify the exact nature of each mutation and its phenotypic consequence. The appropriate test is chosen on case by case depending of the genetic target.

3.1. Conception of a Modified Microorganism

Key modifications to be introduced into the central metabolism to construct the modified microorganism, able to evolve *in vivo* under appropriate metabolic pressure, are selected on case by case basis according to the choice of metabolic pathway to be modified or

created. For an optimization of the key modifications to introduce, the computational OptKnock framework can be used to explore and finally suggest gene deletion strategies to ensure that the cellular objective (i.e., biomass formation, substrate consumption) is directly coupled to the flux in the metabolic pathway/enzyme to improve or create (18). This coupling is accomplished by ensuring that the metabolite of interest becomes an obligatory byproduct of growth by "shaping" the connectivity of the metabolic network. OptKnock has been shown to identify not only straightforward but also nonintuitive knockout strategies by considering the entire range of biotransformations in genome-scale metabolic networks of *E. coli*, *S. cerevisiae*, and others organisms. Optgene which is an extension of Optknock can too be used (19). It presents two major advantages, higher speed and ability to optimize for nonlinear objective functions.

From a practical point of view, the construction of the modified microorganism is based on the use of combined technical approaches:

3.1.1. Deletion of Genes Encoding Metabolic Enzymes

Deletion of genes encoding metabolic enzymes and removal of antibiotic resistance genes were according to Datsenko and Wanner (13).

1. Kanamycin and chloramphenicol resistance genes (Km^R or *CmR*) carried by pKD4 or pKD3 plasmids, respectively, were amplified by PCR.
2. Sense and antisense primers contained respectively nucleotide sequences (80 bp) corresponding to the N-terminus and C-terminus of each targeted gene followed by 20 bp (priming site 1 (P1: 5′ATTCCGGGGATCCGTCGACC3′) and priming site 2 (P2: 5′TGTAGGCTGGAGCTGCTTCG3′)) corresponding to the Flp recognition target site (FRT)-antibiotic resistance cassette (13).
3. Single knock-out mutants were constructed by electroporation of amplified DNA fragments into *E. coli* strains harboring pKD46 containing an arabinose-inducible Red recombinase (see Notes 1 and 2).
4. After the first gene replacement by an antibiotic resistance gene, additional chromosomal deletions were achieved by generalized P1 phage transduction from single mutant lysates (20) (see Note 3).
5. Removal of antibiotic resistance genes was by Flp recombinase expressed from pCP20 (14), leaving a FRT scar.
6. Chromosomal deletions and/or integrations were tested for antibiotic resistance and PCR analysis (see Note 3).

3.1.2. Decreasing or Increasing of the Expression of a Metabolic Enzyme or a Complete Metabolic Pathway

It is often necessary to tune the level of many enzymes to achieve the expected flux in one metabolic pathway and force the evolution of the metabolic pathway/enzyme to improve or create. The modulation of expression of chromosomal genes was according to Meynial-Salles et al. (21).

1. The first step was the creation of a library of expression cassettes with controlled degeneracy artificial promoters. This step included the design of two 100-mer oligonucleotides (OL_1 and OL_2) containing various specific regions (called pS_1, pS_2, H_1, and P_1) (Fig. 1). The reverse oligonucleotide was degenerated on specific bases of the sequence coding for a very short

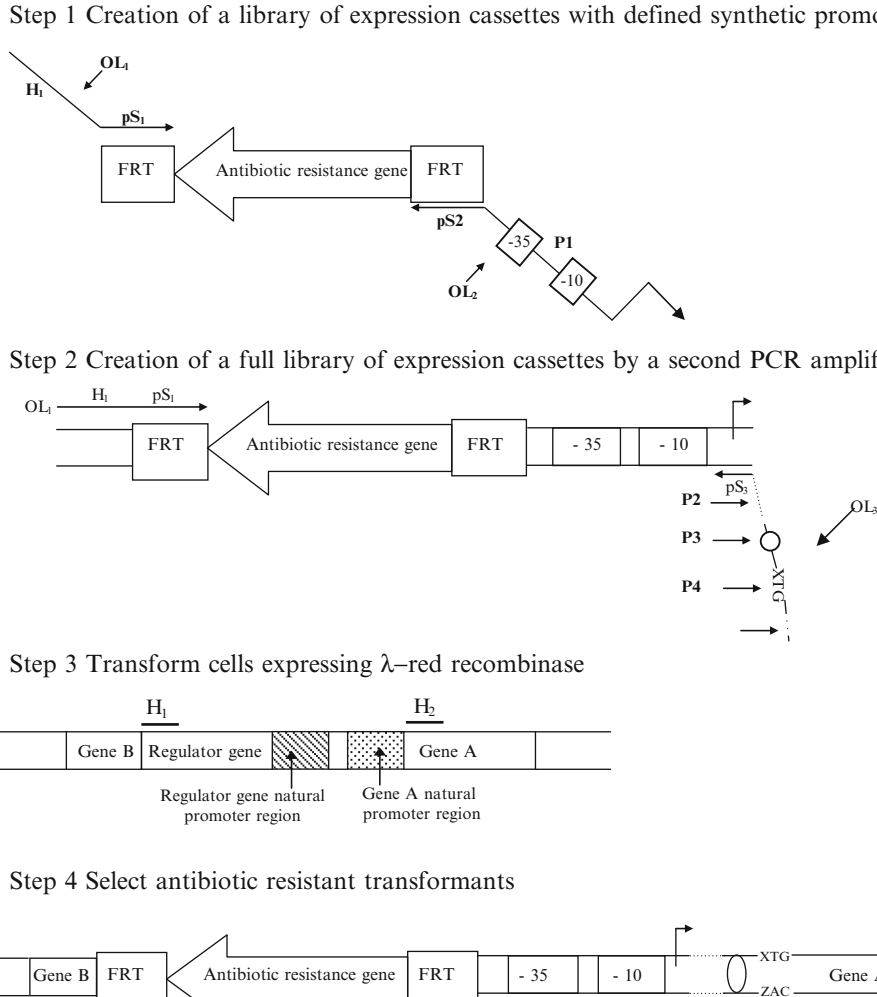


Fig. 1. Strategy for the creation of a library of clones with different level of expression of a chromosomal gene pS_1 and pS_2 is the priming sites for pKD3 or pKD4. H1 and H2 are the homology extensions. P_1 is the degenerated part of the reverse oligonucleotide OL_2 containing synthetic short GI promoters. OL_3 can contain up to three degenerated sequences, P_2 for mRNA stabilizing sequences, P_3 for ribosome binding sites, and P_4 for start codons.

artificial promoter (P_1) to create a library of promoters. The template plasmid pKD3 or pKD4 (6) and the two oligonucleotides (OL_1 and OL_2) were then used in a PCR reaction (Fig. 1) producing a first set of expression cassettes.

2. The second step was the extension of the initial promoter library by adding modified and defined RBS and/or mRNA stabilizing sequences and/or start codons depending on the range of expression level needed. This PCR step used the first set of expression cassettes as a template. The forward oligonucleotide is the same as that used for the first step (OL_1). The reverse oligonucleotide (OL_3) contains a priming sequence (pS_3) for the artificial promoter, defined and degenerated sequences encoding (1) mRNA stabilizing sequences (P_2), (2) ribosome binding sites (P_3), (3) start codons (P_4) and a flanking sequence (H_2) homologous to the chromosomal DNA downstream of the start codon (Fig. 1). After the second PCR reaction, a full library of expression cassettes was obtained.
3. Mutants were constructed by electroporation of amplified DNA fragments into *E. coli* strains harboring pKD46 containing an arabinose-inducible Red recombinase (see Notes 1 and 2).
4. The result of the *in vivo* recombination was the replacement of all or part of the upstream region of a coding sequence containing the elements involved in its expression by a PCR-generated library of expression cassettes.
5. Chromosomal promoter region modifications were tested for antibiotic resistance and PCR analysis (see Note 3).

3.1.3. Overexpression of Heterologous or Nonheterologous Genes Encoding a Complete Metabolic Pathway

1. Overexpression of heterologous or nonheterologous genes can be alternatively achieved from a stable single copy (or multicopy) replicative plasmid carrying the gene(s) of interest overexpressed under the appropriate promoter or by direct introduction into the genome overexpressed under a strong or inducible promoter according to Datsenko and Wanner (13).
2. Plasmid or DNA amplification fragments (expressing genes) were introduced by electroporation into the *E. coli* cells.

3.2. In Vivo Metabolic Evolution

3.2.1. Evolution Using Serial Batch Subcultures

1. Evolution was in 50 mL of minimal synthetic medium in 500-mL Erlenmeyer baffled flasks (or in 100-mL anaerobic vials under anaerobic conditions), at 37°C, on a gyratory shaker at 200 rpm.
2. At the beginning of the evolutionary process, a preculture was grown in LB medium, harvested by centrifugation, washed in the same volume of fresh minimal medium (lacking carbon source, thiamine and $FeSO_4$), and used for inoculation to an optical density 600 nm (OD_{600}) of 0.1.

3. Serial evolution subculture growth was regularly determined by OD_{600} measurement, and cells were diluted to an $OD_{600\ nm}$ of 0.1 into fresh medium before they entered stationary phase (see Note 4).
4. Evolution experiment is stopped once a stable specific growth rate/specific substrate consumption rate is achieved after four successive batch subcultures (6, 12).
5. Samples were frozen in 20% (v/v) glycerol solution from each serial evolution subculture.

3.2.2. Evolution Using Continuous Cultures

1. Evolution was in 300 mL of minimal synthetic medium in 500-mL glass stirred bioreactor equipped with automatic pumps to renew the liquid culture at a constant volume with nutrient medium inflow, such that *E. coli* must counteract dilution by growing at least at an equal rate (chemostat device, Sartorius) . The temperature and pH were respectively regulated at 35°C and pH6.5 (by addition of 6 M NH_4OH).
2. At the beginning of the evolutionary process, a preculture was grown in LB medium, harvested by centrifugation, washed in the same volume of fresh minimal medium (lacking carbon source, thiamine and $FeSO_4$), and used for inoculation to an optical density 600 nm (OD_{600}) of 0.1.
3. After 24 h of batch conditions, culture was continuously fed with fresh minimal medium at a low dilution ($D \approx \sim 0.05\ h^{-1}$).
4. During the evolution experiment, dilution rate was gradually increased once a steady state was achieved and up to a stable specific substrate consumption rate was measured.
5. Samples were frozen in 20% (v/v) glycerol solution from each steady state.

3.3. Selection of Evolved Clones

1. The evolved population (10^{-5} or 10^{-6} dilution depending on the biomass concentration) was then spread onto agar minimal medium containing a unique carbon source and a cosubstrate if needed.
2. The medium used to select the evolved clone was identical to the defined medium on which the initial microorganism grew.
3. A single colony (see Note 5) was selected from the plate that was incubated and resuspended in a low volume of defined medium and inoculated into 500-mL Erlenmeyer flasks (for aerobic experiment) or into 100-mL anaerobic vials (anaerobic experiment) containing 50 mL of the same defined medium used for metabolic evolution experiments. The flasks were incubated at 37°C using a stir bar for mixing.

4. Cell growth was monitored as $OD_{600\ nm}$ and specific growth rate and specific consumption rate were evaluated. Among clones having the same specific growth rate/specific substrate consumption rate as the evolved population, one was selected to determine the genetic basis of the metabolic evolution.

3.4. Analytical Procedures and Physiological Parameters

1. *E. coli* cell growth was monitored as OD_{600} using a spectrophotometer.
2. Substrate and by-products concentrations were determined all along cultures by High-Performance Liquid Chromatography (Agilent Technologies 1200). Samples were centrifuged and filtrated (0.2 μM) before loaded onto a Bio-Rad Aminex HPX-87H column (300 × 7.8 mm). Detection was done by refractometry. Operating conditions were mobile phase 5 mM sulfuric acid, flow-rate 0.6 mL/min, temperature 48°C. External calibration is used to quantify products and substrate concentrations.
3. Physiological parameters (maximum growth rate, specific glucose consumption rate) were determined during exponential growth phase.

3.5. Analysis of the Mutation Found

1. Genomic and/or plasmid DNA were extracted from the selected evolved clone to be sequenced and identify genetic mutations.
2. To evaluate the contribution of individual mutation found by sequencing, each mutation was reintroduced one by one into the unevolved strain according to Datsenko and Wanner (13).
3. In parallel, individual wild-type genotype was reintroduced into the evolved clone according to Datsenko and Wanner (13).
4. Specific growth and specific consumption rates of each newly reconstructed strain were evaluated in the same defined medium used for evolution experiments.

3.6. General Comments and Conclusion

Among the several examples of directed *in vivo* evolution of *Escherichia coli* the authors did, it was showed:

1. Only a small number of mutations were accumulated during *in vivo* evolution under metabolic pressure.
2. Marked changes in phenotype can be mediated by as few as one or two mutations.
3. None of the mutations were found in the intergenic region.
4. Main mutations was found to occur in coding or promoter regions and always made sense from a metabolic point of view.

4. Notes

1. Single knock-out mutants or gene insertion mutants accomplished by Red-mediated recombination must be selected at 30°C (13).
2. pKD46 plasmid containing an arabinose Red recombinase can be lost by subculturing recombinant cells at 37°C without antibiotics (13).
3. After achieving additional chromosomal deletions by generalized P1 phage transduction from single mutant lysates (20), the presence of all previous deletions must be checked by PCR.
4. Serial batch subcultures must be done to maintain at exponential phase by daily passage of cultures into fresh medium prior to entry into stationary phase to avoid cell lyses (OD of the culture at passage is depending of the period of the evolution). Inoculum at time of passage is adjusted to account for increasing growth rates over evolutionary time (slope of log plot) ($OD_{600} \approx 0.1$) (6, 12).
5. Twice successive isolations on mineral medium agar plates are done to select pure clones from the evolved population.

Acknowledgments

This work was supported through several independent collaborations with both Genencor and Metabolic Explorer companies, FP5 European project (contract n° QLRT-1999-01364), and the French National Research Agency (contract PRIB-2005 Bioglycol).

References

1. Atwood, K. C., Schneider, L. K., Ryan, F. J., (1951) Selective mechanisms in bacteria. Cold Spring Harbor Symp. *Quant. Biol.* **16**, 345–355.
2. Dykhuizen, D. E., (1990) Experimental studies of natural selection in bacteria. *Annu. Rev. Ecol. Syst.* **21**, 373–398.
3. Hall, B.G., Zuzel T., (1980) Evolution of a new enzymatic function by recombination within a gene. *Proc. Natl. Acad. Sci. USA.* 77, 3529–3533.
4. Hall, B.G., (1981) Changes in the substrate specificities of an enzyme during directed evolution of new functions. *Biochemistry.* **20**, 4042–4049.
5. Lu, Z., Cabiscol, E., Obradorsi, N., Tamarit, J., Ros, J., Aguilari, J., Lin, E.C.C., 1998. Evolution of an *Escherichia coli* Protein with Increased Resistance to Oxidative Stress. *J. Biol. Chem.* **273**, 8308–8316.
6. Lee D-H and Palsson B. (2010) Adaptive evolution of *Escherichia coli K12* MG16555 during growth on a nonnative carbon source, L 1.2 propanediol. *Appl. Environ. Microb.* **76**, 4158–4168.
7. Membrillo-Hernandez, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., Lin, E.C.C. (2000) Evolution of the *adhE* Gene Product of *Escherichia coli* from a Functional Reductase to a Dehydrogenase. *J. Biol. Chem.* **275**, 33869–33875.
8. Hua Q., Joyce A.R., Palsson B. and Fong S. S. (2007) Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Appl. Environ. Microb.* **73**, 4639–4647.

9. Applebee M. K., Herrgård MJ, Palsson BØ. (2008) Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli*. *J. bacteriol.* 190, 5087–5094.
10. Meynial-Salles I.,Forchammer N., Croux C. Girbal L., and Soucaille P. (2007) Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. *Met Eng.* 9, 152–159.
11. Chateau M., Gonzalez B., Meynial-Salles I., Soucaille P. Zink O. (2005) Method for the preparation of an evolved microorganism for the creation or the modification of metabolic pathways: US patent 205/0054060.
12. Auriol C., Bestel-Corre G., Claude J. B., Soucaille P. and Meynial-Salles I. (2011) Stress induced evolution of *Escherichia coli* points to new concepts in respiratory cofactor selectivity. *Proc. Natl. Acad. Sci USA.* **108**, 1278–1283.
13. Datsenko, K. A., Wanner, B. L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA.* **97**, 6640–6645.
14. Cherepanov PP, Wackernagel W (1995) Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene.* **158**, 9–14.
15. Kues U. and Stalh U. (1989) Replication plasmids in Gram-Negative Bacteria. *Microbiological reviews.* **53**, 491–516.
16. Herring C.D. CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BØ (2006) Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. *Nat Genet.* **38**, 1406–1412.
17. Conrad T.M., Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BØ. (2009) Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* **10**, R118.
18. Burgard A., Pharkya P. and Maranas C. (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization: *Biotechnol. Bioeng.* **84**, 647–657.
19. Patil K. R., Rocha I., Forster J. and Nielsen J. (2005) Evolutionary programming as a platform for in silico metabolic engineering: BMC Bioinformatics, **6**. 1–12.
20. Miller JH (1992) A short course in bacterial genetics. A laboratory manual and handbook for *Escherichica coli* and related bacteria.
21. Meynial-Salles I., Cervin M. A. and Soucaille P. (2005). New tool for metabolic pathway engineering in *E. coli*: one step method to modulate the expression of chromosomal genes: *Appl. Environ. Microbiol.* **71**, 2140–2144.

Chapter 7

Bioluminescent Reporter Genes for Promoter Discovery

Tina K. Van Dyk

Abstract

Discovery of promoter elements with previously unknown regulated responses is important for metabolic engineering. For example, promoters responsive to the end product can be useful to regulate expression with increasing levels of product. In addition, such promoters can be used as screens for production strain with increased titers. Use of reporter genes, such as a bioluminescent reporter *luxCDABE*, can facilitate promoter discovery. Here, protocols for analysis of genome-wide *luxCDABE* reporter gene collections in *Escherichia coli* are provided. Further, a protocol for using a selected *para*-hydroxycinnamic (pHCA)-responsive promoter as detection assay for bioproduced pHCA is provided.

Key words: Bacterial bioluminescence, Luminometry, *lux* reporter gene fusions, LuxArray

1. Introduction

The choice of promoter to drive gene expression is critically important in metabolic engineering, yet the range of promoter availability is limited. In particular, promoters that respond to the concentration of end product in the medium could be valuable to drive expression of tolerance genes or other genes required as the product accumulates. Promoters responsive to fermentation end products can be discovered in the production host strain by genome-wide analysis of transcriptional responses to treatment with externally added chemical. DNA microarray analysis or RNA sequencing is frequently used for such an approach. However, genome-wide collections of whole cell biosensors carrying reporter gene fusions can also be used for this application. An advantage of using reporter genes is that selected biosensor cells can readily be used for further characterization of the transcriptional response and for such applications as developing a biosensor for screens of the end product production.

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_7,

Of the commonly used reporter genes, bacterial bioluminescence encoded by the *luxCDABE* operon has the advantage of measurement without disruption of the cell allowing continuous monitoring of expression. Furthermore, the sensitivity and large dynamic range of light detection equipment enables measurement of very low levels and very high levels of gene expression. Genome-wide collections of *lux* reporter gene fusions have been developed for several bacterial species, such as *Escherichia coli* (1), *Salmonella enterica serovar typhimurium* (2, 3), and *Staphylococcus aureus* (4). An example of promoter discovery using the *E. coli* LuxArray (5) is given in this chapter. This example is with regard to bioproduced *para*-hydroxycinnamic acid (pHCA), a bi-functional aromatic compound (6).

2. Materials

2.1. Overnight Cultures

1. Costar#3595 96-well flat bottomed plates with low evaporation lids.
2. LB medium supplemented with 100 mg/L amplicillin.
3. Microplate centrifuge.
4. 96-well inoculating device.
5. Tupperware containers.

2.2. Growth Cultures

1. Costar#3595 96-well flat bottomed plates with low evaporation lids.
2. Microplate shaker, such as IKA Schuttler MTS 4 shaker.
3. Humidified Box to contain microplate shaker.
4. Incubator set at 37°C.

2.3. Quantitation of Responses to Chemical Stresses

1. Microplate luminometer, such as Dynex MLX.
2. 96-well white luminometer plates (Dynex).
3. Incubator set at 37°C.

3. Methods

The format initially implemented for the *E. coli* LuxArray entailed printing the array of whole cell biosensors on solid medium (5). However, issues arose such as cross-illumination between spots on the array that interfered with accurate detection of promoter activity. Thus, an alternative liquid format, herein described, was developed. This protocol employs growth and luminescence assays in a

series of standard 96-well microplates and white luminometer plates designed to minimize cross talk between wells. Furthermore, measurement of light employing a microplate luminometer allows a greater dynamic range of measurements than possible with a cooled CCD camera that had been used for the solid medium format. The biosensors in the LuxArray are organized into 18 96-well microplates; thus, this protocol is designed to test a third of the array per day.

3.1. Overnight Culture Plates

1. Pipette 100 μL LB medium supplemented with 100 mg/L ampicillin into all wells of six 96-well microplates (see Note 1).
2. Remove six LuxArray working plates from –80°C and quick thaw at 37°C for ~10 min (see Note 2).
3. Spin down working plates and inoculate the microplates using a 96-well inoculation device (see Note 3), then cover plates with low evaporation lids.
4. Grow overnight in Tupperware containers lined with wet paper towels without shaking at 37°C (see Note 4).

3.2. Growth Culture Plates

1. Pipette 135 μL LB medium into six flat bottom 96-well microplates with low evaporation lids.
2. Note which wells in the overnight plates have little or no growth (see Note 5).
3. Pipet 15 μL of cells from first overnight plate into first growth plate (see Note 1).
4. Use moderate shaking at setting 400 on IKA Schuttler MTS 4 shaker, 2.0 h, 37°C in humidified box.
5. Seven minutes after Plate 1 inoculation, repeat steps 2–4 for second growth plate.
6. Additionally, repeat steps 2–4 for each of the remaining growth plate staggered by 7 min.

3.3. Quantitation of Responses to Chemical Stresses

1. Prewarm microplate luminometer to 37°C.
2. During 2.0 h incubation, pipette 80 μL LB into six 96-well white luminometer plates and 80 μL LB + chemical at 1.25× the desired final concentration into six additional white luminometer plates. For experiments with two concentrations of chemical, prepare six additional plates with 80 μL LB + higher concentration of chemical at 1.25× the desired final concentration (see Notes 1, 6, and 7).
3. At the 2.0 h time point, deliver 20 μL cells from growth Plate 1 into the prepared control (LB) and LB + chemical plates (see Note 1).
4. Take 0 time bioluminescence measurement in microplate luminometer with a single cycle assay, then incubate covered at

37°C without shaking, on incubator shelf, and measure bioluminescence at 45 min intervals until time 135 min.

5. Perform steps 3 and 4 for each growth plate at end of 2-h incubation.

3.4. Data Analysis

1. Normalize data to correct for differing cell growth profiles caused by the chemical treatment, by summing the bioluminescence from the total array at each time point and calculating a normalization factor using the equation: Normalization factor = total array signal at time zero LB control/total array signal at time x and condition y. Multiply each measurement by the normalization factor to yield a normalized signal for each reporter at each time point for each treatment (see Note 8).
2. Process the normalized data so that the time course of bioluminescence for each chemical treatment can be examined (see Note 8).
3. Examine the data to find upregulated or downregulated genes responsive to the chemical treatment (see Note 9). For example, Figure 1 shows the response of a fusion in the LuxArray, lux-a.pk035.c7, to two concentrations of pHCA. Although the growth inhibition as estimated by the inhibition of total bioluminescence of the array was modest at 13% for 5 mM pHCA and 27% for 10 mM pHCA, the response of this fusion was dramatic upregulation of expression. The pHCA-responsive

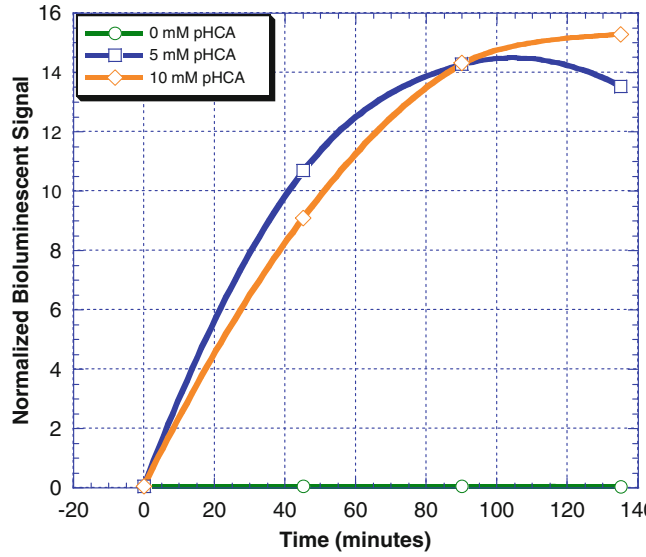


Fig. 1. Response of lux-a.pk035.c7 to pHCA. The normalized bioluminescent signal during the time course of a liquid Lux Array experiment is shown. *Circles*, control untreated. *Squares*, treated with 5 mM pHCA. *Diamonds*, treated with 10 mM pHCA. The time zero normalized bioluminescent signal was 0.061, 0.056, and 0.056, respectively.

promoter element in fusion lux-a.pk035.c7 discovered by analysis of the LuxArray and in DNA microarray experiments was subsequently shown to be the promoter for the *aaeXAB* operon encoding an efflux system for aromatic carboxylic acids (7). An upstream LysR-family regulator controls expression in response to aromatic carboxlic acids such as pHCA and *para*-hydroxybenzoate (7).

3.5. Use of a Bioluminescent Gene Fusion to Detect Biologically Produced pHCA

1. Use a bioluminescent pHCA sensor strain DPD2439, which carries the lux fusion plasmid from lux-a.pk035.c7 in a *tolC–E. coli* host strain. Grow this strain in LB with 100 mg/L ampicillin at 37°C with shaking until growth is in the log phase.
2. Add 50 µL of bioluminescent sensor strain to 50 µL of cell culture supernatant from pHCA producing strains in wells of microplate luminometer plates. For a negative control, use 100 µL of biosensor strain alone.
3. Measure bioluminescence for 1 h with automated readings every 3 min in microplate luminometer at 37°C, with shaking.
4. Examine data for upregulation of gene expression for evidence of pHCA in the culture medium. For example, Fig. 2 shows the responses from the culture supernatants of two pHCA producing strains. Both show an upregulation of the bioluminescent response consistent with pHCA in the culture medium.

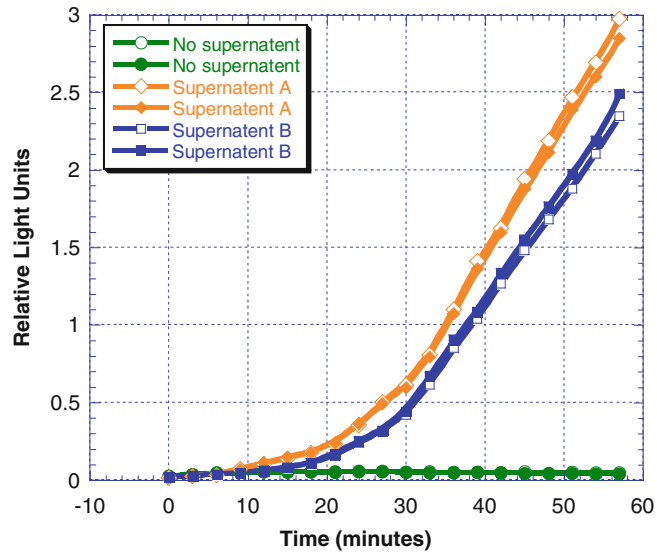


Fig. 2. Response of lux-a.pk035.c7 to biologically produced pHCA. The time course of the bioluminescent signal (relative light units, dimensionless) in response to two sources of biologically produced pHCA is shown, for duplicate samples. *Circles*, control untreated. *Diamonds*, supernatant A. *Squares*, supernatant B.

4. Notes

1. Automated pipetting devices, such as Multimek (Beckman Scientific), can be used for 96-well plate filling and inoculation of growth and chemical stress plates.
2. Thaw each working set only 3 times then discard. On the third thaw, grow overnights 18–20 h. We found it useful to make new sets of working plates upon the first thaw of the last set of working plates.
3. A Biomek 2000 with a HDRT flat-head pin set is useful for inoculation of the overnight culture plates.
4. Shaking the microplates is not necessary for overnight growth, if at least 16 h is allowed.
5. Occasionally, a few wells have poor growth in the overnight culture. These should be excluded from further analysis.
6. Select subinhibitory concentrations of chemicals to test. It is generally useful to have inhibition of total array bioluminescence in the range of 20–70%.
7. It is helpful to consistently label the luminometer plates (e.g., at front and right edges) for quickly moving plates back and forth from the incubator to microplate luminometer in the correct orientation.
8. Macros in Microsoft Excel are valuable for collating, organizing, and normalizing the large amounts of data generated in these experiments.
9. Programs for analysis of microarray data can typically be adapted to use with LuxArray data sets.

References

1. Van Dyk TK, Wei Y, Hanafey MK, Dolan M, Reeve MJG, Rafalski JA, Rothman-Denes LB, LaRossa RA (2001) A genomic approach to gene fusion technology Proc Natl Acad Sci USA 98:2555–2560
2. Goh EB, Yim G, Tsui W, McClure J, Surette MG, Davies J (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics Proc Natl Acad Sci USA 99:17025–17030
3. Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica serovar typhimurium* by high-throughput screening of a random promoter library J Bacteriol 185:4973–4982
4. Mesak LR, Qi S, Villanueva I, Miao V, Davies J (2010) *Staphylococcus aureus* promoter-*lux* reporters for drug discovery J Antibiot 63: 492–498
5. Van Dyk TK, DeRose EJ, Gonye GE (2001) LuxArray, a high-density, genomewide transcription analysis of *Escherichia coli* using bioluminescent reporter strains J Bacteriol 183: 5496–5505
6. Sariaslani S, Van Dyk T, Huang L, Gatenby A, Ben-Bassat A (2005) Renewable resources for production of aromatic monomers In Hou CT (Ed) Handbook of Industrial Catalysis. CRC Press, Boca Raton
7. Van Dyk TK, Templeton LJ, Cantera KA, Sharpe PL, Sariaslani FS (2004) Characterization of the *Escherichia coli* AaeAB Efflux Pump, a Metabolic Relief Valve? J Bacteriol 186: 7196–7204

Chapter 8

Recombination-Based DNA Assembly and Mutagenesis Methods for Metabolic Engineering

Xiquan Liang, Lansha Peng, Billyana Tsvetanova, Ke Li, Jian-Ping Yang, Tony Ho, Josh Shirley, Liewei Xu, Jason Potter, Wieslaw Kudlicki, Todd Peterson, and Federico Katzen

Abstract

In recent years there has been a growing interest in the precise and concerted assembly of multiple DNA fragments of diverse sizes, including chromosomes, and the fine tuning of gene expression levels and protein activity. Commercial DNA assembly solutions have not been conceived to support the cloning of very large or very small genetic elements or a combination of both. Here we summarize a series of protocols that allow the seamless, simultaneous, flexible, and highly efficient assembly of DNA elements of a wide range of sizes (up to hundred thousand base pairs). The protocols harness the power of homologous recombination and are performed either in vitro or within the living cells. The DNA fragments may or may not share homology at their ends. An efficient site-directed mutagenesis protocol enhanced by homologous recombination is also described.

Key words: Recombineering, Recombinational, Yeast, Mutation, Double-strand break repair, Synthetic biology

1. Introduction

The long standing ambition of converting a digital sequence of a chromosome stored in a computer into its biological counterpart in a single step requires that the classical recombinant DNA technology needs to evolve to cover a wide range of DNA sizes and number of fragments (for a brief summary of advanced cloning systems see Note 1).

The number and size of fragments that can be simultaneously assembled is limited due to poor cloning efficiency. In addition, most of the in vitro methods for joining two or more DNA fragments require specific sequences that leave watermarks or scars at

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834, DOI 10.1007/978-1-61779-483-4_8, © Springer Science+Business Media, LLC 2012

the end of the process. In our opinion, the most viable strategies are those that make use of in vitro or in vivo double-strand break repair mechanisms.

In this chapter we provide a set of assembly and mutagenesis protocols based on homologous recombination useful for assembling and editing small, intermediate, and large DNA fragments. Examples of other similar approaches have been previously described (1–6).

2. Materials

2.1. Yeast Cloning

1. DNA fragments to assemble.
2. Stitching oligonucleotides (only if necessary; up to three pairs per assembly reaction, see section 3.1).
3. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.
4. pYES1L linear cloning vector (Life Technologies, Carlsbad, CA, part of cat # A13286 or A13287) or your own yeast-adapted cloning vector.
5. MaV203 competent yeast cells (Life Technologies, Carlsbad, CA, part of cat # A13286 or cat # 11445012) or equivalent.
6. PEG/LiAc solution (Life Technologies, Carlsbad, CA, part of cat # A13286) or any other DNA condensing reagent.
7. 0.9% NaCl solution (sterile).
8. DNase-, RNase-Free water.
9. CSM-Trp agar plates (Life Technologies, Carlsbad, CA, part of cat # A13286 or cat # A13292) or equivalent.
10. Pair of diagnostic primers for each DNA junction including the cloning vector.
11. Yeast lysis buffer (Life Technologies, Carlsbad, CA, part of cat # A13286) or equivalent.
12. Dimethyl sulfoxide (DMSO).
13. Platinum® PCR SuperMix (Life Technologies, Carlsbad, CA, cat # 11306081) or similar.
14. Plates with yeast colonies containing the plasmid of interest.
15. Glass beads (Life Technologies, Carlsbad, CA, part of cat # A13286) or similar.
16. S.O.C. medium: 20 g/L Bacto Tryptone, 5 g/L Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
17. One Shot® TOP10 Electrocompetent *Escherichia coli* cells (Life Technologies, Carlsbad, CA, part of cat # A13286 or cat # C4040-52) or similar.

18. Electroporation cuvettes.
19. Electroporator.
20. LB plates with the appropriate selection antibiotics, prewarmed to 37°C.
21. Deionized, sterile water.

2.2. In Vitro DNA Assembly

1. DNA fragments for DNA assembly.
2. Linearized *E. coli* vector.
3. In vitro recombination buffer (Life Technologies, Carlsbad, CA, part of cat # A13288) or equivalent.
4. In vitro recombination enzyme mix (Life Technologies, Carlsbad, CA, part of cat # A13288) or equivalent.
5. Deionized, sterile water.
6. One Shot® TOP10 chemically competent *E. coli* (Life Technologies, Carlsbad, CA, part of cat # A13288 or cat # C404010) or similar.
7. S.O.C. medium (formulation above).
8. LB plates with the appropriate selection antibiotics, prewarmed to 37°C.

2.3. Site-directed Mutagenesis

1. Target plasmid DNA.
2. Custom mutagenic oligonucleotide pair (see Note 2).
3. AccuPrime™ Pfx DNA Polymerase (Life Technologies, Carlsbad, CA, cat # 12344024) or similar.
4. CpG Methyltransferase (Life Technologies, Carlsbad, CA, part of cat # A13282).
5. S-adenosyl methionine (Life Technologies, Carlsbad, CA, part of cat # A13282).
6. PCR enhancer (Life Technologies, Carlsbad, CA), part of cat # A13282.
7. In vitro recombination buffer (Life Technologies, Carlsbad, CA, part of cat # A13282) or equivalent.
8. In vitro recombination enzyme mix (Life Technologies, Carlsbad, CA, part of cat # A13282) or equivalent.
9. Deionized, sterile water.
10. 0.5 M EDTA.
11. One Shot® MAX Efficiency® DH5α-T1R competent cells (Life Technologies, Carlsbad, CA, part of cat # A13282 or cat # 18258012) or any competent cells wild type for the McrBC restriction-modification system.
12. S.O.C. medium (formulation above).
13. LB plates with the appropriate selection antibiotics, prewarmed to 37°C.

3. Methods

3.1. Yeast Cloning: DNA Assembly

The strategy relies on the reconstitution of circular plasmids by homologous recombination in yeast using linearized yeast vectors and fragments with overlapping ends. For precedents and previous publications using this approach see Note 3.

For guidelines on how to prepare the vector and inserts for recombination see Note 4.

The approach may be applied to recombine adjacent fragments that do not share end-homology (stitching oligonucleotides approach). For details on this strategy see Note 5.

A fundamental requirement for the Transformation-Associated Recombination (TAR) approach is a *Saccharomyces cerevisiae* replicating vector with a selectable marker. For that purpose we designed an *E. coli–S. cerevisiae* shuttle vector with high DNA capacity (for details of this vector see Fig. 1 and Note 6). Last, we generated a conversion cassette that enables the adaptation of any *E. coli* plasmid to this technology by simply incorporating the necessary yeast DNA replication and selection elements (see Fig. 1 and Note 7).

Typical results using this approach are shown in Fig. 2.

1. Add the following components to a microcentrifuge tube and mix:

Linearized vector	100 ng
DNA fragments	100 ng each (if final construct is ≤25 kb) 200 ng each (if final construct is >25 kb)
Stitching oligonucleotides (if necessary[a])	500 ng each (20 pmol each)

[a]For further details see Note 5.

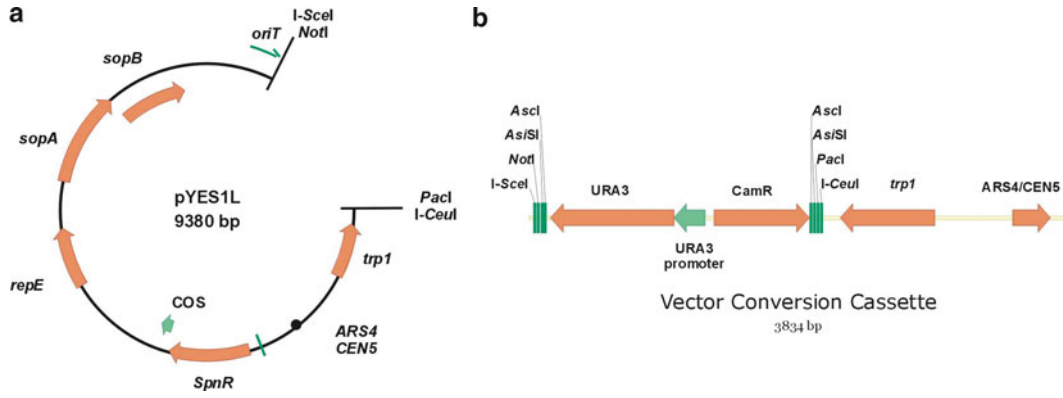


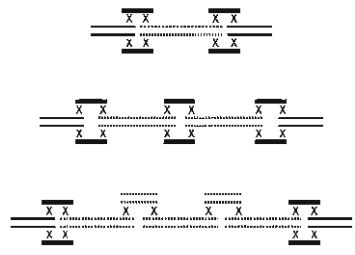
Fig. 1. Map of the constructs for yeast cloning. (**a**) Linearized BAC/YAC vector pYES1L. Circularization of the plasmid results in a functional *E. coli–S. cerevisiae* shuttle episome. (**b**) Scheme of the yeast adaptation cassette. For details and abbreviations see Notes 6 and 7.

a

No of fragments (excluding vector)	No x length of pre-cloned fragments (kbp)*	No x length of PCR-amplified fragments (kbp)	Total size excluding vector (kbp)	Overlap (bp)	Insert (ng)	Cloning Efficiency
3	3 x 10	0 x 0	30	80	100	100%
5	5 x 10	0 x 0	50	80	100	100%
10	10 x 10	0 x 0	100	80	100	50%
20	8 x 10	12 x 0.5-2.5	100	80	100	58%
20	8 x 10	12 x 0.5-2.5	100	80	200	83%
1	0 x 0	1 x 0.7	0.7	30	100	100%
1	0 x 0	1 x 10	10	80	200	100%
1	0 x 0	1 x 10	10	30	200	100%
10	0 x 0	10 x 5	50	30	200	92%

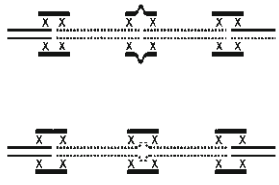
*[2] Fragments were initially cloned into pACYC184 and excised by *Not*1 digestion.

b



Fragment Size	Oligo Size	Cloning Efficiency
1 x 1 kbp	60mer	94%
2 x 5 kbp	80mer	75%
3x5 + 2x0.5 kbp	60mer	37%
3x5 + 2x0.5 kbp	80mer	75%

c



Fragment Size	Oligo Size	Cloning Efficiency
2 x 5 kbp	80mer 10 bp insertion	63%
2 x 5 kbp	80mer 20 bp insertion	50%
2 x 5 kbp	80mer 12 bp deletion	87%

Fig. 2. Yeast cloning. (**a**) Assemblies using fragments with end-homology. Experiments were performed using the protocol depicted in section 3.1. The recipient plasmid was pYES1L. The asterisk indicates that fragments were initially cloned into pACYC184 and excised by *Not*1 digestion. (**b**) Bridging DNA fragments with stitching oligonucleotides. One or more DNA fragments were bridged with the vector or adjacent fragment using double-stranded oligonucleotides perfectly complementary to the bridged molecules' ends. Hundred to two hundred nanograms of each DNA fragment plus 40 pmol of double-stranded oligonucleotides were transformed into MaV203 yeast competent cells and processed as indicated in the text. (**c**) Bridging and editing DNA fragments with stitching oligonucleotides, adjacent DNA fragments were bridged with imperfect double-stranded 80 bp oligonucleotides that generate insertions or deletions. DNA fragments are represented by *dotted lines*, vector (ends only) by *thin lines*, and oligonucleotides by *thick lines.*

If the total volume of the DNA mix is smaller than 10 μL, proceed to step 2.

If the total volume of the DNA mix is larger than 10 μL, reduce total volume to 5–10 μL using a SpeedVac® or a centrifugal filter device. Do not let the pellet dry completely.

2. Add 100 μL of 30°C thawed MaV203 cells into the DNA mix (the volume of the DNA mix should be ≤10 μL). Mix well by tapping the tube.
3. Add 600 μL of the PEG/LiAc solution to the DNA/competent cell mixture. Mix by inverting the tube 5–8 times until all of the components are homogeneous.
4. Incubate the mixture in the 30°C water bath for 30 min. Invert the tube occasionally (every 10 min) to resuspend the components.
5. Add 35.5 μL of DMSO to the tube. Mix by inverting the tube 5–8 times.
6. Heat-shock the cells by incubating the tube in the 42°C water bath for 20 min. Invert the tube occasionally to resuspend the components.
7. Centrifuge the tube at 1,800 rpm (200–400 × *g*) for 5 min.
8. Carefully discard the supernatant from the tube and resuspend the cell pellet in 1 mL of sterile 0.9% NaCl by gentle pipetting.
9. Plate 100 μL of the transformed cells onto CSM-Trp agar plates. For final constructs of >60 kb, we recommend that you centrifuge the remaining 900 μL of the transformation mixture, remove ~750 μL of the supernatant, resuspend the cell pellet in the remaining 100–150 μL of supernatant, and plate all cells onto another CSM-Trp agar plate to ensure that you have sufficient number of colonies to screen.
10. Incubate the cells at 30°C for 3 days and proceed to screening for the correct clone (see below).

3.2. Yeast Cloning: Screening for the Positive Clone

The fastest way to screen for yeast colonies containing the correct assembled construct is by performing colony-PCR assays using pair of diagnostic primers that amplify each single expected junction (for guidelines to design diagnostic oligonucleotides see Note 8)

1. Aliquot 15 μL of lysis buffer into PCR tubes or plates.
2. Pick individual yeast colonies one at a time using a sterile 20 μL pipette tip. Leave the tip in the PCR tube or the well until all the colonies have been picked.
3. Resuspend the cells by pipetting up and down 3 times.
4. Transfer 5 μL of each cell suspension into fresh PCR tubes and store at 4°C until verified that the colony is positive (see below).

5. Heat the remaining cells (10 μL) at 95°C for 5 min in a thermocycler and place them on ice. Briefly centrifuge the PCR tubes or plates to bring down condensed water.
6. Add 40 μL of deionized sterile water to each lysate and pipette up and down 3–5 times to mix.
7. Set up a PCR master mix for each junction and aliquot 49.5 μL of it into fresh PCR tubes or plates.
8. Add 0.5 μL of each diluted yeast lysate (from step 4) into each PCR tube or well. Do not exceed 0.5 μL of lysed yeast cells for 50 μL of PCR volume.
9. Vortex to mix the contents and briefly centrifuge to bring down all liquid.
10. Perform PCR cycling in a thermocycler.
11. Load 10 μL onto an agarose gel to visualize the PCR products. Sequencing these PCR products is recommended.

3.3. Yeast: E. coli Transfer

The protocol below usually yields 50–100 *E. coli* colonies containing the expected vector (up to 110 kbp) per 1 μL of lysed cells.

1. Aliquot 4–5 glass beads into a fresh PCR tube and add 10 μL of yeast lysis buffer.
2. Add 5 μL of the cell suspension that was stored into the Lysis Buffer/Glass beads mix (section 3.1). Pipette up and down 3–5 times to mix.
3. Vortex the cells at room temperature for 5 min. Do not heat the lysed cells.
4. Add 1 μL of the lysed cells (from step 3, above) into a vial of electrocompetent cells and mix gently. Do not add more than 1 μL of the lysed cells to avoid arcing during electroporation.
5. Transfer the cells to the chilled electroporation cuvette on ice.
6. Electroporate the cells following the manufacturer's recommended protocol.
7. Add 250 μL of prewarmed S.O.C. medium to each vial.
8. Transfer the solution to a 15-mL snap-cap tube and shake for at least 1 h at 37°C.
9. Spread 10–50 μL from each transformation on a prewarmed LB plate supplemented with the appropriate selection antibiotic.
10. Invert the selective plate(s) and incubate at 37°C overnight.

3.4. In Vitro DNA Assembly

The most important features of the method presented here are: (1) only 15 bp of end-homology is required between adjacent fragments; (2) it efficiently assembles multiple fragments; and (3) it readily works with standard transformation protocols allowing high-throughput cloning. For guidelines on how to prepare the vector and inserts for recombination (see Note 4). Typical assembly results are shown in Figs. 3 and 4.

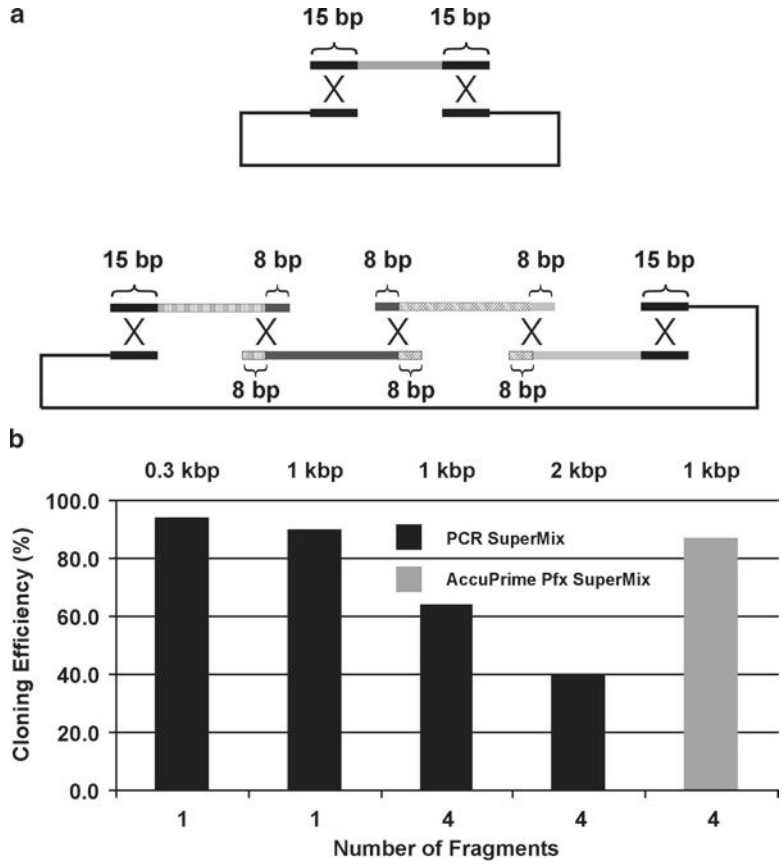


Fig. 3. In vitro DNA assembly. (**a**) Scheme of the approach. In this particular example, the vector (pUC19) was linearized with the restriction endonucleases *Kpn*I and *Pst*I. Indicated are the fragment overlaps used in each case. Identity among the fragments and the vector was generated by the addition of 15 nucleotides to the 5′ end of the oligonucleotides used to generate the first and last fragments. Identity between adjacent fragments was generated with 8 nt tails added to the oligonucleotides used to PCR amplify the fragments. Drawings are not to scale. (**b**) Fragments of the indicated size were amplified using a standard PCR polymerase (PCR SuperMix, Life Technologies, Carlsbad, CA) or a proofreading thermostable DNA polymerase (AccuPrime Pfx SuperMix, Life Technologies, Carlsbad, CA) and recombined into *Pst*I-*Kpn*I linearized pUC19 plasmid.

1. In a micro-centrifuge tube, add the components below in the order they are listed:

Insert(s)	20–200 ng each
Linearized vector	100 ng
5× reaction buffer[a]	4 μL
Deionized water	to 18 μL
10× enzyme mix[a]	2 μL

[a]As an example we list the in vitro recombination buffer and enzyme mix in the GENEART® Seamless Cloning and Assembly Kit (Life Technologies, Carlsbad, CA). Other kits are available from different commercial sources.

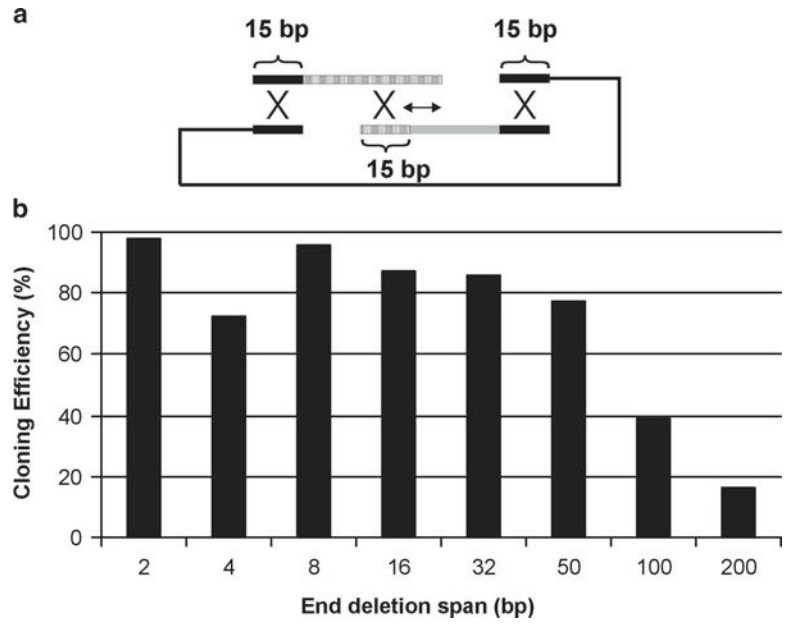


Fig. 4. In vitro junction editing. (**a**) Two DNA fragments were simultaneously recombined into a vector. One of the fragments shared 15 bp of homology with the other one at different distances from the end. (**b**) Cloning efficiencies from the experiments detailed in (**a**). Assemblies resulted in constructs bearing up to 200 bp deletions.

For optimum results use a 2:1-M ratio of insert:vector

2. Incubate at room temperature for 30 min.
3. Immediately use a 6–8-μL aliquot from the reaction above to transform 50 μL competent *E. coli* cells, following standard protocols. *Important*: do not use electrocompetent cells.
4. Plate the cells on LB agar plates with the corresponding antibiotics.

3.5. Site-directed Mutagenesis

With the advent of synthetic biology and rational design, the manipulation of genes to produce enzymes with subtle differences and the modification of promoters to finely tune metabolic flows, rely even more on robust site-directed mutagenesis approaches (for a summary of commercial site-directed mutagenesis approaches see Note 9).

We applied our homologous recombination approach (section 3.1) to join the ends of a single DNA molecule, thereby enabling a highly efficient site-directed mutagenesis strategy. The system relies on the inherent properties of a CpG methyltransferase, a high fidelity thermostable DNA polymerase, recombination enzymes, and the *E. coli* McrBC restriction-modification system. The DNA methylation and amplification steps are combined into

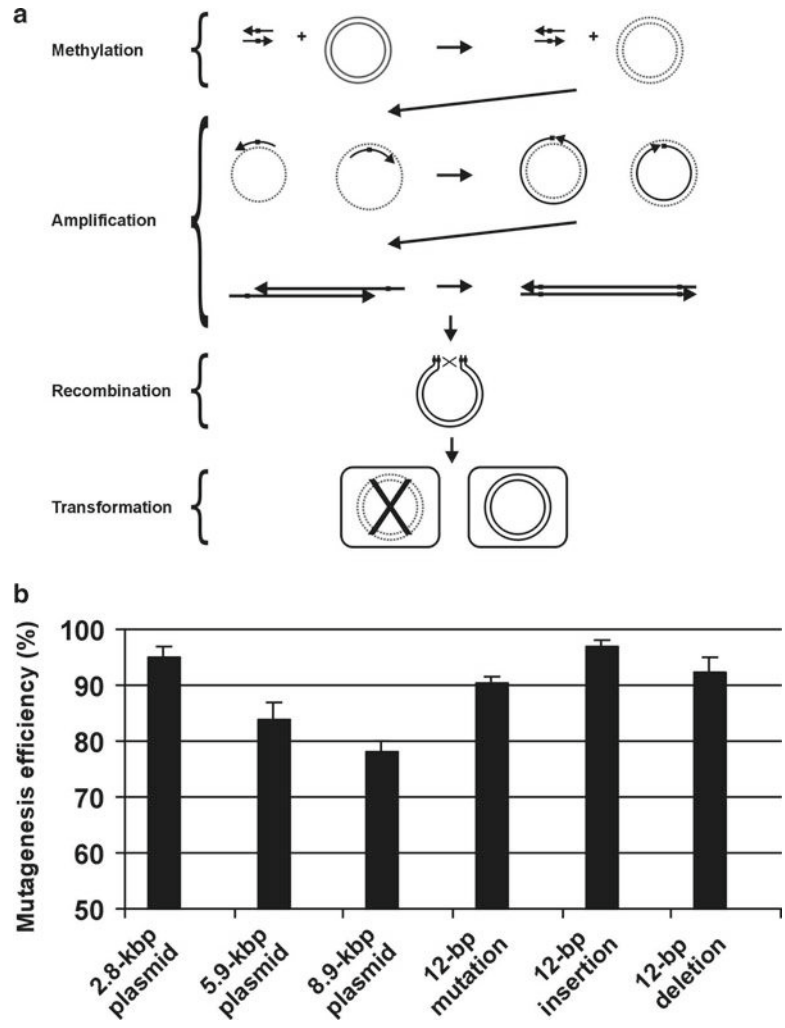


Fig. 5. Site-directed mutagenesis. (**a**) Strategy's workflow. Template strands are shown as *gray lines*. Methylated strands are shown as *gray dotted lines*. Oligonucleotides and new strands are shown in *black*. (**b**) Plasmids with frameshift mutations in the *lacZ*α gene were subjected to the mutagenesis protocol described in the text, transformed into DH5α-T1R cells, and then plated onto LB agar ampicillin X-gal plates. Primer pairs were designed to revert mutations to a wild-type *lacZ*α allele. The mutagenesis efficiency was calculated by the ratio of blue/total colonies.

a single reaction followed by a 10-min recombination step. This short in vitro recombination reaction of PCR products increases the colony output by three to tenfold. Finally, the products are transformed into a host strain that degrades the methylated DNA template (Fig. 5).

1. In a PCR tube, add the components below:

10× AccuPrime™ Pfx Reaction mix	1×
10× PCR enhancer	1×
Oligonucleotide pair	0.3 μM each
Plasmid DNA	20 ng
CpG DNA methyl transferase	4 U
S-adenosyl methionine	160 μM
AccuPrime™ Pfx	1 U
Deionized, sterile water	to 50 μL

2. Perform PCR using the following parameters:

37°C	12–20 min[a]
94°C	2 min
(a) 94°C	20 s
(b) 57°C	30 s
(c) 68°C	30 s/kb of plasmid 12–18 cycles of (a), (b), and (c)[b]
68°C	5 min
4°C	as needed

[a]Perform methylation of the plasmid at 37°C for 12–20 min. We recommend 12 min for 2.8–4 kb plasmids and 20 min for 4–14 kb plasmids.
[b]The cycling parameters specify a 30-s extension for each 1 kb of DNA. For optimal mutagenesis efficiency, we recommend 12–15 cycles for 2.8–4 kb plasmids and 18 cycles for 4–14 kb plasmids.

3. Analyze 5 μL of the product on a 0.8% agarose gel.
4. Setup the recombination reaction in a tube by adding the following reagents in the order below:

In vitro recombination buffer	1×
Deionized, sterile water	to final reaction volume of 20 μL
PCR sample	4 μL
In vitro recombination enzyme mix	1×

5. Mix well and incubate at room temperature for 10 min.
6. Stop the reaction by adding 1 μL 0.5 M EDTA. Mix well and place the tubes on ice.
7. Immediately transform *mcrBC*$^+$ *E. coli* cells, following standard protocols. *Important*: do not use electrocompetent cells.
8. Plate the cells on LB agar plates with the corresponding antibiotics

4. Notes

1. Advanced cloning systems for assembling multiple DNA fragments have been developed during the past 2 decades. Some of them such as RecA-independent recombination, Red/ET recombination, Gateway, or loxP-based are practical and fast but they work with a limited number of fragments (7–12). Other approaches, such as ligation-independent cloning, overlap PCR, SLIC, or the use of Type IIS restriction enzymes, require considerable design and/or preparation time (13–16).
2. For designing custom mutagenic oligonucleotide pair for site-directed mutagenesis,
   (a) Both primers (forward and reverse) should contain the desired mutation.
   (b) The mutation site should be centrally located on both primers and can be up to 12 bases (deletions, insertions, and/or any substitutions).
   (c) Both primers (forward and reverse) should be approximately 30–45 nucleotides in length, not including the mutation site. Primers longer than 45 nucleotides increase the likelihood of secondary structure formation, which may affect the efficiency of PCR amplification.
   (d) Primers should have an overlapping region at the 5′ ends of 15–20 nucleotides, for efficient end-joining of mutagenesis product.
   (e) For most applications, DNA oligonucleotides purified by desalting are generally sufficient, although oligonucleotides purified by HPLC or PAGE may increase the mutagenesis efficiency.
3. The approach described in section 3.1 relies on the yeast's powerful ability to take up and recombine DNA fragments (17, 18). It was originally described as Transformation-Associated Recombination (TAR) and used to generate yeast artificial chromosomes (YACs) (19–23), and bacterial chromosomes (5, 24–27). The method presented here is a derivative of that one described by Raymond and coworkers (1) (for important modifications to the Raymond method see Note 10). End-homology can be added manually to the fragments by PCR-amplifying the elements with oligonucleotides bearing additional sequences at their 5′ end, or can be exposed by excising the fragments with restriction endonucleases from larger DNA entities. Residual nucleotides derived from the original restriction site do not interfere and are readily eliminated during the recombination process (for important considerations on homologous sequences see Notes 4 and 11).

4. Guidelines for efficient DNA recombination
   (a) Fragments for assemblies can be generated by excising inserts from preexisting plasmids or by PCR amplification using oligonucleotides with 5′ overhangs that provide the required homology with the adjacent fragments. For assemblies performed in vitro, adjacent fragments must have an end-overlap of 15 bp. For assemblies performed in yeast, the overlap must be of 30 bp (for constructs up to 60 kbp) or 50 bp (for constructs larger than 60 kbp).
   (b) The PCR fragment size must be between 100 bp and 5 kbp. Larger fragments are more susceptible to damage in a gel extraction procedure. Furthermore, many PCR enzymes are not processive enough to amplify fragments >5 kb. Therefore, for fragments larger than 5 kbp, we recommend assembling multiple smaller fragments rather than trying to amplify a single large PCR molecule. After generating the inserts, PCR products must be verified by gel electrophoresis. If multiple bands are observed, the correct fragment must be purified by excising the band from an agarose gel.
   (c) Best cloning efficiencies are obtained if the total construct size (including the vector) does not exceed 13 kbp for in vitro assemblies, and 110 kbp for in vivo recombination.
   (d) Linearized cloning vectors may be prepared using restriction enzymes or by PCR amplification. Digestion with two restriction enzymes is a very efficient way to linearize the cloning vector. A double digest followed by PCR amplification virtually results in no cloning background.
   (e) When recombining a PCR-amplified fragment into a restriction-linearized vector, sequences contributing to the required homology must be added to the 5′ ends of the oligonucleotides. Alternatively, in adjacent PCR fragments the required homology can be obtained by splitting the necessary sequences into two. For example, the first half of the homologous sequences may be added to the reverse primer of fragment 1 and the second half to the forward primer of fragment 2 (for a visual representation see Fig. 3).
   (f) The cloning strategy, including the design of the DNA fragments and required oligonucleotides can be greatly facilitated using the free DNA Oligo Designer web tool (www.invitrogen.com/DesignDNAassembly).
5. The oligonucleotide stitching approach can be followed in those cases where adjacent DNA fragments do not share end-homology. In this case, the necessary homology is provided in

trans by complementary oligonucleotides that overlap both fragments, thereby serving as recombination linkers (stitching oligonucleotides) (28, 29). An important feature of the stitching oligonucleotide approach is that it allows editing the fragment junctions, thus generating required imperfections. The approach is particularly well suited for reusing fragments in a new sequence context, or for cloning DNA targets that cannot be readily amplified by PCR. It is also particularly useful when the removal of end sequences such as restriction sites or primer tails are required. It also permits the opposite type of alterations such as the addition of foreign sequences to insert restriction sites, small tag coding regions or small watermarks. Stitching oligonucleotides are readily designed by the free Oligo Designer Webtool (www.invitrogen.com/DesignDNAassembly) according to the following rules:

(a) Each junction between adjacent fragments requires two oligonucleotides for oligonucleotide stitching, a sense and an antisense oligonucleotide.

(b) Up to five fragments plus a vector can be assembled using stitching oligonucleotides, provided that not more than three junctions are formed by the stitching oligonucleotides and the remaining junctions are produced by shared end-terminal homology.

(c) Oligonucleotides used for oligonucleotide stitching of up to three nonhomologous fragments of <5 kb must be 80-mers (i.e., they must have a 40-bp overlap with each adjacent fragment).

(d) Prepare stitching oligonucleotide stocks at a final concentration of 100 μM in 1× TE buffer, pH 8.

(e) Stitching oligonucleotides used for insertion editing must have a 30-nucleotide overlap with each adjacent fragment in addition to the insertion bases (for a total length of up to 80-mer, including up to 20 insertion bases).

(f) Stitching oligonucleotides used for deletion editing must have a 40-nucleotide overlap with each adjacent fragment, annealing up to 6 nucleotides from the junction into each fragment, thus leaving up to 6 bp at the end of each fragment to be deleted during transformation-associated recombination.

6. Replication elements routinely used in yeast plasmids consist of either an autonomously replicating sequence (ARS) and a centromere (CEN) or replication and partitioning sequences from the endogenous 2-μm plasmid (32–34). The most common selectable markers are those that restore prototrophy for an essential metabolite in auxotrophic cells. Our shuttle vector that has ARS4 and CEN5 as replication elements plus the *trp1*

gene encoding the yeast phosphoribosylanthranilate isomerase as a selectable marker that complements strains deficient in the synthesis of tryptophan such as those harboring homozygous *trp1-901* alleles (Fig. 1a). In order to maximize DNA capacity the F' replication origin and accessory genes from the mini-F plasmid (reviewed by ref. (35)) were cloned. Other elements such as a spectinomycin resistance gene (SpnR), an origin of transfer (*oriT*) for plasmid mobilization, and the cos site from bacteriophage λ were also included (Fig. 1).

7. We designed a linear adaptation cassette with all the features necessary for yeast cloning and replication (Fig. 1b). The fragment contains all the yeast-related features described for the plasmid pYES1L plus the *ura3* gene encoding yeast's orotidine 5-phosphate decarboxylase (36). This counter-selectable marker becomes lethal when 5-fluoroorotic acid (5-FOA) is added to the media, as it converts it into the toxic compound 5-fluorouracil. During development we learned that the use of this feature is not really necessary, as the frequency of plasmid recircularization is negligible. The cassette contains also the bacterial chloramphenicol acetyl transferase (or chloramphenicol resistance gene, CamR) which facilitates plasmid adaptation by selecting recombinants in agar plates containing chloramphenicol and the corresponding antibiotic specific to the vector backbone. Convenient rare restriction sites were added in order to linearize the final adapted molecule for cloning in yeast.
8. We recommend designing oligonucleotide pairs (forward and reverse) at a distance of 100–250 bp from the ends of each DNA fragment (including the cloning vector) so that the colony-PCR products would be 200–500 bp in size and span the junctions between the fragments.
9. Site-directed mutagenesis kits commercially available use, at least, one of the following approaches: (1) the isolation of single strand template DNA and the generation of the mutation using one complementary primer (37); (2) the design of two sets of PCR primers that overlap the mutation site, the amplification of the template by two PCR reactions, and then the cloning of the two PCR fragments and the vector by three piece ligation (38); (3) the PCR amplification of a plasmid using complementary oligonucleotides and the subsequently elimination of the template molecule (39, 40).
10. Features that distinguish the approach described in (section 3.1) from previously described ones include:
    (a) The yeast strain used in this method is MaV203 (MATα *leu2-3,112 trp1-901*; *his3Δ200*; *ade2-101*; *cyh2R*; *can1R*; *gal4Δ*; *gal80Δ*; *GAL1::lacZ SPAL10::URA3 HIS3*$_{UASGAL1}$*:: HIS3@LYS2)*, a derivative from a cross between two nonisogenic strains, PCY2 and MaV99 (30, 31).

(b) This method uses *chemically* competent yeast cells.

(c) Recommended overlaps between fragments are 30 bp for constructs <60 kbp, and 50 bp for constructs >60 kbp.

11. The recombinogenic properties of *S. cerevisiae* strongly promotes DNA recombination between any pair of highly homologous sequences 50 bp or larger. Therefore, we recommend verifying that the fragment sequences do not share patches of DNA identity other than those at the end of adjacent fragments. To avoid using fragments with unwanted similarity we internally blast our sequences using the Oligo Designer Webtool (www.invitrogen.com/DesignDNAassembly), a free online software that detects potential patches of internal homology, designs oligonucleotides for recombination, and provides a GenBank annotated sequence of the assembled construct.

Disclosure

Products are For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

References

1. Raymond C. K., Pownder T. A. and Sexson S. L. (1999) General method for plasmid construction using homologous recombination. *Biotechniques* **26**, 134–8, 140–1
2. Willer D. O., Yao X. D., Mann M. J. and Evans D. H. (2000) In vitro concatemer formation catalyzed by vaccinia virus DNA polymerase. *Virology* **278**, 562–9
3. Hamilton M. D., Nuara A. A., Gammon D. B., Buller R. M. and Evans D. H. (2007) Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. *Nucleic Acids Res* **35**, 143–51
4. Zhu B., Cai G., Hall E. O. and Freeman G. J. (2007) In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* **43**, 354–9
5. Gibson D. G., Benders G. A., Andrews-Pfannkoch C., Denisova E. A., Baden-Tillson H., Zaveri J., et al. (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215–20
6. Gibson D. G., Young L., Chuang R. Y., Venter J. C., Hutchison C. A., 3rd and Smith H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–5
7. Cheo D. L., Titus S. A., Byrd D. R., Hartley J. L., Temple G. F. and Brasch M. A. (2004) Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: functional analysis of multi-segment expression clones. *Genome Res* **14**, 2111–20
8. Hartley J. L., Temple G. F. and Brasch M. A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res* **10**, 1788–95
9. Bubeck P., Winkler M. and Bautsch W. (1993) Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res* **21**, 3601–2
10. Zhang Y., Buchholz F., Muyrers J. P. and Stewart A. F. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**, 123–8
11. Datsenko K. A. and Wanner B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97**, 6640–5
12. Liu Q., Li M. Z., Leibham D., Cortez D. and Elledge S. J. (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr Biol* **8**, 1300–9
13. Lebedenko E. N., Birikh K. R., Plutalov O. V. and Berlin Yu A. (1991) Method of artificial

DNA splicing by directed ligation (SDL). *Nucleic Acids Res* **19**, 6757–61

14. Aslanidis C. and de Jong P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**, 6069–74
15. Li M. Z. and Elledge S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4**, 251–6
16. Gao X., Yo P., Keith A., Ragan T. J. and Harris T. K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res* **31**, e143
17. Orr-Weaver T. L., Szostak J. W. and Rothstein R. J. (1981) Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci USA* **78**, 6354–8
18. Hinnen A., Hicks J. B. and Fink G. R. (1978) Transformation of yeast. *Proc Natl Acad Sci USA* **75**, 1929–33
19. Larionov V., Kouprina N., Eldarov M., Perkins E., Porter G. and Resnick M. A. (1994) Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast* **10**, 93–104
20. Ma H., Kunes S., Schatz P. J. and Botstein D. (1987) Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201–16
21. Marykwas D. L. and Passmore S. E. (1995) Mapping by multifragment cloning in vivo. *Proc Natl Acad Sci USA* **92**, 11701–5
22. Ebersole T., Okamoto Y., Noskov V. N., Kouprina N., Kim J. H., Leem S. H., et al. (2005) Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. *Nucleic Acids Res* **33**, e130
23. Larionov V., Kouprina N., Graves J., Chen X. N., Korenberg J. R. and Resnick M. A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc Natl Acad Sci USA* **93**, 491–6
24. Gibson D. G. (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* **37**, 6984–90
25. Lartigue C., Vashee S., Algire M. A., Chuang R. Y., Benders G. A., Ma L., et al. (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–6
26. Gibson D. G., Benders G. A., Axelrod K. C., Zaveri J., Algire M. A., Moodie M., 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 USA* **105**, 20404–9
27. Gibson D. G., Glass J. I., Lartigue C., Noskov V. N., Chuang R. Y., Algire M. A., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–6
28. DeMarini D. J., Creasy C. L., Lu Q., Mao J., Sheardown S. A., Sathe G. M., et al. (2001) Oligonucleotide-mediated, PCR-independent cloning by homologous recombination. *Biotechniques* **30**, 520–3
29. Raymond C. K., Sims E. H. and Olson M. V. (2002) Linker-mediated recombinational subcloning of large DNA fragments using yeast. *Genome Res* **12**, 190–7
30. Vidal M., Brachmann R. K., Fattaey A., Harlow E. and Boeke J. D. (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc Natl Acad Sci USA* **93**, 10315–20
31. Chevray P. M. and Nathans D. (1992) Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc Natl Acad Sci USA* **89**, 5789–93
32. Struhl K., Stinchcomb D. T., Scherer S. and Davis R. W. (1979) High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc Natl Acad Sci USA* **76**, 1035–9
33. Clarke L. and Carbon J. (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287**, 504–9
34. Murray J. A. (1987) Bending the rules: the 2-mu plasmid of yeast. *Mol Microbiol* **1**, 1–4
35. Kline B. C. (1985) A review of mini-F plasmid maintenance. *Plasmid* **14**, 1–16
36. Boeke J. D., LaCroute F. and Fink G. R. (1984) A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**, 345–6
37. Hutchison C. A., 3rd, Phillips S., Edgell M. H., Gillam S., Jahnke P. and Smith M. (1978) Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **253**, 6551–60
38. Stemmer W. P. and Morris S. K. (1992) Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. *Biotechniques* **13**, 214–20
39. Kunkel T. A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA* **82**, 488–92
40. Hemsley A., Arnheim N., Toney M. D., Cortopassi G. and Galas D. J. (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res* **17**, 6545–51

Chapter 9

Ethanol-Tolerant Gene Identification in *Clostridium thermocellum* Using Pyro-Resequencing for Metabolic Engineering

Shihui Yang, Dawn M. Klingeman, and Steven D. Brown

Abstract

Classic strain development that combines random mutagenesis and selection has a long history of success in generation of biocatalysts with industrially designed traits. However, the genetic loci contributing to the phenotypic strain changes are difficult to identify prior to genome sequencing technology advancement. In this chapter, we present the approach using Roche 454 next-generation pyro-resequencing to identify the genotypic changes such as single nucleotide polymorphisms (SNP) associated with an ethanol-tolerant strain of *Clostridium thermocellum*. The parameters used to filter the pyro-resequencing output for SNP identification are also discussed. These can help researchers to identify the genotypic change of other biocatalysts for strain improvement through metabolic engineering.

Key words: 454 pyrosequencing, Next-generation sequencing, Single nucleotide polymorphism, Genotyping, *Clostridium thermocellum*, Biofuel, Consolidated bioprocessing

1. Introduction

Genome sequencing provides opportunities for fundamental insights to genetic makeup that will greatly facilitate strain development (1). We have recently proposed a paradigm to identify and characterize the process-relevant traits created by classical strain development through the integration of systems biology and next-generation sequencing (NGS) approaches with genetics tools (2). With the development of new technology, NGS could deliver fast, inexpensive, and accurate genome information (see recent reviews (3–5)) for metabolic modeling and engineering (6).

The first NGS platform was commercialized by Roche 454 in 2005, which measures the release of inorganic pyrophosphate by

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_9,

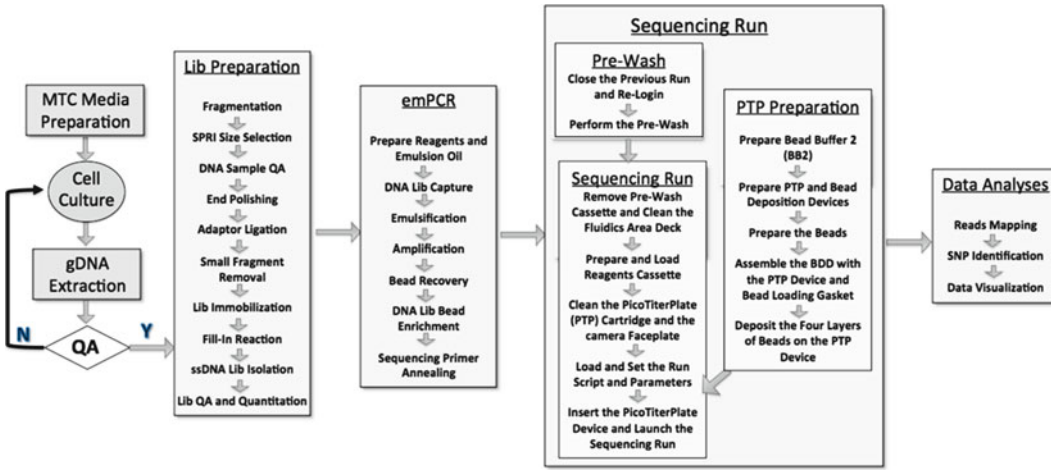


Fig. 1. Workflow to identify the genetic changes associated with mutant strain for metabolic engineering using the Roche 454 next-generation sequencing technique. *gDNA* genomic DNA; *Lib* library; *QA* quality assessment; *ssDNA* single strand DNA.

converting it into visible light during DNA synthesis and includes steps of DNA library preparation, emulsion PCR, DNA sequencing, and data analysis (3–5, 7) (see Fig. 1). The Illumina Genome Analyzer is currently the most widely used NGS platform, which is based on the conventional Sanger DNA sequencing chemistry using four color cyclic reversible terminators that are blocked at the 3′ end by modified nucleotides such as 3′-*O*-azidomethyl dNTPs (3–5, 8). It has three major steps of DNA library preparation, clonal clusters generation, and sequencing. Other NGS platforms include the ligation based SOLiD™ system from Applied Biosystems and the single molecule sequencing platform of Helicos® Genetic Analysis System from Helicos BioSciences Corporation (3, 4). In addition, the third generation sequencing techniques such as PacBio RS system from Pacific Biosciences and Nanopore Sequencing techniques licensed by companies like Oxford Nanopore Technologies and NobleGen Biosciences from Harvard, UCSC, and other universities are coming online to directly sequence single RNA molecules and identify genome-wide methylation in massively parallel formats (3, 4).

The goal of all new sequencing technologies is to increase throughput and yield while reducing cost. Currently, the average read length of a Roche 454 titanium sequencing run is about 400 bp, which will be extended up to 1,000 bp in the future. The 454 sequencing run takes about 10 h and can sequence eight samples per run with the capability for more samples with barcoding. Illumina and SOLiD™ platforms take 3–6 days run time and the reads are shorter (less than 100 bp) than 454 ones. However, both Illumina and SOLiD™ can generate hundreds of times more reads than 454 (about 100–400 million reads per run) for extensive coverage. Each platform can deliver very accurate results with

minimal sequencing coverage of at least 10–15 fold (9) although each platform also has its limitations (3, 4). For example, Roche 454 has a high indel rate in homopolymers, and Illumina has errors for detection of substitutions, especially after a "G" base (5, 10). Amplification bias during template preparation could also cause underrepresentation of AT-rich and GC-rich regions for Illumina system (3–5, 11, 12).

Clostridium thermocellum is a thermophilic, obligate anaerobic Gram-positive bacterium that is a consolidated bioprocessing (CBP) biocatalyst candidate for converting cellulosic biomass directly into ethanol. Wild-type strains produce ethanol as well as organic acids and growth is inhibited at relatively low ethanol concentrations (5 g/L) (13, 14). Ethanol-tolerant mutants of *C. thermocellum* (EA) have previously been isolated to tolerate ethanol concentrations up to 80 g/L (15), but the mechanism of tolerance has remained elusive. In this chapter, we present details using next-generation sequencing methods (2) to identify genetic changes in an ethanol-tolerant mutant (EA) of *C. thermocellum* ATCC27405.

2. Materials

2.1. Strains and Medium

1. Strain: wild-type *C. thermocellum* ATCC27405 was obtained from the American Type Culture Collection, and the ethanol-tolerant mutant (EA) strain was obtained from Dr. Herbert J. Strobel at University of Kentucky, and has been described previously to be able to tolerate 50 g/L ethanol consistently (15).
2. MTC Medium (see Note 1) with cellobiose: preparation see Subheading 3.1.

   Solution A: 6.24 g/L Cellobiose (Final concentration: 5.0 g/L), resazurin (0.001 g/L, using 20 g/L stock solution).

   Solution B Stock (25×): 50 g/L potassium citrate monohydrate ($C_6H_5O_7K_3$), 31.25 g/L citric acid monohydrate ($C_6H_8O_7{\cdot}H_2O$), 25 g/L sodium sulfate anhydrous (Na_2SO_4), 25 g/L potassium phosphate monobasic (KH_2PO_4), and 62.5 g/L sodium bicarbonate ($NaHCO_3$).

   Solution C stock (50×): 75 g/L ammonium chloride (NH_4Cl), 100 g/L urea (CH_4N_2O).

   Solution D stock (50×, anaerobic): 50 g/L magnesium chloride hexahydrate ($MgCl_2{\cdot}6H_2O$), 10 g/L calcium chloride dihydrate ($CaCl_2{\cdot}2H_2O$), 5 g/L ferrous chloride tetrahy-

drate ($FeCl_2 \cdot 4H_2O$), and store at room temperature. Add 50 g/L L-cysteine HCl monohydrate ($C_3H_7NO_2S \cdot HCl \cdot H_2O$) powder and autoclave just before use.

Solution E Vitamin stock (50×): 1 g/L pyridoxol hydrochloride, 0.05 g/L riboflavin, 0.05 g/L nicotinamide, 0.025 g/L lipoic acid, 0.2 g/L 4-aminobenzoic acid, 0.2 g/L d-biotin, 0.00125 g/L folic acid, 0.1 g/L vitamin B12, and 0.2 g/L thiamine HCl.

Solution F stock (1,000×): 0.5 g/L manganese chloride tetrahydrate ($MnCl_2 \cdot 4H_2O$), 0.5 g/L cobalt chloride hexahydrate ($CoCl_2 \cdot 6H_2O$), 0.2 g/L zinc sulfate heptahydrate ($ZnSO_4 \cdot 7H_2O$), 0.05 g/L copper sulfate pentahydrate ($CuSO_4 \cdot 5H_2O$), 0.05 g/L boric acid (H_3BO_3), 0.05 g/L sodium molybdate dihydrate ($Na_2MoO_4 \cdot 2H_2O$), 0.05 g/L nickel (II) chloride hexahydrate ($NiCl_2 \cdot 6H_2O$), and 10 g/L citric acid monohydrate ($C_6H_8O_7 \cdot H_2O$).

3. *Glycerol stock solution* (230 mL): 115 mL H_2O, 115 mL glycerol, 0.2 g cysteine, 0.05 g ammonium chloride, 0.04 g sodium chloride, 0.05 g sodium bicarbonate, and 1 mL MTC solution.
4. Serum bottle and bottle stopper.

2.2. General Chemical and Equipment

1. PCR machine.
2. NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc., MA).
3. Heating block (1.5 mL tubes).
4. Isopropanol.
5. Ethanol.
6. 50 mM EDTA (pH 8.0).
7. 10 mg/mL lysozyme.
8. 10N NaOH.

2.3. 454 Titanium Shot-Gun Pyrosequencing

2.3.1. Library Preparation

1. GS FLX Titanium Library Preparation Kit (Roche 05 233 747 001).
2. Compressed Nitrogen (Research Grade 5.0) using a dual-stage regulator (TechAir 2123351-000).
3. Qiagen MinElute PCR purification kit (Qiagen 28004).
4. AMPure XP beads (Beckma Coulter A63880).
5. Agilent Bioanalyzer 2100 and lapchip kits (Agilent 5067-1513, 5067-1506, 5067-1504).
6. QuantiFluor™-ST Fluorometer (Promega, CA) or a 96-well Plate Fluorometer.
7. Magnetic particle concentrator (MPC) (Invitrogen 123 21D).
8. 3M sodium acetate pH 5.2 (Sigma S-7899).

2.3.2. emPCR

1. Qiagen Tissue Lyser.
2. GS Titanium emPCR kits: Small (Roche 05618444001) and Large volume (Roche 0618428001).
3. GS Titanium emPCR filters SV 64 pc (Roche 05233674001).
4. GS Titanium emPCR breaking Kits (Roche 05233658001).
5. Beckman Coulter Bead Counter.
6. Blunt, flat tip needle, 16 gauge.
7. 10 mL Disposable Syringe with Luer-Lok.

2.3.3. Sequencing

1. Thermolyme lab quake rotator (Thermolyne 400110/400220).
2. The GS FLX Titanium Sequencing Kit XLR70 (Roche 05233526 001).
3. GS FLX Titanium PicoTiterPlate (PTP) Kit 70×75 (Roche 05233682 001).
4. Zeiss Moistened Cleaning Tissue.
5. Katadyn Micropur MP1 tablets.
6. Sparkleen soap solution.
7. Maintenance wash kits (Roche 04932358001).

3. Methods

3.1. Anaerobic MTC Medium Preparation

For each 50 mL anaerobic MTC in serum bottle:

1. Add 40 mL Solution A into a serum bottle, autoclave at 121°C for 15 min; sparge 15 min or with nitrogen gas while hot, cool.
2. Add 50 g/L L-cysteine HCl monohydrate powder to Solution D, cap and crimp, autoclave at 121°C for 15 min just before use (can be autoclaved with the above serum bottles with 40 mL Solution A at the same time).
3. Make cocktail containing 5 mL 5% MOPS, 2.4 mL Solution B, 1.2 mL Solution C, 1.2 mL Solution E, and 0.05 mL Solution F.
4. When Solution D cools, add 1.2 mL Solution D (with L-cysteine HCl monohydrate added and autoclaved) into above cocktail. Cap, crimp, and mix well.
5. Add 10 mL cocktail containing Solution D from above step to each serum bottle 40 mL.
6. Sparge again with nitrogen gas for 10 min, the medium should be reduced completely as colorless or very light yellow color broth, store at 4°C.

3.2. Genomic DNA Extraction

The method is based on the technical manual of Promega Wizard® Genomic DNA

Purification kit.

1. Add 1 mL overnight culture to a 1.5 mL microcentrifuge tube.
2. Centrifuge at 13,000 × *g* for 2 min to pellet the cells. Remove the supernatant.
3. Resuspend the cells thoroughly in 480 μL of 50 mM EDTA.
4. Add lysozyme to the resuspended cell pellet in a total volume of 120 μL, and gently pipette to mix.
5. Incubate the sample at 37°C for 30–60 min. Centrifuge for 2 min at 13,000 × *g* and remove the supernatant.
6. Add 600 μL of Nuclei Lysis Solution. Gently pipette until the cells are resuspended.
7. Incubate at 80°C for 5 min to lyse the cells, then cool to room temperature.
8. Add 3 μL of RNase Solution to the cell lysate. Invert the tube 2–5 times to mix.
9. Incubate at 37°C for 15–60 min. Cool the sample to room temperature.
10. Add 200 μL of Protein Precipitation Solution to the RNase-treated cell lysate. Vortex vigorously at high speed for 20 s to mix the Protein Precipitation Solution with the cell lysate.
11. Incubate the sample on ice for 5 min.
12. Centrifuge at 13,000–16,000 × *g* for 3 min.
13. Transfer the supernatant containing the DNA to a clean 1.5 mL microcentrifuge tube containing 600 μL of room temperature isopropanol.
14. Gently mix by inversion until the thread-like strands of DNA form a visible mass.
15. Centrifuge at 13,000–16,000 × *g* for 2 min.
16. Carefully pour off the supernatant and drain the tube on clean absorbent paper (see Note 2). Add 600 μL of room temperature 70% ethanol and gently invert the tube several times to wash the DNA pellet.
17. Centrifuge at 13,000–16,000 × *g* for 2 min. Carefully aspirate the ethanol.
18. Drain the tube on clean absorbent paper and allow the pellet to air-dry for 10–15 min.
19. Add 100 μL of DNA Rehydration Solution to the tube and rehydrate the DNA by incubating at 65°C for 1 h. Periodically mix the solution by gently tapping the tube (see Note 3).

20. Quantify the DNA using NanoDrop Spectrophotometer, and check the DNA quality by running about 50 ng in a 1% agarose gel.
21. If the DNA has a 260/280 and 260/230 ratio greater than 1.7 and no smear in the gel, store the DNA at –20°C for future use.

3.3. Library Preparation for Titanium Shot-Gun Pyrosequencing

The method is based on the technical manuals of 454 GS FLX Titanium General Library Preparation Method Manual April 2009. Also see Notes 4 and 5 for general guidance.

3.3.1. DNA Fragmentation Using Nebulization Method

1. Dilute 3–5 μg of sample DNA (see Notes 6 and 7) with TE Buffer to a final volume of 100 μL, assemble nebulizer and then add 500 μL of Nebulization Buffer with the 100 μL diluted DNA, and mix thoroughly by swirling or pipetting up and down.
2. Direct 30 psi (2.1 bar) of nitrogen through the Nebulizer for 1 min to nebulize the DNA.
3. Collect and measure the volume of nebulized material with a micropipette (see Note 8).
4. Add 2.5 mL of Qiagen's Buffer PBI directly into the Nebulizer cup, and swirl to collect all material droplets and mix the sample.
5. Purify the nebulized DNA using two Qiagen MinElute PCR Purification Kit columns.
6. Do not use any extra PBI buffer.
7. Multiple spins may be necessary to load sample.
8. After the PE dry spin rotate the column 180° and spin an additional 30 s.
9. Elute each column with 10 μL of room temperature EB.
10. Pool elutes for a total volume of ~20 μL.

3.3.2. Select DNA Fragment Size Using Double SPRI Method

1. Measure the volume of pooled elutes from Subheading 3.3.1, step 10.
2. Add Buffer EB (Qiagen) to a final volume of 100 μL.
3. Vortex the AMPure beads (see Notes 9 and 10) to achieve a uniform suspension then transfer the amount of AMPure beads appropriate for the Double SPRI method as determined per calibration of the lot in use to the nebulized and purified sample (100 μL), and vortex to mix.
4. Incubate for 5 min at room temperature.
5. Pellet the beads with the MPC for several minutes and leave the tube in the MPC for all wash steps.

6. Remove and discard the supernatant and wash the beads twice with 500 µL of 70% Ethanol, incubating for 30 s each wash.
7. After the second wash remove and discard all the supernatant and allow the AMPure beads to air-dry completely (cracks in the bead pellet are visible when the beads are dry).
8. Remove the tube from the MPC, and add 24 µL of 10 mM Tris–HCl, pH 8.0 (Qiagen's EB).
9. Vortex to resuspend the beads and elute the nebulized DNA from the AMPure beads.
10. Place the tube in the MPC to pellet the beads.
11. Transfer and keep the supernatant containing the purified nebulized DNA in a fresh microcentrifuge tube.

3.3.3. Assess DNA Sample Quality

1. Run 1 µL of the size-selected material on an Agilent 2100 DNA 1000 LabChip.
2. If the DNA meets the criteria (see Note 11), proceed to next steps.

3.3.4. Polish Fragment Ends

1. In a microcentrifuge tube, add reagents in the following order to a final volume of 50 µL: ~23 µL nebulized size-selected DNA fragments, 5 µL 10× Polishing Buffer, 5 µL BSA, 5 µL ATP, 2 µL dNTP Mix, 5 µL T4 PNK, and 5 µL T4 DNA polymerase.
2. Mix well and incubate at 12°C for 15 min, and then immediately continue incubation at 25°C for an additional 15 min.
3. Purify the fragments using 250 µL of Qiagen PBI Solution and one Qiagen MinElute PCR Purification Kit column.
4. After adding PE Solution, dry spin rotate the column 180° and spin an additional 30 s.
5. Elute the sample in 10 µL of room temperature Buffer EB (Qiagen).

3.3.5. Ligate Adaptor to the Polished DNA

1. In a microcentrifuge tube, add reagents in the following order to a final volume of 40 µL: ~10 µL Polished DNA, 20 µL 2× Ligase Buffer, 5 µL Adaptors, and 5 µL Ligase.
2. Mix well, spin briefly, and incubate the ligation reaction at 25°C for 15 min.
3. Purify the ligation products using 200 µL of PBI (Qiagen) and one column from Qiagen MinElute PCR Purification Kit eluting in 50 µL of room temperature Buffer EB (Qiagen).
4. After adding PE Solution, dry spin rotate the column 180° and spin an additional 30 s.
5. Elute in 50 µL of room temperature Buffer EB (Qiagen).

3.3.6. Remove Small DNA Fragments

1. Remove small DNA fragments using AMPure bead purification as described in Subheading 3.3.2.
2. Elute in 25 μL of 10 mM Tris–HCl, pH 8.0 (Qiagen's EB).

3.3.7. Immobilize Library to Magnetic Streptavifin-Coated Beads

1. Transfer 50 μL of Library Immobilization Beads to a fresh 1.5 mL tube.
2. Pellet the beads and remove the buffer using the MPC.
3. Wash the Library Immobilization Beads twice with 100 μL of 2× Library Binding Buffer using the MPC.
4. Resuspend the beads in 25 μL of 2× Library Binding Buffer.
5. Add 25 μL purified ligated DNA to the tube of washed Library Immobilization Beads.
6. Mix well and place on a tube rotator at room temperature (22°C) for 20 min.
7. Using the MPC, wash the immobilized Library twice with 100 μL of Library Wash Buffer.
8. Remove all remaining Library Wash Buffer from the pelleted beads, and remove the tube from the MPC.

3.3.8. Fill-In Gaps at the Non-Phosphorylated Ends of the Roche Provided A/B Adaptors

1. In a microcentrifuge tube, add reagents in the following order to a final volume of 50 μL: 40 μL water, 5 μL 10× Fill-in Polymerase Buffer, 2 μL dNTP Mix, and 3 μL Fill-in Polymerase.
2. Add the 50 μL fill-in reaction mix prepared in step 1 to the tube containing the library-carrying beads; mix well and incubate at 37°C for 20 min.
3. Wash the immobilized Library twice with 100 μL Library Wash Buffer using the MPC.
4. Remove all remaining Library Wash Buffer from the pelleted beads, and remove the tube from the MPC.

3.3.9. Single-Stranded DNA Library Isolation

1. Add 50 μL Melt Solution (125 μL of 10N NaOH in 9.875 mL Water, see Note 12) to the washed library-carrying beads.
2. Vortex well and pellet the beads away from the 50 μL supernatant using the MPC.
3. Carefully remove and transfer the supernatant to the freshly prepared neutralization solution (500 μL Qiagen's PBI buffer with10 μL 3M Sodium Acetate at pH 5.2).
4. Repeat above steps, and pool the two Melt Solution washes together in the same tube of neutralization solution.
5. Purify the neutralized DNA library using one column from Qiagen MinElute PCR Purification Kit. Do not use additional PBI.

6. Perform two PE washes, dry spin rotate the column 180° and spin an additional 30 s.
7. Elute with 15 μL room temperature TE Buffer (from GS FLX Titanium General Library Preparation Kit).

3.3.10. DNA Library Quality Assessment and Quantitation

1. Run 1 μL DNA library on a Bioanalyzer RNA Pico 6000 LabChip to assess the library quality.
2. Quantitate the DNA library (1 μL, in triplicate) by fluorometry using the Invitrogen RiboGreen method.
3. If the DNA meets the criteria (see Note 11), and is ≥5 ng proceed to next steps.

3.3.11. Library Primary Dilution and Storage

1. Determine the amount of input DNA that gives an enriched bead yield of approximately 8% as determined in a emulsion titration assay (See GS FLX Titanium General Library preparation Method Manual and emPCR Method Manual-Lib-L SV) (October 2009 (Rev. Jan 2010)).
2. Store the library in aliquots at −15 to −25°C.

3.4. emPCR

The method is based on the technical manuals of Lib-L Large Volume (LV), GS FLX Titanium Series October 2009 (Rev. Jan 2010), the written preparation here is for a 70×75 2 region plate.

3.4.1. emPCR Preparation for Two Cups

1. Thaw, vortex, and centrifuge all reagents except for the enzymes which should be left at −20°C.
2. Prepare the emPCR Additive by heating the tube at 55°C for 5 min then centrifuge to remove remaining particulates.
3. Prepare Emulsion Oil (see Notes 13 and 14) by shaking at 28 Hz for 2 min on the TissueLyser.
4. Dilute 2 mL of the 5× Mock Amplification Mix with 8 mL water and vortex.
5. Add 5 mL of Mock Amplification Mix dilution to each cup of Emulsion Oil.
6. Invert the cup 2–3 times and shake on the TissueLyser at 28 Hz for 5 min.

3.4.2. Live Amplification Mix preparation

1. In a tube, add reagents for the Live Amplification Mix in the following order to a final volume of 3,915 μL: 1,200 μL water, 1,500 μL emPCR additive, 780 μL 5× Amplification Mix, 230 μL Amplification Primer, 200 μL emPCR Enzyme Mix, and 5 μL PPiase.
2. Vortex the Live Amplification Mix for 5 s and store on ice (see Note 15).

3.4.3. DNA Library Capture

1. Prepare 1× Capture Bead Wash Buffer TW by mixing 1 mL 10× Capture Bead Wash Buffer TW with 9 mL Water.
2. Vortex the tubes of DNA Capture Beads.
3. Pellet the beads in a bench top minifuge by spinning for 10 s, rotating the tube 180°, and spinning again for 10 s.
4. Carefully remove and discard the supernatant without disturbing the bead pellet.
5. Wash each tube of beads *twice* with 1 mL 1× Capture Bead Wash Buffer TW. Vortex to resuspend the beads, spin, and discard the supernatant after each wash.
6. Thaw an aliquot of the DNA library to be amplified.
7. To each tube of washed DNA Capture Beads, add the correct volume of the DNA library that will provide optimal amplification (e.g., per titration) to the bead pellet.
8. Vortex the tubes for 5 s to mix their contents.
9. Prepare the DNA Capture Beads with library DNA mixes for individual emulsion reactions.
10. Transfer the captured library mixes from their DNA Capture Beads tubes to 15 mL tubes.
11. Add 3.75 mL of the Live Amplification Mix to each 15 mL tube.
12. From there, pipette 1 mL of the mix to the DNA Capture Beads tubes that contained the captured library, vortex, and transfer back the solution into the 15 mL tubes.
13. Repeat two times and then vortex.

3.4.4. Emulsify to Create Tiny Microreactors a nd Amplify

1. Pour the captured library into a prepared emulsion cup, invert the cups 3 times to mix.
2. Shake in the TissueLyser at 12 Hz for 5 min.
3. After emulsification, dispense the emPCR amplification mixes into 96-well thermocycler plates, at 100 µL per well; cap the wells.
4. Place them in a PCR machine, and start the amplification (1× 4 min at 94°C; 50× 30 s at 94°C; 4.5 min at 58°C, 30 s at 68°C; 10°C on hold) with the heated lid on.

3.4.5. Emulsion Collection and Initial Washes

1. Aspirate the emulsions from the 96-well thermocycler plates, 8 wells at a time using the Vacuum-Assisted Emulsion Breaking Set up and collect them in the two 50 mL collection tubes, using a slow circular motion of the transpette tips at the bottom of the wells.
2. Rinse the wells *twice* with 100 µL isopropanol per well. Aspirate the rinse and turn the transpette upside-down to retrieve as much material as possible.

3. *SLOWLY* aspirate an additional 5 mL isopropanol to collect remaining beads.
4. Turn off the vacuum, remove and cap the two 50 mL collection tubes containing the amplified DNA beads.

3.4.6. Bead Washes and Recovery (see Notes 16 & 17)

1. Vortex each tube and mix the 50 mL collection tubes in pairs by transferring their contents back and forth, until the bead suspensions are of equivalent amount of emulsion oil.
2. Add isopropanol to a final volume of 40 mL in each collection tube and vortex.
3. Pellet the beads in a centrifuge at 930 × *g* for 5 min and carefully remove the supernatant.
4. Add 35 mL of Enhancing Fluid XT and vortex well to resuspend, centrifuge, and discard the supernatant as in step 3.
5. Add 35 mL of isopropanol and vortex well, centrifuge, and discard the supernatant as in step 3, repeat one more time.
6. Add 35 mL of ethanol and vortex well, centrifuge, and discard the supernatant as in step 3.
7. Add 35 mL of Enhancing Fluid XT and vortex well, centrifuge, and discard the supernatant as in step 3, leaving approximately 2 mL of Enhancing Fluid XT.
8. Transfer the DNA bead suspension using a 1,000 µL pipette in two 1.7 mL tubes for each emulsion cup processed (a total of 4 tubes per full LV kit).
9. Spin-rotate-spin and discard the supernatant.
10. Repeat steps 8 and 9 until the entire DNA bead suspension has been transferred.
11. Rinse each of the 50 mL collection tubes with 600 µL Enhancing Fluid XT, vortex, and add this rinse to the 1.7 mL tubes. Spin-rotate-spin and discard the supernatant.
12. Thoroughly rinse each bead pellet *twice* with 1 mL Enhancing Fluid XT. Spin-rotate-spin and discard the supernatant.
13. Add 1 mL Enhancing Fluid XT to each bead pellet and vortex.

3.4.7. Indirect Enrichment

1. Spin down beads and remove supernatant.
2. Add 1 mL of Melt Solution per tube of beads, vortex, and incubate for 2 min at room temperature then spin and discard supernatant.
3. Repeat step 2 once.
4. Add 1 mL of Annealing Buffer XT per tube of beads and vortex.
5. Spin-rotate-spin and discard the supernatant.

6. Repeat steps 4 and 5 once.
7. Add the following per tube: 45 µL of Annealing Buffer XT, 25 µL of Enrichment Primer and vortex.
8. Incubate at 65°C for 5 min and then promptly cool on ice for 2 min.
9. Wash with 800 µL of Enhancing Fluid XT per tube, vortex, spin-rotate-spin, and discard supernatant.
10. Add 1 mL of Enhancing Fluid XT per tube, vortex, spin-rotate-spin, and discard the supernatant.
11. Repeat step 10.
12. Resuspend each tube in 800 µL of Enhancing Fluid XT and vortex.

3.4.8. Prepare the Enrichment Beads

1. Vortex the provided Enrichment beads for 1 min to resuspend, and prepare 160 µL of beads per emulsion cup (320 µL for two cups) by completing the following steps.
2. Pellet the beads using a MPC and discard the supernatant.
3. Wash the beads 2 times by adding 1 mL of Enhancing Fluid XT, vortex then pellet in the MPC, and discard the supernatant.

3.4.9. Enrich and Collect DNA Library Bead

1. Resuspend the beads by adding 160 or 320 µL (two emulsion cups) of Enhancing Fluid XT to the beads.
2. Add 80 µL washed Enrichment Beads to each tube of amplified DNA and vortex.
3. Rotate the tubes on the Labquake at room temperature (+15 to +25°C) for 5 min.
4. Place the tubes in the MPC, and wait 3–5 min to pellet the Enrichment Beads.
5. Invert the MPC several times.
6. Carefully discard the supernatant from each tube using a 1,000 µL pipette, taking care not to draw off any brown Enrichment Beads.
7. Wash the beads with Enhancing Fluid XT until there are no visible beads remaining in the supernatants (see manual for details).
8. Remove the tubes of enriched beads (enrichment tubes) from the MPC and resuspend each bead pellet in 700 µL of Melt Solution.
9. Vortex for 5 s, and place the enrichment tubes in the MPC until the Enrichment Beads have pelleted.
10. Transfer the supernatants containing enriched DNA beads from each sample (>1.4 mL) into a single 1.7 mL collection tube.

11. Spin-rotate-spin the single 1.7 mL collection tube, and discard the supernatants.
12. Again add 700 µL of Melt Solution to the enrichment tubes.
13. Vortex for 5 s, and place the enrichment tubes in the MPC until the Enrichment Beads have pelleted.
14. Transfer the supernatants containing enriched DNA beads into the same 1.7 mL collection tube, discard the enrichment tubes.
15. Spin-rotate-spin the 1.7 mL collection tube and discard the supernatants.
16. Add 1 mL of Annealing Buffer XT per collection tube and vortex for 5 s.
17. Spin-rotate-spin and discard the supernatant.
18. Repeat steps 16 and 17 *two more times*.
19. Resuspend each bead pellet in 200 µL of Annealing Buffer XT.

3.4.10. Anneal Sequencing Primer

1. Add 50 µL of Sequencing Primer to each collection tube and vortex.
2. Place the collection tubes in a heat block at 65°C for 5 min, and then promptly on ice for 2 min.
3. Add 800 µL of Annealing Buffer XT per collection tube and vortex for 5 s. Spin-rotate-spin, and discard the supernatant.
4. Repeat step 3 two times with 1 mL of Annealing Buffer XT.
5. Add 1 mL of Annealing Buffer XT to each bead pellet and vortex.
6. Store the beads at +2 to +8°C and sequence them within 2 weeks.

3.5. Titanium Shot-Gun Pyrosequencing

The method is based on the Sequencing Method Manual GS FLX Titanium Series October 2009 (Rev. November 2010) and written for a 2 region 70×75 PicoTitre Plate.

3.5.1. Initial Step Up for Sequencing

1. Thaw sequencing reagent cassette.
2. Perform a prewash.
3. Prepare Bead Buffer 2 (BB2) by adding 1.2 mL of Titanium Supplement CB and 34 µL of Apyrase to the bottle of pre-chilled Titanium Bead Buffer.
4. Mix by gently inverting the BB2 and keep on ice throughout the procedure.

3.5.2. Prepare the PicoTiterPlate (PTP) and Bead Deposition Devices (BDD)

1. Submerge the PTP device in BB2 for at least 10 min at room temperature.
2. Wash the bead loading gasket for 30 s in a Sparkleen solution, rinse thoroughly with nanopure water and let air-dry.

3.5.3. Prepare the Beads

Note: The volumes for each plate type and gasket are different. The below protocol is valid only for a 70×75 plate paired with a large 2 region gasket

1. Prepare the Packing Beads by washing three times in 1 mL of BB2 and centrifuging at 10,000 rpm (9,300×*g* RCF) for 5 min. After BB2 is added, vortex to break up aggregates, until a uniform suspension is achieved.
2. After the third wash, add 550 μL BB2 per tube, resuspend beads by vortexing, and keep on ice.
3. Prepare the DNA Library Beads (from emPCR) by vortexing to achieve a uniform suspension.
4. Transfer the appropriate volume (μL) to give 2,000,000 DNA Library beads per region to two 2 mL tubes using a separate tube for each loading region.
5. Add 20 μL of Control DNA Bead suspension to each DNA library bead tube.
6. Spin-rotate-spin the combined Library and Control DNA Beads 1 min at 10,000 rpm.
7. Being careful not to disturb the bead pellet, remove and discard supernatant. Bring the final volume to 50 μL for each tube.
8. In a separate appropriately labeled 15 mL tube prepare the DNA Bead Incubation Mix (DBIM) to a final volume of 2,020 μL by adding 1,570 μL of BB2, 150 μL of Polymerase Cofactor and 300 μL of DNA Polymerase.
9. Vortex gently to ensure thorough mixing.
10. Add 950 μL DBIM to each tube of DNA/Control Beads (50 μL) for a total of 1,000 μL per tube.
11. Vortex well and incubate the samples on the lab rotator at room temperature for minimum of 15 min but do not exceed 50 min.
12. Discard the leftover DBIM.
13. Proceed to preparation of the Enzyme Bead and PPiase Beads during this incubation.

3.5.4. Prepare the Enzyme and PPiase Beads (Bead Layers 1, 3 and 4)

These two types of beads can be washed in parallel, in separate tubes. Make sure to change pipette tips to avoid contaminating them with one another.

1. Vortex then pellet the Enzyme Beads and the PPiase Beads using a MPC for 30 s.
2. Invert the MPC several times and wait another 30 s.
3. Carefully remove and discard the supernatants.
4. Remove the tubes from the MPC.
5. Wash the Enzyme and PPiase Beads three times with 1 mL of BB2 for each tube of Enzyme Beads and 500 µL of BB2 for the tube of PPiase Beads.
6. Vortex and wash the beads using the MPC, as in steps 1–3.
7. After the third wash, add 1 mL of BB2 to each tube of Enzyme Beads and 500 µL of BB2 to the tube of PPiase Beads.
8. Vortex the tubes and keep on ice.
9. In a 15 mL tube prepare the beads for layer 1 to a final volume of 3,800 µL by adding 3,250 µL of BB2 and 550 µL of Enzyme beads.
10. In an appropriately labeled 15 mL tube prepare the beads for layer 3 to a final volume of 3,800 µL by adding 2,500 µL BB2 and 1,300 µL Enzyme Beads.
11. In an appropriately labeled 15 mL tube prepare the beads for layer 4 to a final volume of 3,800 µL by adding 3,340 µL of BB2 and 460 µL of PPiase Beads.

3.5.5. Combine the DNA and Packing Beads (Bead Layer 2)

1. In an appropriately labeled 15 mL tube prepare bead layer 2 to a final volume of 1,700 µL by adding 265 µL Prepared Packing beads and 435 µL BB2 to the DNA beads (1,000 µL from above).
2. Repeat step 1 for additional regions.
3. Vortex and rotate the DNA and Packing Beads mixes at least 5 min at room temperature.
4. Discard any unused packing beads.

3.5.6. Assemble the BDD with the PTP Device and Bead Loading Gasket (see Notes 18–25)

1. Remove the PTP device from the BB2 bath.
2. Wipe the back of the PTP device with a Kimwipe, so that it is completely dry.
3. Place the PTP device onto the Bead Deposition Devices (BDD) base.
4. Insert PTP gasket.
5. Carefully place the BDD top over the assembled BDD base/PTP device/Gasket.
6. Close BDD. When you hear a "click," the latches should be firmly seated in the grooves, providing the correct amount of pressure to maintain a liquid-tight seal.

3.5.7. Deposit Bead Layer 1: The Enzyme Beads Pre-Layer

1. Vortex the layer 1 bead suspension 5 s.
2. Using a pipette and tip of the proper size, promptly load the bead suspension onto the first region of the PTP device, through the port hole on the BDD top. Make sure to use a single, smooth dispensing action to ensure even distribution of the beads over the entire region of the PTP device.
3. Repeat steps 1–3 for all loading regions of the PTP device.
4. Cover the loading ports and the vent holes with port seals.
5. Centrifuge the loaded PTP device in the BDD for 5 min at 1,620 × *g* (2640 RPM for the Beckman Coulter X-12 or X-15 centrifuges).
6. During centrifugation remove the Pre-Wash Cassette and clean the Fluidics Area Deck.

3.5.8. Deposit Bead Layer 2: The DNA and Packing Beads

1. When the centrifugation is complete, return to the BDD and remove the supernatants through the port holes on the BDD.
2. Remove the tube of bead suspension for layer 2 from the rotator.
3. Load the beads into the BDD.
4. Repeat step 3 for all loading regions of the PTP device.
5. Cover the loading ports and the vent holes.
6. Centrifuge the loaded PTP device in the BDD for 10 min at 1,620 × *g* RCF (2,640 rpm for the X-12 or the X-15 centrifuges).

3.5.9. Prepare and Load the Sequencing Reagents Cassette

1. Add 6.6 mL of Titanium Supplement CB and 1,000 μL of DTT to each bottle of Titanium Buffer CB. Recap and gently invert the bottles to mix.
2. Place the Sequencing Reagents tray, kept in the cold, in the right-hand side of the Reagents cassette.
3. Dilute 5 μL of the PPiase reagent in 45 μL Inhibitor TW reagent in a 1.7 mL tube, and vortex.
4. Supplement the four reagents from the Sequencing Reagents tray with their appropriate additives.
5. Invert the tray 20 times to uniformly mix the contents of all tubes, and place the tray back in the Reagents cassette with the lone Apyrase tube at the far end.
6. Remove the caps from all bottles and tubes.
7. Load the Reagents cassette into the instrument.
8. Lower the sipper carefully, and close the exterior fluidics door.

3.5.10. Deposit Bead Layer 3: The Enzyme Beads Post-Layer

1. When the centrifugation is complete, return to the BDD and remove the supernatants through the port holes on the BDD.

2. Vortex the bead suspension for layer 3 for 5 s to obtain a uniform suspension.
3. Using a pipette and tip of the proper size, draw the appropriate amount of bead suspension for the loading regions to be used.
4. For the third and fourth layers, these volumes may be more than needed to fill the loading regions, due to remaining fluid volume leftover on the PTP device. This is not a problem.
5. Pipette up and down three times to resuspend the beads completely (avoid creating bubbles) and promptly load the bead suspension onto the first region of the PTP device, through the port hole on the BDD top.
6. Repeat steps 3–5 for all loading regions of the PTP device.
7. Cover the loading ports and the vent holes.
8. Centrifuge the loaded PTP device in the BDD for 5 min at 1,620 × *g* RCF (2,640 rpm for the Beckman Coulter X-12 or X-15 centrifuges), as above.

3.5.11. Clean the PicoTiterPlate Cartridge and the Camera Faceplate (see Note 26)

1. Click the *Unlock* button from the Instrument Tab.
2. Open the camera door by pulling gently on the handle.
3. Remove the spent PTP device from the PTP cartridge by first pressing the PTP frame spring latch to lift the frame from the cartridge, and then sliding out the used PTP device.
4. Close the PTP frame and remove the used PTP cartridge seal from the PTP cartridge.
5. Wet a Kimwipe with 50% ethanol and wipe the surface of the cartridge to remove any bead and reagent residue. Allow the cartridge to air-dry completely.
6. Use a new Zeiss premoistened cleaning tissue to gently wipe the camera faceplate.
7. Allow the camera faceplate to air-dry completely.
8. Moisten a Kimwipe with a 10% user-prepared solution of Tween-20, and wipe the surface of the PTP cartridge with it.

3.5.12. Load and Set the Run Script and Other Run Parameters

1. On the instrument computer, the GS Sequencer main window is displayed, with the Instrument Tab active.
2. Click the Start button in the Global Action area, which opens the Run Wizard's first window: Choose a procedure.
3. In the Run Wizard's first window, select the Sequencing Run option, and click the Next button, which opens the Run Wizard's second window: Enter IDs and barcodes.
4. Enter the barcode of the PTP device to be used in this Run, click Next, and open the Run Wizard's third window: Enter Run name and Run Group.

5. Enter a specific, unique name for this Run. Then find and select your Run Group in the Run group list. Click Next and open the Run Wizard's fourth window: Choose sequencing kit.
6. Select the type of Sequencing kit to be used in this Run (For the GS FLX Titanium chemistry, the only valid choice is XLR70Ti), and click Next to open the Run Wizard's fifth window: Choose PTP type.
7. Select the bead loading gasket to be used in this Run, and click Next, which opens the Run Wizard's sixth window: Choose number of cycles.
8. Select the number of nucleotide cycles appropriate for this Run, and click Next, which opens the Run Wizard's seventh window: Choose Run Processing type.
9. Select the data processing scheme of "Full processing for Shotgun and Paired End" for this Run, and click Next, which opens the Run Wizard's eighth window: Request data backup.
10. Select the "Backup Run and Processor data upon completion" checkbox, and click Next, which opens the Run Wizard's ninth window: Run comments.
11. Enter any comments about the Run, and click Next, which opens the tenth, and last, window: Insert new PicoPiterPlate device.

3.5.13. Deposit Bead Layer 4: The PPiase Beads

1. When the centrifugation is complete, return to the BDD and remove the supernatants through the port holes on the BDD.
2. Vortex the bead suspension for layer 4 for 5 s to obtain a uniform suspension.
3. Using a pipette and tip of the proper size, draw the appropriate amount of bead suspension for the loading regions you are using.
4. Pipette up and down three times to resuspend the beads completely (avoid creating bubbles) and promptly load the bead suspension onto the first region of the PTP device, through the port hole on the BDD top.
5. Repeat steps 2–4 for all loading regions of the PTP device.
6. Cover the loading ports and the vent holes with new BDD port seals provided in the GS FLX Titanium PTP Kit 70×75.
7. Centrifuge the loaded PTP device in the BDD for 5 min at 1,620×*g* RCF (2,640 rpm for the Beckman Coulter X-12 or X-15 centrifuges), as above.

3.5.14. Insert the PicoTiterPlate Device and Launch the Sequencing Run (see Note 27)

1. Install the cartridge seal with the square ridge on the seal is facing up, and drop the seal in the cartridge groove.
2. Press the PTP cartridge spring latch to lift the PTP frame from the cartridge.
3. When centrifugation is complete return to the BDD and remove the supernatants through the port holes on the BDD.
4. With a pipettor, gently draw out and discard all the supernatant from the centrifuged bead layer 4, through the port holes in the BDD top.
5. Remove the PTP device from the BDD.
6. Slide the PTP device into the PTP cartridge frame, face down. Make sure that the PTP device notch is on the lower right-hand corner, matching the notch in the frame.
7. Close the PTP frame, making sure it is properly secured by the latch.
8. Wipe the back of the PTP device with a Kimwipe. Use a downward motion, toward the camera faceplate, to avoid sliding the PTP device within the frame, or out of it.
9. Close the camera door. Click the Start button in the Run Wizard's Insert new PTP device window to start the sequencing Run (see Note 28).
10. When the sequencing Run is complete, a message will appear in the Status area of the GS Sequencer application window.

3.6. SNP Identification and Annotation

3.6.1. Pyrosequencing Reads Mapping

The GS Reference Mapper application aligns sequencing reads against a reference sequence, with or without associated annotations. Mapping generates consensus sequences of the reads that align against the reference and also computes statistics for variations found in the reads, relative to the reference. The input data can come from one, several or all the PTP regions of the 454 Run(s) of interest. It also allows the inclusion of one or more FASTA files of reads in the analysis, such as reads obtained using Sanger sequencing ("Sanger reads"), as well as the inclusion of one or more Paired End read files. The application can be accessed via a graphical user interface (GUI) or from a command line interface (CLI). The analysis using GUI-based gsMapper is briefly described below and the reader can refer to the 454 Genome Sequencer FLX System Software Manual (version 2.3) for details.

1. Open the gsMapper application.
2. Click the *New Reference Mapping Project* text link to create a new mapping project (the *New* button on the right side of the window performs the same action); type a name for the project in the "Name" text field and select the drive and folder to save the project result; then click the drop-down menu to specify the sequencing type of cDNA or Genomic DNA.

3. Click the *Project Tab* to add or remove read data and/or references. There are two sub-tabs for adding read data: one for adding 454 reads (*GS Reads* sub-tab) and one sub-tab for *FASTA Reads* to add non-454 reads (e.g., Sanger reads) in FASTA format (see Note 29). In addition, the *References* sub-tab is used to add reference files to a project.
4. Click the button + to open the *Select GS Read Data* window and select the data in their directory and pick the sff file(s) to input the read files. To delete GS reads or reference data from a project, click on the sub-tab that lists the *References, GS Reads* or *FASTA Reads*, select the data file(s) to be removed, then click the Remove button -.
5. Customize Project with the Parameters Tab. The Parameters Tab is organized as three sub-tabs of *Input*, *Computation* and *Output*. The input parameters and output data options that can be specified are different for the Genomic and cDNA sequence types. Each sequence type will have its own set of parameters under the project sub-tabs. See 454 Genome Sequencer FLX System Software Manual (version 2.3) for details. Use default value unless specific parameters are needed.
6. Computing the Mapping. When at least one Read Data file and one Reference file have been added to the project and all parameters are within acceptable limits, the *Start* button becomes enabled. Click the *Start* button to save the project parameter settings and carry out the mapping computation of the current project. When the mapping computation is complete, the message "Ready for analysis" appears in the upper right corner of the application window.

3.6.2. SNP Identification

After a successful mapping, the data for the various columns of the reads in the *GS reads* and *FASTA reads* sub-tabs will be updated. Placing the mouse pointer over any cell in the table brings up a tool tip with additional information about that item. A successful mapping also updates data in the *Overview* tab.

1. Viewing Mapping Output with the *Result Files* Tab. Click the *Result Files* tab to view the various files generated by the GS Reference Mapper software. Using the list in the left-hand panel of the tab, click on the name of the output file to examine; its content will be displayed on the right-hand panel of the tab.
2. Click the *454 HCDiffs* sub-tab to obtain the Information about local differences between reads and the reference and the details can be obtained from 454AllDiffs.txt file in the mapping output file directory. This file contains the list of high confidence variations (of at least three nonduplicate

reads) relative to the reference sequence or to other reads aligned at a specific location. Single nucleotide polymorphisms (SNPs), insertion–deletion pairs, multihomopolymer, insertion or deletion regions, and single-base overcalls and undercalls are reported. The GS Reference Mapper application uses a combination of flow signal information, quality score information and difference type information to determine if a difference is high confidence. The leftmost columns (from Reference Accession Number to Total Depth) contain general information on the high confidence differences (HCDiffs) that were found in the computed data, between the reads and the reference sequences. The center columns (from Reference Amino Acids to Known SNP Info) show gene annotation or known SNP file information on the regions of the HCDiffs. For such information to be available, Genome Annotation and Known SNP databases must have been specified (either on the Parameters Tab or by using a Golden Path reference sequence). The remaining columns on the right (from Percent Forward to Total Num Reverse Reads) contain a detailed breakdown of the computation data shown on the left.

3. The *Structural Variations* sub-tab shows locations of larger-scale changes relative to the reference that are indicated by a group of reads. Entries in the table are classified as either Rearrangement points or Rearrangement regions containing a section listing the high confidence rearrangement points, followed by a section listing the high confidence rearrangement regions.
4. The two text files of *454 HCDiffs* and *Structural Variations* as mentioned above can be imported into excel, and the possibility of potential SNPs will be assessed based on the reads depth and the frequency as described previously (2). The HCDiffs after filtration will be used as the SNP candidates.
5. SNP Annotation: Genome annotation file can be added in the gsMapper application during the Mapping, which will generate the details about the SNPs.
6. SNP visualization: the SNP information can be converted into gff file and the SNP information across the genome can be viewed by any genome viewing software such as SignalMap from NimbleGen (Roche, CA). The SNPs identified in this study for *C. thermocellum* ethanol-tolerant EA strain are shown in Fig. 2.

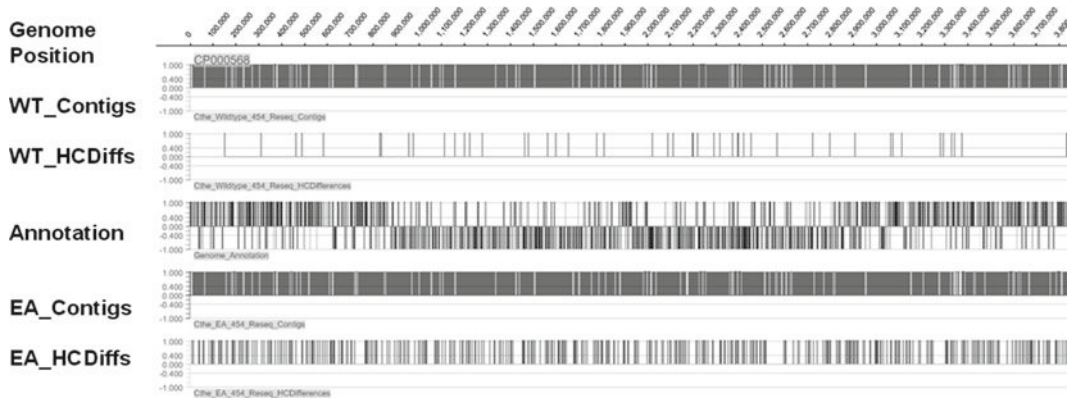


Fig. 2. Summary of pyrosequencing data for *Clostridium thermocellum* wild-type strain ATCC27405 (WT) and the ethanol-tolerant mutant strain (EA). The columns for each lane are: Genome positions, 149 contiguous regions of DNA (WT_Contigs) and 72 putative high confidence differences (WT_HCDiffs) that mapped to the wild-type reference genome from 454 pyrosequencing data for the wild-type strain, genome annotation, and 153 contiguous regions of DNA (EA_Contigs) and 500 putative high confidence differences (EA_HCDiffs) that mapped to the wild-type reference genome from 454 pyrosequencing data for the ethanol-tolerant EA strain.

4. Notes

1. Unless stated otherwise, all solutions should be prepared in water that has a resistivity of 18.2 MΩ cm and the water used in 454 pyrosequencing is molecular biology grade.
2. During genomic DNA extraction, the DNA pellet may be lost during isopropanol precipitation, use extreme care when removing the isopropanol to avoid losing the pellet.
3. Remember to mix the samples periodically during the rehydration step, and do not leave the DNA at 65°C overnight.
4. *Sodium hydroxide* present in the 454 sequencing Melt Solution and *Sodium chlorite* tablet used in 454 sequencing system are highly corrosive chemical that may cause burns if it contacts eyes or skin. Read the Material Safety Data Sheet and wear appropriate personal protective equipment.
5. Be as exact as possible in pipetting. Accuracy is critical for 454 library preparation.
6. Sample DNA used for 454 titanium library preparation should be: double-stranded; $OD_{260/280} \geq 1.8$; concentration ≥5 ng/μL, and fragment size >1.5 kb.
7. Library preparation involves many pipetting steps and manipulations. Although the input DNA requirement is 5 μg per

library it is best to have 10–15 μg in case there is a failure at the library preparation step to avoid having to return to DNA isolation.

8. After nebulization, total recovery should be greater than 300 μL. Do NOT collect any material that may have lodged outside the Nebulizer: this material may not have been completely fragmented and could cause problems later on.
9. Due to significant variability in the size exclusion characteristics between individual lots of AMPure beads, a calibration of each lot is necessary before these beads can be used for the preparation of DNA libraries in the GS FLX Titanium Sequencing system. Make sure always use calibrated AMPure beads for the preparation of your DNA libraries.
10. The exact incubation volume and number of beads are critically important to the success of size exclusion steps using AMPure beads. Make sure to pipette the exact volumes of all components of such reactions, as listed in the procedure.
11. The recovered nebulized and size-selected material should have a mean size (top of the peak) between 500 and 800 bp, with less than 10% of the library peak below 350 bp and less than 10% above 1,000 bp.
12. During single-stranded DNA library isolation, the fresh-made Melt Solution (mixing 125 μL 10N NaOH in 9.875 mL Water) has a shelf life of only 7 days (at +2 to +8°C). Buffer PBI contains a pH indicator. The solution should promptly return to its neutral/acidic yellow color after the addition of each Melt Solution wash and mixing. If it does not, add an extra 5 μL of the 3M Sodium acetate pH 5.2 solution to fully neutralize the melts.
13. During the emulsion oil preparation step, if the TissueLyser makes a loud rattling sound, turn it off immediately and make sure that all tubes and adapters are screwed tightly in place.
14. At the step of emulsion dispensing, some emulsion may have lodged in the tube lids. Pipette it up from the lids but *DO NOT* spin the emulsion tubes in an attempt to reclaim this material as this would risk breaking the emulsions; and dislodge air bubbles that may be present at the bottom of the wells by gently capping the plate.
15. After the amplification reaction, *DO NOT* freeze the DNA beads, instead leave the amplification reactions at 10°C (up to 16 h). In addition, check all wells for emulsion breakage (i.e., faint yet distinct layers with clearer middle layer). If the emulsion in any well appears broken, discard the entire well and do not recover the beads from it.
16. During Bead recovery avoid mixing of individual libraries unless they have been barcoded.

17. All 1.7 mL tubes used for bead washes and recovery must be siliconized. Use the correct pipette tips to load beads: 2 mL tip for large regions.
18. Prior to loading any bead layer pipette the beads up and down 2–3 times to get an even suspension of the beads (avoid spilling or creating bubbles) and promptly load them onto the region of the PTP device, through the port hole on the BDD top.
19. Minimize the time interval between loading the beads and starting the centrifugation.
20. Load the beads promptly for an even distribution in the PTP device.
21. Avoid injecting air into the BDD.
22. Use a single, even injection to fill each loading region of the BDD.
23. Fill the BDD completely but do not overflow the loading regions. Discard any excess bead mix.
24. Change gloves after loading the Reagents cassette into the instrument before the step of Deposit Bead Layer 3. It is useful to save this bead layer 2 supernatants, for troubleshooting purposes in case the sequencing Run does not produce the expected number of reads.
25. Don't over load the regions of a PTP. This may cause shifting in the gasket. Especially in the two region gasket.
26. When cleaning the PTP Cartridge and the Camera Faceplate, always be extremely careful when handling or working near the camera face. Never touch the camera face with anything other than Zeiss moistened cleaning tissue or Lens paper from Thorlabs.
27. When inserting the PTP Device for sequencing run, make sure that the square ridge on the seal is facing up, and drop the seal in the cartridge groove. If necessary, gently tap the seal into place with a gloved hand. DO NOT wipe the seal with anything, not even with a Kimwipe, as this could damage or stretch the seal.
28. After the run is initialed, monitor the instrument until the sequencing run is under way and the Status LED located above the camera door on the instrument is blinking green (also shown at the upper-left corner of the GS Sequencer window). If the instrument encounters any problems during the initiation of the Run, a message describing the issue will appear in the Status area of the GS Sequencer window, and user intervention will be required.
29. During gsMapping (*see* Subheading 3.6.1), although in many cases the file name extension for FASTA files may be *.fna*, there

is no standardized extension for such files. If you cannot see your FASTA file, try setting the "Files of Type" field to "All Files."

Acknowledgments

The authors would like to thank Meghan M. Drake for her careful review and suggestions. The BioEnergy Science Center is a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy.

References

1. Jeffries, T. W. (2005) Ethanol fermentation on the move. *Nat. Biotechnol.* **23**, 40–41.
2. Yang, S., Land, M. L., Klingeman, D. M., Pelletier, D. A., Lu, T-YS., Martin, S. L., Guo, H-B., Smith, J. C., and Brown, S. D (2010) Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **107**, 10395–10400.
3. MacLean, D., Jones, J. D., and Studholme, D. J. (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.* 7, 287–296.
4. Metzker, M. L (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31–46.
5. Duan, J., Heikkila, J. J., and Glick, B. R (2010) Sequencing a bacterial genome: an overview. In: Mendez-Vilas, A. (ed) *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, Volume* 2. Formatex Research Center, Badajoz, Spain.
6. Widiastuti, H., Kim, J. Y., Selvarasu, S., Karimi, I. A., Kim, H., Seo, J. S., and Lee, D. Y. (2011) Genome-scale modeling and in silico analysis of ethanologenic bacteria *Zymomonas mobilis*. *Biotechnol. Bioeng.* **108**, 655–665.
7. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L.,A., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.
8. Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., et al. (2008) Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **105**, 9145–9150.
9. Smith, D. R., Quinlan, A. R., Peckham, H. E., Makowsky, K., Tao, W., Woolf, B., et al. (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* **18**, 1638–1642.
10. Dohm, J. C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105.
11. Hillier, L. W., Marth, G. T., Quinlan, A. R., Dooling, D., Fewell, G., Barnett, D., et al. (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods* **5**, 183–188.
12. Harismendy, O., Ng, P. C., Strausberg, R. L., Wang, X., Stockwell, T. B., Beeson, K. Y., et al. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, R32.
13. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506–577.
14. Herrero, A. A., and Gomez, R. F. (1980) Development of ethanol tolerance in *Clostridium thermocellum*- effect of growth temperature. *Appl. Environ. Microbiol.* **40**, 571–577.
15. Williams, T. I., Combs, J. C., Lynn, B. C., and Strobel, H. J. (2007) Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. *Appl. Microbiol. Biotechnol.* **74**, 422–432.

Chapter 10

Use of Proteomic Tools in Microbial Engineering for Biofuel Production

Shaoming Mao*, Kaizhi Jia*, Yanping Zhang, and Yin Li

Abstract

The production of biofuels from renewable sources by microbial engineering has gained increased attention due to energy and environmental concerns. Butanol is one of the important gasoline-substitute fuels and can be produced by native microorganism *Clostridium acetobutylicum*. To develop a fundamental tool to understand *C. acetobutylicum*, a high resolution proteome reference map for this species has been established. To better understand the relationship between butanol tolerance and butanol yield, we performed a comparative proteomic analysis between the wild-type strain DSM 1731 and its mutant Rh8 at acidogenic and solventogenic phases, respectively. The 102 differentially expressed proteins that are mainly involved in protein folding, solvent formation, amino acid metabolism, protein synthesis, nucleotide metabolism, transport, and others were detected. Hierarchical clustering analysis revealed that over 70% of the 102 differentially expressed proteins in mutant Rh8 were either upregulated (e.g., chaperones and solvent formation related) or downregulated (e.g., amino acid metabolism and protein synthesis related) in both acidogenic and solventogenic phase, which, respectively, are only upregulated or downregulated in solventogenic phase in the wild-type strain.

Key words: Comparative proteomics, Strain engineering, Bio-based chemical, Biofuel, *Clostridium acetobutylicum*

1. Introduction

Due to the world's mushrooming energy demands and the challenge of reducing greenhouse gas emissions from fossil fuels, biofuel produced from renewable resources, as a partial solution, received much more attention (1–3). Butanol is an important chemical and can be used as a directed replacement of gasoline or a fuel additive for its promising characteristics such as low volatility, less hydroscopic and corrosive, and so on (4, 5). While, the key

* These authors contributed equally to this paper.

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_10, © Springer Science+Business Media, LLC 2012

problems limiting the large-scale bioproduction of butanol is the cost of substrate and butanol toxicity/inhibition of the fermenting microorganisms, resulting in a low butanol titer in the fermentation broth (6–8). These problems can be solved by the implementation of innovative reactor or process design strategies, but it is more cost effective to directly engineer microbes (6, 9, 10).

As the physiological mechanisms of substrate utilization and butanol toxicity of butanol-producing microbes is poorly characterized, random approaches such as traditional mutagenesis and screening and genome shuffling has been adopted for rapid engineering of these microbes (9). To make their desirable industrial phenotypes transferable and tractable for metabolic engineering, a system-level understanding of metabolism is required (11). Proteomics is such an effective and efficient tool to monitor the changes of functional proteins simultaneously in mutant and wild-type strains (11–14). By systematically analyzing these data, the potential molecular mechanisms for desirable industrial phenotypes could be identified (12). Moreover, the information would provide rationale to cost-effectively engineer optimal industrial microbes such as eliminating bottlenecks in microbial metabolic pathways and increasing the stress resistance (12, 15).

The aim of this protocol is to obtain fundamental data for understanding the biological mechanism on butanol tolerance and yield by establishing a comprehensive cytoplasmic proteome reference map for *Clostridium acetobutylicum*. Using the established protein profiling protocol, a comparative proteomic analysis between *C. acetobutylicum* DSM 1731and its mutant Rh8 (a genome-shuffled strain derived from strain DSM 1731 and its butanol-tolerant mutants) was carried out (12). With the assistance of the proteome reference map, the differentially expressed proteins between the wild type and different mutant strains, or between different culture conditions, could be revealed. These information will not only help us to better understand the relevant physiological mechanisms, but also provide useful information of target proteins for further metabolic engineering, with the aim to improve the biofuel production efficiency.

2. Materials

2.1. Bacterial Strains

1. Wild-type *C. acetobutylicum* DSM 1731 (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany).
2. Mutant Rh8, a genome-shuffled strain derived from strain DSM 1731 and its butanol-tolerant mutants (12).

2.2. Culture and Fermentation

1. Anaerobic chamber (Bactron™, Sheldon manufacturing, USA).
2. Clostridial growth medium (CGM) (per liter of distilled water): 0.75 g KH_2PO_4, 0.75 g K_2HPO_4, 0.4 g $MgSO_4 \cdot H_2O$, 0.01 g $MnSO_4 \cdot H_2O$, 0.01 g $FeSO_4 \cdot 7H_2O$, 1.0 g NaCl, 2.0 g asparagines, 5.0 g yeast extract, 2.0 g $(NH_4)_2SO_4$, and 50 g glucose.
3. Reinforced Clostridium Medium (RCM) plates (per liter of distilled water): 10 g tryptone, 10 g beef extract, 3 g yeast extract, 5 g glucose, 10 g starch soluble (A.R.), 5 g NaCl, 3 g $CH_3COONa \cdot 3H_2O$, and 0.5 g L-cysteine hydrochloride monohydrate.
4. BioFlo 110 fermentors (New Brunswick Scientific, Edison NJ) containing 4.0 L CGM.
5. Unico UV-2000 Spectrophotometer (Shanghai, China).

2.3. Cell Lysis, Sample Precipitation, Resuspension, and Determination

1. TE buffer: 10 mM Tris–HCl, 5 mM EDTA, pH 7.5.
2. 5 mM PMSF (phenylmethanesulfonyl fluoride).
3. Xinchen JY92-IID Ultrasonic cracker (Beijing Hongdaxinchen Biotechnology CO., LTD, Beijing, China).
4. RNaseA and DNaseI.
5. Acetone supplemented with 0.1% DTT (w/v), placed at –20°C at least 15 min prior to use.
6. Sample lysis: 8 M urea, 4% CHAPS, and 65 mM DTT.
7. 2D Quant Kit (GE Healthcare, Uppsala, Sweden).

2.4. Two-Dimensional Polyacrylamide Gel Electrophoresis

1. Ettan IPGphor 3 system (GE Healthcare, Uppsala, Sweden) using 24 cm IPG strips (GE Healthcare).
2. PlusOne Drystrip Cover Fluid (GE Healthcare).
3. Equilibration buffer 1: 6 M urea, 50 mM DTT, 30% glycerol, 50 mM Tris–HCl, pH 8.8.
4. Equilibration buffer 2: 6 M urea, 100 mM iodoacetamide, 30% glycerol, 50 mM Tris–HCl, pH 8.8.
5. Polyacrylamide storage solution: 30% (w/v) acrylamide, 0.8% (w/v) methylene bisacrylamide.
6. Tris–HCl: 1.5 M Tris, pH 8.8.
7. Electrophoresis buffer: 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3.
8. Denaturing solution: 0.7% agarose, 0.1% SDS, 192 mM glycine, 25 mM Tris–HCl (pH 8.8), and 0.001% bromophenol blue (BPB).
9. CBB (Coomassie brilliant blue) solution: 0.03% coomassie brilliant blue R-250 (Amresco, Solon, OH), 23.75% (v/v) ethanol, and 8% acetic acid.

10. Image Scanner III (GE healthcare).
11. Image Master 6.0 2-D platinum software (GE healthcare).
12. KMC support (TIGR MeV, version 4.5.1).

2.5. In-Gel Digestion and Detection

1. Spot picker (The Gel Company, San Francisco, CA).
2. Proteomics sequencing grade trypsin (10 ng/μL; Cat.: T6567, Sigma-Aldrich, St. Louis, MO).
3. Speed Vac (Thermo Savant, Waltham, MA).
4. ACN (Acetonitrile) solution: 50 mM ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO), 50% ACN.

2.6. Maldi-TOF MS and MS/MS Analysis

1. 0.1% TFA (trifluoroacetic acid).
2. CHCA (α-Cyano-4-hydroxycinnamic acid) matrix solution: 5 mg/mL CHCA in 50% ACN/0.1% TFA.
3. Applied Biosystems 4700 Proteomics Analyzer MALDI-TOF/TOF (Applied Biosystems, Framingham, MA).
4. MASCOT 2.0 search engine (Matrix Science, London, U.K.).
5. GPS Explorer 3.5 (Applied Biosystems).
6. MASCOT 2.0 search engine (Matrix Science, London, U.K.).

2.7. Comparative Proteomics Analysis

TIGR Multiexperiment Viewer (MeV, version 4.5.1, Dana-Farber Cancer Institute, http://www.tm4.org/mev/).

3. Methods

Proteomics offers a powerful tool to monitor global cellular events by directly visualizing vast amount of gene expression products and understand the cellular status at the protein level, which cannot be deciphered from either genome or transcriptomic analysis (16, 17). First, this technique extracts complex protein mixtures from microbial cells; and then, proteins were separated according to two independent properties in two discrete steps (18). In the first step, isoelectric focusing (IEF) was used to separate proteins according to their isoelectric points (pI); subsequently, protein mixture was separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Each spot on the resulting two-dimensional gel potentially corresponds to a single protein species in the sample. Thus thousands of different proteins can be separated, and their information about pI and molecular weight can be obtained. Furthermore, these proteins can be identified by tryptic digestion and MALDI-TOF MS and then were used to reconstruct the metabolic network of *C. acetobutylicum* to evaluate the quality and the coverage of the proteome reference map (12). The wide

coverage of the identified proteins in the metabolic pathways indicates a good application potential of using this proteome reference map to understand the clostridial metabolism-related physiology. By analyzing the differentially expressed protein spots in 2D gel among different growth stages, or between the wild type and the various mutants, the detailed physiological mechanism of substrate utilization and product tolerance in engineered biofuel-producing microorganisms could be clarified (12).

3.1. Sample Preparation

3.1.1. Cell Culture

1. Colonies of strains DSM 1731 and RH8 were picked from plates at least 4 days old and inoculated into 10 mL CGM.
2. After heat-shocked at 80°C for 10 min, the precultures were cultured at 37°C for 12 h.
3. Subsequently, they were transferred into a 250 mL bottle containing 100 mL CGM at 37°C. When the cell density reached an OD_{600} of 2.0, cells were harvested for proteomics analysis.

3.1.2. Fermentation Experiments

1. 200 mL seed cultures were inoculated into the fermentor. The initial pH of the fermentation was 6.5, and the pH was allowed to drop to 5.0 as the culture progressed.
2. Subsequently, the pH was automatically maintained at or above 5.0 by adding 6 M ammonium hydroxide. Cells were harvested for proteomic analysis, when cell density achieved OD_{600} of 2.0 (as acidogenic cells) and 5.0 (as solventogenic cells).

3.1.3. Extraction of Total Protein Using Acetone

1. Cells were harvested by centrifugation at 10,000 × *g* at 4°C for 10 min and washed three times with 45 mL of ice-cold TE buffer.
2. The resulting *C. acetobutylicum* DSM 1731 and Rh8 cell pellets were resuspended in 20 mL of lysis buffer. The cells were sonicated on ice for 30 min using the following conditions: 5 s of sonication with a 5-s interval, set at 50% duty cycle. After adding 10 mg/mL RNaseA and 1 U/μL DNaseI, the cell lysate was incubated at 37°C for 30 min to degrade nucleic acids.
3. The resulting lysate was collected and centrifuged at 10,000 × *g* at 4°C for 15 min. The supernatant was diluted with three volume of ice-cold acetone. The protein was precipitated for 16 h at −20°C, and then centrifugated at 15,000 rpm at 4°C for 30 min. The supernatant was decanted, and the pellet was washed twice with ice-cold acetone, followed by removal of all the acetone.
4. The protein precipitants were solubilized in sample lysis buffer, followed by vortexing and centrifugation of the sample at 10,000 rpm for 10 min (4°C). Protein concentration was measured by using the 2-D Quant Kit, and 1 mg aliquots were stored at −80°C.

3.2. Two-Dimensional Polyacrylamide Gel Electrophoresis

2-DGE was carried out using precast IPG strips on an IPGphor unit followed by 12.5% SDS-PAGE on a vertical electrophoresis unit.

3.2.1. First-Dimension Isoelectric Focusing Using pH 4–7 IPG Strips

1. For IPG strips of pH 4–7, approximately 1 mg of protein per sample was mixed with sample lysis buffer containing 0.5% (v/v) pH 4–7 IPG buffer to bring to a final volume of 450 μL. A trace of BPB was added and the mixture centrifuged at 10,000 rpm for 10 min, followed by pipetting into 24 cm strip holder tray placed into the IPGphor unit.
2. IPG strip were carefully placed onto the protein samples and then overlayed with about 2 mL Cover Fluid. The protein sample was run using the in-gel sample rehydration technique. IEF was performed using the following voltage program: 30 V constant for 12 h, gradient to 200 V for 4 h, gradient to 1,000 V for 4 h for a total of 65,000 Vh.

3.2.2. First-Dimension Isoelectric Focusing Using pH 6–11 IPG Strips

1. For IPG strips of pH 6–11, approximately 1 mg protein per sample (100 μL) was loaded on a previously rehydrated strip (rehydrated for 12 h) by anodic cup loading.
2. IEF was performed using the following voltage program: 150 V constant for 4 h, gradient to 300 V within 2 h, gradient to 600 V within 2 h, gradient to 8,000 V within 30 min, and then 8,000 V until a total of 32,000 Vh had been achieved. The temperature was maintained at 20°C for IEF.

3.2.3. SDS-PAGE

1. Following IEF, the IPG strips were immediately used for the second dimension. Each strip was equilibrated in 10 mL equilibration buffer 1 for 15 min and then equilibrated in 10 mL equilibration buffer 2 for another 15 min.
2. Equilibrated IPG strips were subsequently placed on the top of 12.5% SDS-PAGE gels. A denaturing solution was loaded onto the gel strips.
3. After agarose solidification, electrophoresis was performed in the electrophoresis buffer at 16°C for 1 h at 1 W/gel, followed by 5–6 h at 10 W/gel until the BPB reached the bottom. Six gels were run in parallel on Ettan DALT six electrophoresis system. For each sample, a minimum of three IPG strips and corresponding SDS-PAGE was used under the same conditions.
4. To visualize the protein spots, the 2D gels were stained with CBB solution. An example result was seen in Fig. 1a, b.

3.3. 2-DE Gel Image analysis

1. The gels were scanned at 300 dpi resolution using Image Scanner III. Comparative analysis of the protein spots were performed using Image Master 6.0 2-D platinum software.

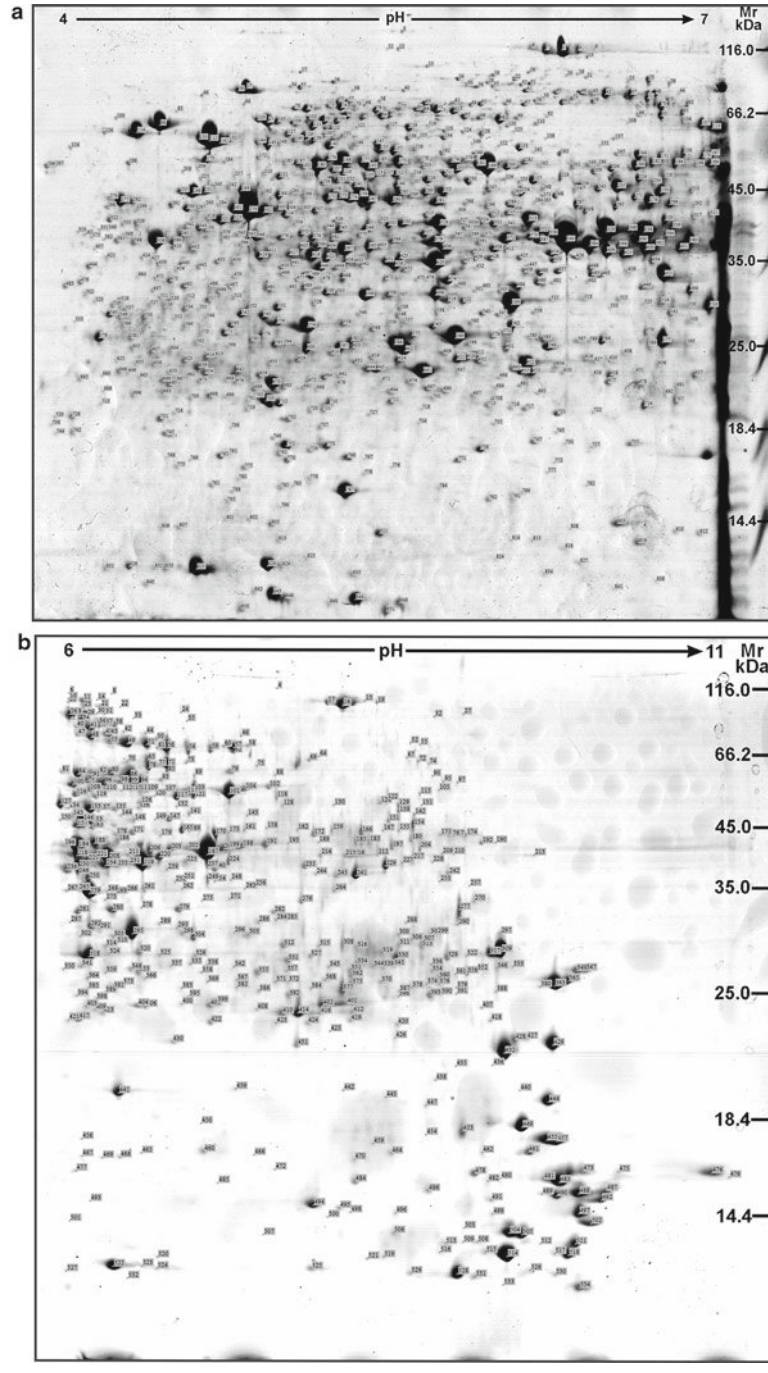


Fig. 1. (**a**) Reference map of the cytoplasmic proteins from *C. acetobutylicum* DSM 1731 within p*I* ranges of 4–7. Approximately 800 μg of protein were subjected to IEF with an IPG 4–7 strip (24 cm separation length) and resolved in the second dimension using 12.5% SDS-PAGE gels. The separated proteins were stained with coomassie blue. (**b**) Reference proteome map of the cytoplasmic proteins from *C. acetobutylicum* DSM 1731 within pI ranges of 6–11. Approximately 800 μg of protein were subjected to IEF with an IPG 6–11 strip (18 cm separation length) and resolved in the second dimension using 12.5% SDS-PAGE gels. The separated proteins were stained with coomassie blue (12) (reproduced from Fig. 1. in ref. (9) with permission from American Chemical Society).

2. All images were submitted to automatic spot detection. The spots were checked manually to eliminate any possible artifacts including background noise and streaks (see Note 1).
3. All images were aligned and matched by using the common spots present in all images as landmarks, to detect potential differentially expressed proteins (see Note 2). Only protein spots showing reproducible changes in protein abundance, by multiple experiments (at least three biological repetitions and two technical replicates), were considered as biomarkers associated with wild-type strain and mutant. Statistical analysis was performed using Student *t*-test. $P < 0.05$ was considered significant.

3.4. Protein Expression Profile Analysis

Hierarchical clustering analysis was used to group proteins exhibiting similar expression profiles. The relative volume (% Vol) of each protein spot, obtained from the Image Master 6.0 2-D platinum software, was used for hierarchical cluster analysis. The protein profile normalization was managed. The differentially expressed proteins were grouped on the basis of similarity in expression profile by using KMC support. An example result was seen in Fig. 2.

3.5. In-Gel-Digestion

1. The coomassie blue-stained protein spots were manually excised using a spot picker. The spots were transferred to Eppendorf tubes, sealed, and stored at −80°C until further processing.
2. One-hundred microliters of 50% ACN solution, and the mixtures were incubated with occasional vortexing for 30 min (see Note 3).
3. The spots were then dehydrated with 100 μL of ACN at room temperature for 15 min.
4. The gel spots were dried under vacuum and rehydrated in 20 μL trypsin, in which lysine residues have been reductively methylated, leading to resistance to autolysis (see Note 4).
5. After rehydration at 4°C for 45 min, excess trypsin solution was removed, and 10 μL of 50 mM ammonium bicarbonate was added followed by incubation at 37°C for 16 h.
6. Peptides were extracted twice by the addition of 30 μL of 5% formic acid (v/v)/50% ACN (v/v) solution and vortexing for 30 min.
7. The peptide samples were concentrated by using Speed Vac to approximately 10 μL and stored at −20°C for mass spectrometric analysis.

3.6. Maldi-TOF MS and MS/MS Analysis

1. A 0.4 μL aliquot of the concentrated tryptic peptide mixture in 0.1% TFA was mixed with 0.4 μL of CHCA matrix solution and spotted onto a freshly cleaned target plate. After air drying,

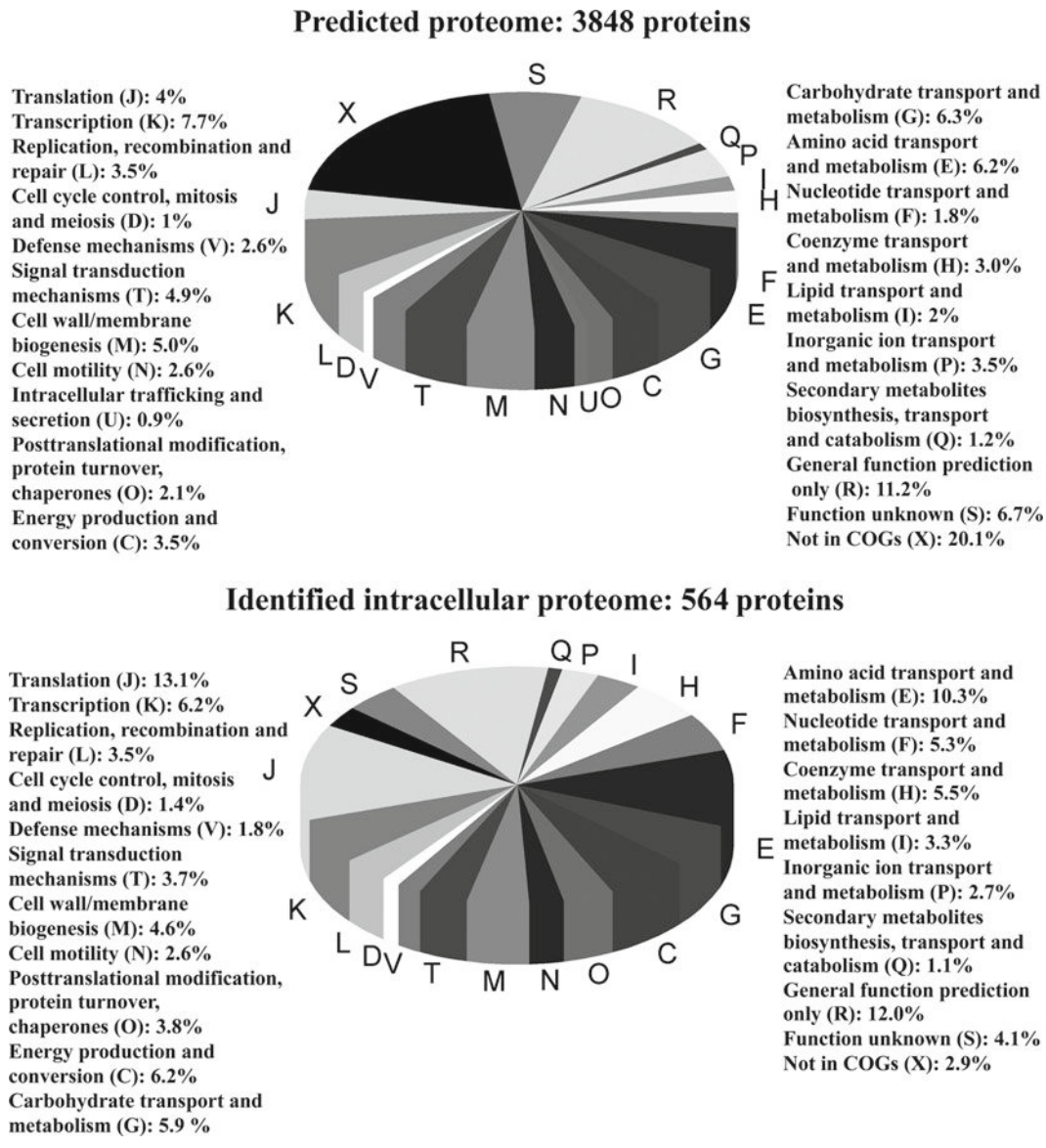


Fig. 2. Representation of the distribution into COG categories of the predicted and identified *C. acetobutylicum* DSM 1731 proteins. Grouping of proteins into COGs was carried out according to the classification scheme provided by the GenBank database (12) (reproduced from Fig 2. in (9) with permission from American Chemical Society).

the crystallized spots were analyzed on the Applied Biosystems 4700 Proteomics Analyzer MALDI-TOF/TOF (see Note 5).

2. The MS/MS mass spectra were acquired by the data-dependent acquisition method with the ten strongest precursors selected from one MS scan. All MS and MS/MS spectra were obtained by accumulation of at least 1,000 and 3,000 laser shots, respectively. Neither baseline subtraction nor smoothing was applied to recorded spectra.

3. MS and MS/ MS data were analyzed and peak lists were generated using GPS Explorer 3.5. MS peaks were selected between 850 and 3,700 Da and filtered with a signal-to-noise ratio greater than 20. A peak intensity filter was used with no more than 50 peaks per 200 Da.
4. MS/MS peaks were selected based on a signal-to-noise ratio greater than 10 over a mass range of 60–20 Da below the precursor mass. MS and MS/MS data were analyzed using MASCOT 2.0 search engine to search against the *C. acetobutylicum* protein sequence database (10,159 sequences; 3,116,366 residues) downloaded from NCBI database on March 20 2008 (see Note 6). For all proteins successfully identified by Peptide Mass Fingerprint and/or MS/MS, Mascot score greater than 53 (the default MASCOT threshold for such searches) was accepted as significant (p value <0.05). An example result was seen in Fig. 3.

3.7. Comparative Proteomic Analysis

1. At the acidogenic and solventogenic phases, the total proteins of butanol-tolerant mutant Rh8 and its original strain 1731 was exacted and run on 2D DIGE (pH 4–7 and 6–11).
2. Based on the established proteome reference map, comparative proteomic analysis between DSM 1731 and its mutant Rh8 was performed. An example result was seen in Fig. 4.
3. After identifying significantly differentially expressed proteins in DSM 1731 and its mutant Rh8, their different expression profiles were analyzed and categorized, the hierarchical clustering analysis and heat map visualization were performed with TIGR MeV (see Note 7). An example result was seen in Fig. 5.

   By comparative proteomic analysis between the wild-type strain DSM 1731 and its mutant Rh8 in acidogenesis and solventogenesis, 102 differentially expressed proteins were identified, which are mainly involved in protein folding, solvent formation, amino acid metabolism, protein synthesis, nucleotide metabolism, transport, and others. Chaperone proteins such as Hsp18, GrpE, and ClpC, and solvent formation related proteins such as THL, AdhE1, CtfA/B, Adc, and BdhA/B, were only upregulated in solventogenic phase in the wild-type strain DSM 1731. However, in mutant Rh8, these proteins were upregulated in acidogenic phase and further upregulated in solventogenic phase. In addition, amino acid metabolism and protein synthesis related proteins were only downregulated in solventogenic phase in the wild-type strain DSM 1731. However, in mutant Rh8, these proteins were downregulated in acidogenic phase, and further downregulated in solventogenic phase. This suggests that mutant Rh8 may have developed a mechanism to prepare themselves for butanol challenges before butanol is produced, which finally resulted in an

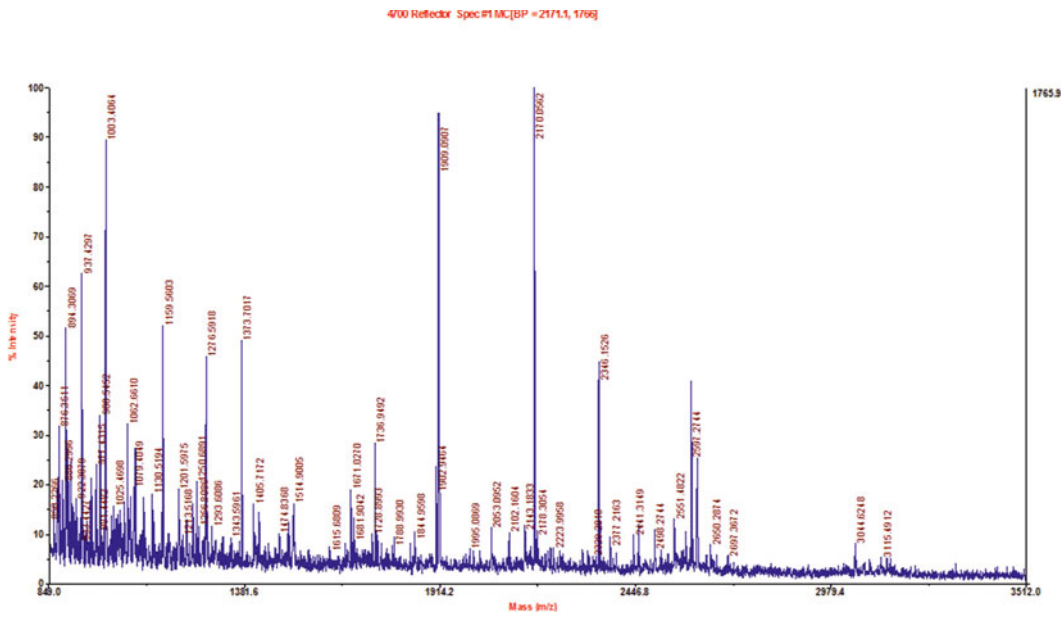


Fig. 3. The peptide mass fingerprinting (PMF) spectrum of glucose-6-phosphate isomerase (12) (reproduced from Supporting Information 3 Spot 1 in (9) with permission from American Chemical Society).

1. gi|15025711|gb|AAK80627.1|AE007 **Mass:** 49759 **Score:** 114 **Expect:** 4e-008 **Peptides matched:** 19

Glucose-6-phosphate isomerase [Clostridium acetobutylicum ATCC 824]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Ions	Peptide
870.3369	869.3296	869.3953	-0.0656	358 -	364	0	---	TMDFVNK + Oxidation (M)
1019.4286	1018.4213	1018.6062	-0.1849	158 -	165	1	---	IFKELLEK
1250.6877	1249.6804	1249.6302	0.0502	146 -	157	0	---	SGTTTEPAIAFR
1671.0266	1670.0193	1669.9515	0.0679	323 -	336	1	---	RQLIETFINVVNPK
1909.0905	1908.0832	1908.0104	0.0728	72 -	90	0	---	NSVDAFIVIGIGGSYLGAR
1909.0905	1908.0832	1908.0104	0.0728	72 -	90	0	5	NSVDAFIVIGIGGSYLGAR
2143.1851	2142.1778	2142.0051	0.1727	91 -	109	1	---	AAIEMFSHSFSSSISKEER
2170.0554	2169.0481	2168.9861	0.0620	189 -	208	0	36	TLSDNEGYETFVVPDDVGGR
2170.0554	2169.0481	2168.9861	0.0620	189 -	208	0	---	TLSDNEGYETFVVPDDVGGR
2178.3057	2177.2984	2177.1956	0.1029	70 -	90	1	---	IRNSVDAFIVIGIGGSYLGAR
2346.1523	2345.1450	2345.0811	0.0640	239 -	258	1	29	EVYSEPDLEKNEAYQYAAAR
2346.1523	2345.1450	2345.0811	0.0640	239 -	258	1	---	EVYSEPDLEKNEAYQYAAAR
2498.2690	2497.2617	2497.1776	0.0842	267 -	286	0	---	SIEMVVNFEPSLHYFGEWWK
2551.4329	2550.4256	2550.3324	0.0932	209 -	232	0	---	YSVLTAVGLLPIAVAGIDIDEMMK + 2 Oxidation (M)
2597.2742	2596.2669	2596.2016	0.0654	299 -	322	0	11	GLFPAAGDFSTDLHSMGQYIQEGR
2597.2742	2596.2669	2596.2016	0.0654	299 -	322	0	---	GLFPAAGDFSTDLHSMGQYIQEGR
2613.2930	2612.2857	2612.1965	0.0893	299 -	322	0	---	GLFPAAGDFSTDLHSMGQYIQEGR + Oxidation (M)
2613.2930	2612.2857	2612.1965	0.0893	299 -	322	0	---	GLFPAAGDFSTDLHSMGQYIQEGR + Oxidation (M)
3043.5828	3042.5755	3042.5007	0.0748	110 -	136	1	---	KNPEIFFCGNNISSTYLADLLEAIEGK + Carbamidomethyl (C)

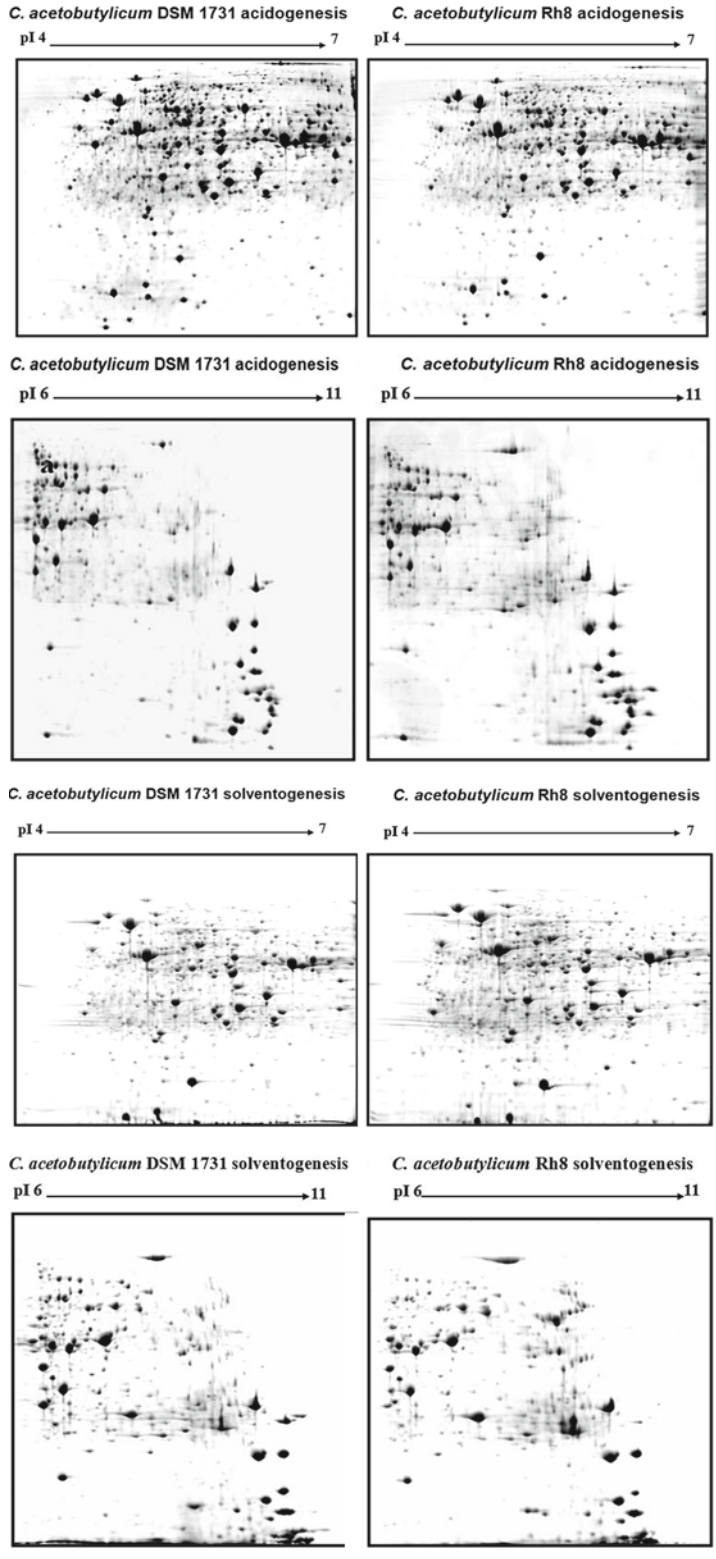


Fig. 4. Gel images of wild-type DSM 1731 and its mutant Rh8 during acidogenesis and solventogenesis, results of one of three technical replicates were shown (12) (reproduced from Figure S1 and S2 in (9) with permission from American Chemical Society).

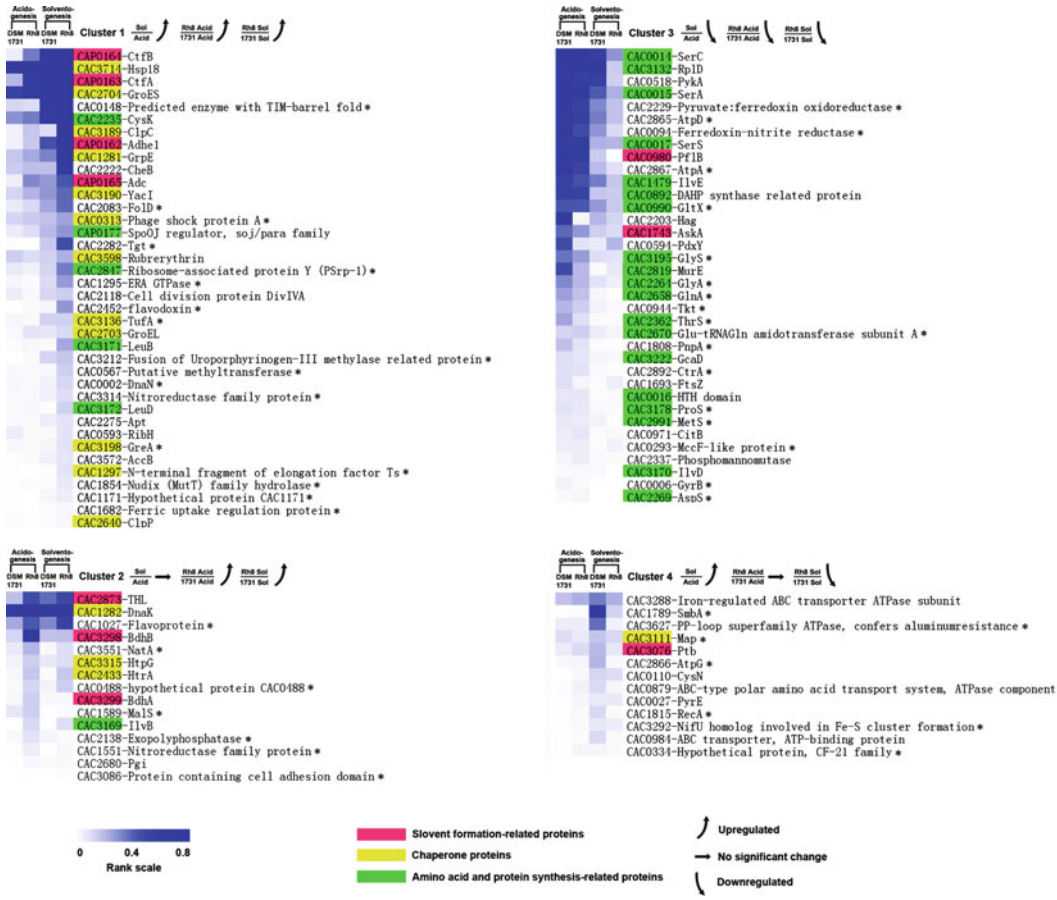


Fig. 5. Hierarchiacal clustering of proteomic data performed using TIGR MeV software. Two phases analysis of differential proteins according to DSM 1731 and Rh8 (12) (reproduced from Fig 5. in (9) with permission from American Chemical Society).

enhanced butanol tolerance and butanol production. With these knowledge we can then, individually or in combination, upregulate those proteins which showed pre-upregulation in acidogenesis by constitutive gene overexpression, or downregulate those proteins which showed pre-downregulation in acidogenesis by antisense RNA technology. Using these targeted genetic modification approaches we can rule out the possible genetic noises introduced by random mutagenesis, thus we can expect to obtain recombinant *C. acetobutylicum* mutants with further improved butanol-producing capability.

4. Notes

1. To obviate batch-to-batch variance, spots that were consistently reproducible in all gel images, including both the biological and technical replicates, were chosen for subsequent analysis.

2. Spots normalization were performed using relative volumes (%Vol) to quantify and compare the gel spots, with the aim to make the data independent of the experimental variations between gels.
3. This process was repeated until all gel spots were completely destained.
4. ACN need to be removed.
5. MS calibration was automatically performed by a peptide standard Kit (Applied Biosystems) containing des-Arg1-bradykinin (m/z 904), Angiotensin I (m/z 1296.6851), Glu1-fibrinopeptide B (m/z 1570.6774), ACTH (1–17, m/z 2903.0867), ACTH (18–39, m/z 2465.1989), and ACTH (7–38, m/z 3657.9294) and MS/ MS calibration was performed by the MS/MS fragment peaks of Glu1-fibrinopeptide B. All MS mass spectra were recorded in the reflector positive mode using a laser operated at a 200 Hz repetition rate with wavelength of 355 nm. The accelerated voltage was operated at 2 kV.
6. Searching parameters were as follows: trypsin digestion with one missed cleavage, variable modifications (oxidation of methionine and carbamidomethylation of cysteine), and the mass tolerance of precursor ion and fragment ion at 0.2 Da for +1 charged ions.
7. The relative expression levels of differential expression proteins were hierarchical clustering analysis according to relative volumes (%Vol) of the protein spots.

References

1. Stephanopoulos, G. (2007). Challenges in engineering microbes for biofuels production. *Science* **315**, 801–804.
2. Rude, M. A. and Schirmer, A. (2009). New microbial fuels: a biotech perspective. *Curr Opin Microbiol* **12**, 274–281.
3. Lee, S. K., Chou, H., Ham, T. S., Lee, T. S. and Keasling, J. D. (2008). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol* **19**, 556–563.
4. Jones, D. T. and Woods, D. R. (1986). Acetone-butanol fermentation revisited. *Microbiol Rev* **50**, 484–524.
5. Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, J. and Jung, K. S. (2008). Fermentative butanol production by *Clostridia*. *Biotechnol Bioeng* **101**, 209–228.
6. Ezeji, T. C., Qureshi, N. and Blaschek, H. P. (2007). Bioproduction of butanol from biomass: from genes to bioreactors. *Curr Opin Biotechnol* **18**, 220–227.
7. Zhang, Y. P., Zhu, Y., Zhu, Y. and Li, Y. (2009). The importance of engineering physiological functionality into microbes. *Trends Biotech* **27**, 664–672.
8. Wackett, L. P. (2008). Biomass to fuels via microbial transformations. *Curr Opin Chem Biol* **12**, 187–193.
9. Jia, K. Z., Zhang, Y. P. and Li, Y. (2010). Systematic engineering of microorganisms to improve alcohol tolerance. *Eng Life Sci* **10**, 422–429.
10. Yan, Y. and Liao, J. C. (2009). Engineering metabolic systems for production of advanced fuels. *J Ind Microbiol Biotechnol* **36**, 471–479.
11. Mukhopadhyay, A., Redding, A. M., Rutherford, B. J. and Keasling, J. D. (2008). Importance of systems biology in engineering microbes for biofuel production. *Curr Opin Biotechnol* **19**, 228–234.
12. Mao, S. M., Luo, Y. A. M., Zhang, T. R., Li, J. S., Bao, G. A. H., Zhu, Y., Chen, Z. G., Zhang, Y. P., Li, Y. and Ma, Y. H. (2010). Proteome

reference map and comparative proteomic analysis between a wild type *Clostridium acetobutylicum* DSM 1731 and its mutant with enhanced butanol tolerance and butanol yield. *J Proteome Res* **9**, 3046–3061.

13. Elias, J. E. and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207–214.
14. Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C. M., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W. and Dunn, M. J. (2006). Guidelines for the next 10 years of proteomics. *Proteomics* **6**, 4–8.
15. Trauger, S. A., Kalisak, E., Kalisiak, J., Morita, H., Weinberg, M. V., Menon, A. L., Poole, F. L., Adams, M. W. W. and Siuzdak, G. (2008). Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J Proteome Res* **7**, 1027–1035.
16. Zhao, B., Yeo, C. C. and Poh, C. L. (2005). Proteome investigation of the global regulatory role of sigma 54 in response to gentisate induction in *Pseudomonas alcaligenes* NCIMB 9867. *Proteomics* **5**, 1868–1876.
17. Dong, M., Yang, L. L., Williams, K., Fisher, S. J., Hall, S. C., Biggin, M. D., Jin, J. and Witkowska, H. E. (2008). A "Tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J Proteome Res* **7**, 1836–1849.
18. GEHealthcare (2004). 2-D Electrophoresis principles and methods. pp. 10–13.

Chapter 11

Metabolic Engineering of Antibiotic-Producing Actinomycetes Using In Vitro Transposon Mutagenesis

Andrew R. Reeves and J. Mark Weber

Abstract

A program of mutation and screening, with stepwise reverse engineering or "decoding" of the improved strain, is a way to better understand the genetics and physiology of the strain improvement process. As more is learned about the genetics of strain improvement, it is hoped that more fundamental principles will emerge about the types of mutations and genetic manipulations that reliably lead to higher producing strains. This will accelerate the construction of higher producing strains by metabolic engineering in the future. In this chapter, a detailed tagged mutagenesis approach is described using in vitro transposon mutagenesis which allowed the successful identification of key genes involved in macrolide (erythromycin) antibiotic biosynthesis.

Key words: In vitro transposon mutagenesis, Reverse engineering, Actinomycete, Protoplast transformation, Erythromycin, *Aeromicrobium erythreum*, Microtiter fermentation

1. Introduction

Industrial microbiologists have been using the classical mutate-and-screen approach to strain improvement (1, 2) for over 60 years, and although the method works well for producing improved strains, very little knowledge has been gained over the years of how it works. Microbial geneticists, on the other hand, have accumulated a vast array of genetic tools and information over the same period of time, but have had limited success in converting their knowledge to commercial use.

Recently, the application of rapidly evolving technologies, including in vitro transposon mutagenesis (3, 4) and especially low-cost comparative genomic sequencing (5), has made it possible to determine how the classical mutate-and-screen method actually generates better strains. The idea goes beyond the reverse engineering of an improved strain to reverse engineering the actual

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_11,

process for producing improved strains. Once the process is reverse engineered over a cycle or several cycles, the principles that are subsequently revealed should allow the direct construction of improved strains in the future, rather than relying on the luck of random mutagenesis. It may even be possible that the information gained will have applications to other strain improvement processes in higher organisms. Work in this field has already begun to produce interesting results (6–10).

Using the in vitro transposition approach, each improved strain generated by random mutagenesis is reverse engineered before proceeding to the next round of mutagenesis. If comparative genomic sequencing were used instead (11, 12), then each strain in the lineage of a commercial process would be compared by sequence analysis to the strain that preceded it. By focusing on the early stages of strain improvement, where the biggest gains in production are typically seen, the most important principles should be revealed. Possibly, only one or two rounds of reverse engineering will be necessary before a useful picture emerges of what happens during a mutate-and-screen strain improvement program.

This chapter outlines the materials and methods needed to perform a reverse engineering or "decoding" of the microbial strain improvement process for *Aeromicrobium erythreum* NRRL B-3381, but can be generally applied to any organism of commercial value or process of interest. An important upfront consideration is first to determine whether in vivo transposon mutagenesis works in the organism of interest. If so, then an efficient gene transfer system would need to be developed if none is currently in place. In vivo transposon mutagenesis has advantages and perhaps is preferred over in vitro transposon mutagenesis, for such reasons as reduced cost, time, and labor to get to the screening stage. However, most organisms do not have efficient in vivo transposon expression systems in place. Furthermore, the mutagenesis process has been found to be biased toward "hotspot" insertion sites and thus would not generate a useful, high-quality knockout library. In vitro transposon mutagenesis offers more control over the randomness of the insertions and thus more complete coverage of the genome-wide mutagenesis process. Moreover, in in vitro transposon mutagenesis the quality control process can be more easily quantified and if a reasonably efficient gene transfer system is in place, it is likely a high-quality knockout library will be achieved in the host strain.

A strain at any point in the engineering process can be a candidate for a decoded mutate-and-screen program, but more information is obtained when the procedure begins with a wild-type strain, since then the entire strain improvement process becomes known. If the process starts with an improved strain, then only the later parts of the strain improvement process can be learned, which means the potentially most important improvement steps would remain unknown. Low-cost, next generation comparative genomic

sequencing can be used as an alternative approach using random mutagenesis with UV or chemicals, but these often reveal many mutations per cycle, thus making it difficult to pinpoint the strain improvement targets. This can be especially true if several rounds of random mutagenesis preceded a detailed genomic analysis. The comparative whole genome sequencing approach is most useful for programs in the very early if not the beginning stage of the strain improvement process when fewer neutral, deleterious, and beneficial mutations have accumulated.

Since an in vivo transposition method is not available for erythromycin-producing organisms, including *A. erythreum*, or for most actinomycetes in general, an in vitro transposition method is used instead. *A. erythreum* and many other filamentous actinomycetes have gene manipulation tools in place. The actinomycete DNA is mutated in *Escherichia coli* then transferred back into the host of choice in a later gene replacement step. In vitro transposon insertion kits are available for *E. coli* from Epicentre Technologies (Madison, WI) (EZ-Tn5 <R6Kγ/KAN-2> or EZ-Tn5 <oriV/KAN-2>; Fig. 1a). Although the transposons can be customized with different drug resistance genes, outward reading promoters, and other features (8, Fig. 1b), they can also be used unmodified. The advantage of using them unmodified is that the available transposons are smaller (~2 Kb). Avoiding incorporation of promoters in the transposon simplifies the analysis of the mutant phenotypes which occurs during the reverse engineering process, but limits the possible mutant phenotypes.

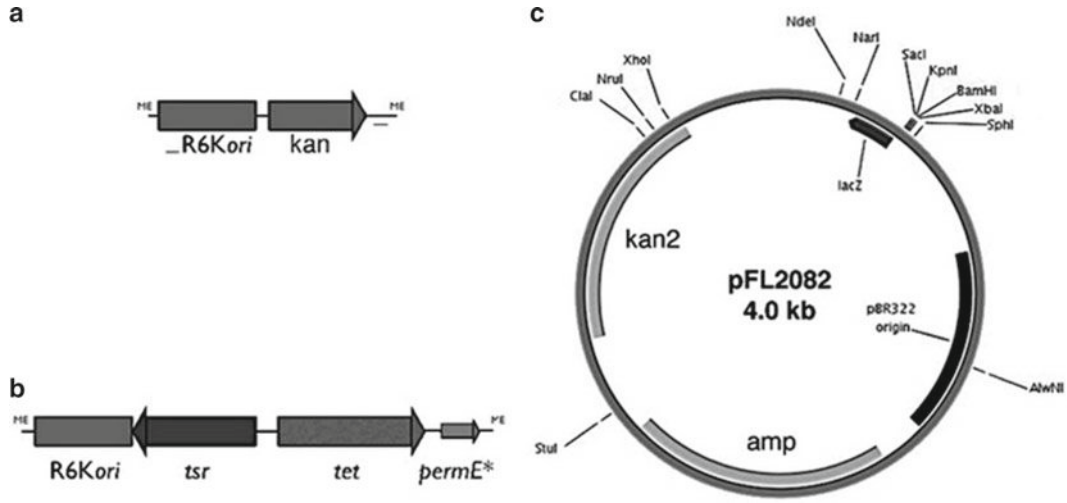


Fig. 1. (**a**) Unmodified EZ-Tn5 <R6Kγori/kan2> transposon from Epicentre Technologies. (**b**) Modified EZ-Tn5 <R6K*γori*/kan2> derivative used for in vitro transposition in *A. erythreum*. (**c**) Plasmid pFL2082 used in *A. erythreum*. Abbreviations: R6K*γori E. coli* origin of replication; *kan* kanamycin resistance gene from *Tn*916; *tsr* thiostrepton resistance gene; *tet* tetracycline resistance gene*; permE** modified *erm*E promoter; *amp* ampicillin resistance gene; *kan*2 kanamycin resistance gene on transposon.

As described in this chapter, the mutagenic procedure is divided into three major steps corresponding to the formation of two plasmid libraries in *E. coli*, and the final one in the actinomycete of interest. As a guide to the quality of the libraries, an analysis of the randomness of the libraries and mutants is included at each step of the procedure.

In vitro transposon mutagenesis has been successfully used on the erythromycin-producing organism *A. erythreum*, which led to a greater understanding of the role that precursor feeding plays in strain improvement of macrolide-producing organisms. In this study, 3,049 Tn mutants were screened in duplicate fermentations and analyzed for a significant percent increase or decrease in erythromycin titers. Seventy of these mutants showed unusually high or low erythromycin titers compared to the parent strains in a high-throughput microtiter screening process. Twenty-six mutants showed increased titers of greater than 25% above parent controls and 44 showed either a 25% or greater decrease in production or no erythromycin production at all. Analysis of the antibiotic resistance phenotypes of the first 2,267 mutants indicated that 88% of the mutants contained a single-crossover insertion. The remaining 12% were gene-replaced strains. The high percentage of single-crossover strains led to the development of a prescreening procedure (described in Subheading 3.5, Note 28 and in Fig. 4) based on a colorimetric assay that could distinguish gene-replaced (double crossover) strains from strains with an integrated plasmid (single crossover). In total, 1,070 gene-replaced transposon mutants were screened which would represent about 27% coverage of the predicted 4 Mb *A. erythreum* chromosome assuming random insertion events.

Further screening and retesting of selected mutants from the primary screen identified seven high-producing mutants. These mutants consistently produced greater than a 50% increase in erythromycin titers when compared to the parent strain (8). Three of the seven mutants mapped to *mutB*, encoding the alpha subunit of methylmalonyl-CoA mutase. The remaining four mapped to *cobA*, encoding a cob(I)alamin adenosyltransferase, involved in one of the later steps of vitamin B_{12} biosynthesis. The metabolic connection of the *cobA* mutation is that methylmalonyl-CoA mutase is one of the few enzymes in the bacterial cell that has a requirement for vitamin B_{12}.

In summary, an in vitro transposon mutagenesis approach was used in an erythromycin-producing organism to create improved strains. The improved strains were reverse engineered and found to have mutations in genes clearly involved in precursor feeding into the erythromycin biosynthetic pathway. In follow-up studies, *mutB* knockout strains were generated in the commercially important erythromycin-producing organism *Saccharopolyspora erythraea* and found to have similar positive strain improvement effects on erythromycin production (13, 14).

2. Materials

To ensure high-quality libraries, all reagents were of molecular biology grade and only double-distilled water should be used (18 Ω at 25°C).

2.1. Preparation of 10–15 kb Chromosomal Library (Library 1)

1. Growth medium (see Note 1): Modified Soluble complete medium (MSCM), 15.0 g Soluble Starch, 20 g Soytone, 1.5 g Yeast extract, 26.25 g MOPS buffer, 0.1 g Calcium Chloride dehydrate; distilled water to 1 L. Adjust before autoclaving to pH 7.5 with NaOH. After autoclaving add, per liter: 30.0 mL of 50% w/v D-(+)-glucose dissolved in distilled water, 2.0 mL of trace elements solution (15), 5 mg of Multivitamin.

   Trace element solution (added in this order): 0.040 g of zinc chloride ($ZnCl_2$), 0.2 g ferric chloride hexahydrate ($FeCl_3 \cdot 6H_2O$), 0.010 g cupric Chloride dihydrate ($CuCl_2 \cdot 2H_2O$), 0.010 g manganese chloride tetrahydrate ($MnCl_2 \cdot 4H_2O$), 0.010 g sodium borate decahydrate ($Na_2B_4O_7 \cdot 10H_2O$), 0.010 g ammonium molybdate tetrahydrate ($(NH_4)_6Mo_7O_{24} \cdot 4H_2O$), Distilled water, to 1,000 mL.
2. All molecular biology techniques use standard methods and can be adjusted as necessary to the organism of interest. All techniques should be carried out according to the manufacturer's instructions for best results (see Note 2).
3. Lysing solution: 25 mM EDTA in 25 mM Tris pH 8.0 and lysozyme was used at a final concentration of 1 mg/mL.
4. One unit of *Mbo*I used per 2 μg chromosomal DNA. Incubated for 5 min at 37°C. Stop solution: 2 μL 0.5 M EDTA and 5 μL electrophoresis running dye for a 30 μL reaction. Can adjust accordingly. (The *MboI* concentration can be adjusted as needed to favor production of 10–15 kb fragments).
5. One Weiss unit of T4 DNA ligase was used in an overnight incubation at 16–22°C.
6. *Escherichia coli* DH5α electrocompetent cells were used at 2.5 KV and recovered in SOC medium (15).
7. 2XYT medium per liter: 16.0 g Tryptone, 10.0 g Yeast extract, 5.0 g sodium chloride. Adjust to pH 7.0 with NaOH, add distilled water to 1,000 mL. Ampicillin selection was performed at 100 μg/mL in 2× YT agar.
8. X-gal indicator (bromo-chloro-indolylgalactopyranoside) was added to a final concentration of 32 μM. IPTG inducer solution (isopropyl-β-D-thiogalactopyranoside) was used at a final concentration of 160 μM.
9. Single-colony isolates were grown in Luria Bertani (LB) medium-containing ampicillin sodium salt at 100 μg/mL.

2.2. Preparation of Transposon Mutagenized Library in E. coli (Library 2)

All reagents and reactions should be carried out carefully and according to manufacturer's conditions for best results. The amount of DNA can be adjusted up or down but the ratio of transposon to chromosomal DNA needs to be carefully determined for effective mutagenesis.

1. 0.2 μg of library 1 DNA in all in vitro transposition reactions using the EZ::Tn mutagenesis kit (Epicentre Technologies, Madison, WI; see Note 3).
2. 2×YT agar-containing ampicillin sodium salt at 100 μg/mL and kanamycin sulfate at 50 μg/mL.
3. Use PCR primers Kan-2 FP1 and Kan-2 RP1 (Epicentre Technologies, Madison, WI) to identify regions where the transposon inserted as a gauge of randomness.

2.3. Preparation of Protoplasts

1. 2×PT Buffer (per liter): 200.0 g sucrose, 0.50 g potassium sulfate, 10.17 g magnesium chloride hexahydrate, 4.0 mL trace elements solution. Distilled water to 750 mL. At time of use complete by adding to 150 mL of above 2×PT solution: 10.0 mL of 1 M Calcium Chloride dehydrate, 40.0 mL of 0.25 M TES buffer, pH 7.2.
2. 1× modified P Buffer (adapted from ref. (14)): Make up the following basal solution: 103 g sucrose, 0.25 g K_2SO_4, 2.0 g $MgCl_2 \cdot 6H_2O$, 2 mL trace element solution, 800 mL distilled water. Dispense in 80 mL portions and autoclave. Before use add to each flask in order: 10 mL of 3.68% $CaCl_2 \cdot 2H_2O$, 10 mL of 5.73% TES buffer adjusted to pH 7.2 (see Note 4).
3. 50% PEG solution: 50.0 g polyethylene glycol (ave. mol.wt 10,000), Distilled water to 100 mL.
4. 10.3% sucrose solution: 10.3 g sucrose in ddH_2O to 100 mL.
5. 1 mg/mL Lysozyme in distilled water.
6. 0.4% Glucose.

2.4. PEG-Mediated Transformation

1. Fresh or frozen protoplasts can be used. Thaw rapidly at 37°C and aliquot to individual eppendorf tubes.
2. R2T2 agar (per liter): 103 g Sucrose, 0.25 g Potassium sulfate, 6.5 g Yeast extract, 5 g Tryptone, 22 g Agar, 850 mL Distilled water. After autoclaving the following sterile solutions are added: 20 mL of 50% w/v D-(+)-glucose, 25 mL of 1 M Trizma Base (pH 7.0), 5.0 mL 0.5% Potassium phosphate monobasic, 2.5 mL of 1 N Sodium hydroxide, 50 mL 1 M Calcium Chloride dehydrate, 50 mL of 1 M magnesium chloride hexahydrate, 2 mL Trace elements. Plates should be dried to 90% of original weight (see Note 5).
3. 100 mg/mL stock Thiostrepton in 50% ethanol and 50% DMSO.

2.5. Preparation of Host Integration Library (Library 3)

1. DNA concentration greater than 0.2 μg/μL.
2. Thiostrepton at 25 mg/mL in 50% ethanol and 50% DMSO. Store at −20°C.
3. Kanamycin sulfate at 100 mg/mL in sterile distilled water. Store at −20°C.
4. Prepare selection media plates if an assay to distinguish between single-crossover and double-crossover insertions is available (see Note 6; Fig. 4).
5. Use primers KanFP-1 and Kan RP-1 to sequence from ends of the transposon into the adjacent chromosomal DNA (Epicentre Technologies, Madison, WI); (see Note 7).

2.6. Microtiter Screen

1. Alligator magnetic stirrer system (V&P Scientific, San Diego, CA; see Note 8).
2. Power setting 3 using steel magnetic stir rods.
3. 96-deep well plates with 1 mL capacity.
4. Prepare a lid to cover the 96-well microtiter dish lid by taping tightly tissue paper over the lid to create a seal between deep-well plate and lid.
5. Sterile glycerol, 80%.

2.7. Shake Flask Screen

1. Sterile MSCM broth (see Subheading 2.1).
2. Add 1/10 v/v of thawed glycerol stock to sterile medium.
3. Add 1/20 v/v of inoculum into fermentation flask.

2.8. Bioassay

1. *Bacillus subtilis* strain.
2. 1% solution of 2, 3, 5 triphenyltetrazolium chloride (tetrazolium red) in double-distilled H_2O.
3. Tryptic Soy Broth (per liter distilled water): 17 g Tryptone (Pancreatic Digest of Casein), 3 g Soytone (Pancreatic Digest of Soybean Meal), 2.5 g dextrose, 5 g NaCl, and 2.5 g dipotassium phosphate.
4. Erythromycin dilution series made from a 1 M stock using pure erythromycin resuspended in pure ethanol.
5. 245-mm Bioassay dish (Corning Costar, Cambridge, Mass).
6. 1/4-in paper bioassay discs (Schleicher & Schuell, Keene, NH).

2.9. Small-Scale gDNA Prep

1. Lysozyme solution: 2 mg/mL in 0.3 M sucrose, 25 mM Tris buffer pH 8, 25 mM EDTA pH 8, and 50 μg/mL RNase A.
2. 2% SDS solution in water.
3. Phenol: Chloroform: isoamyl alcohol (25:24:1).
4. 3 M sodium acetate pH 5.2.

5. Isopropanol.
6. 70% Ethanol: contains 70 mL of 100% ethanol and 30 mL of ddH_2O.
7. TE buffer: contains 10 mM of Tris base pH 8.0 and 1 mM of EDTA, pH 8.0.

2.10. Plasmid Rescue

1. DNA concentration at least 0.2 µg/µL.
2. Use a frequent cutting restriction enzyme. For high G+C% actinomycetes these could be *Apa*I, *Stu*I, *Bss*HI, *Xho*I, *Sma*I, and *Nde*I.
3. Ligate using T4 DNA ligase and incubate overnight at room temperature.
4. Clean up ligation reaction using QiaQuick kit (Qiagen, Valencia, CA).
5. Transform *E. coli* pir+strain EC100D cells and select with tetracycline at 10 µg/mL.
6. Use transposon-based sequencing primers Kan-2 FP1 and Kan-2 RP1 to localize Tn insertions.

3. Methods

All enzymes or temperature sensitive reagents should be kept as briefly as possible at room temperature or 4°C.

3.1. Preparation of 10–15 kb Chromosomal Library (Library 1)

1. Prepare chromosomal DNA as follows. Resuspend 500 mg of cells wet weight from 3.0 mL of a 2-day MSCM culture in 5.0 mL of lysozyme solution and incubate at 37°C for 30 min or until cells become somewhat translucent.
2. Add 2.5 mL of a 2% SDS solution and mix on vortex for 1 min, this will reduce viscosity and clear solution noticeably.
3. Add 2.5 mL of phenol-chloroform solution (pH 8.0), mix on vortex for 30 s, centrifuge for 2 min in a microcentrifuge.
4. Remove 3.0 mL of the supernatant, leaving the white interface behind.
5. Add 0.1 volume of 3 M sodium acetate, pH 5.2 and mix on vortex, then add 0.6–1 volume of isopropanol and mix by inverting tube several times, a small clump of DNA should form and visibly fall to the bottom of the tube.
6. Centrifuge for 2 min in a microcentrifuge and a small white spot of DNA appears at the bottom of the tube.
7. Carefully pour off supernatant and wash pellet twice with 5.0 mL of 70% EtOH, drain liquid off and air dry for 20 min.

8. Dissolve pellet in 1.0 mL of TE 8.0 or double-distilled water. Mixing and pipetting may be required to get the DNA in solution.
9. Partial digestion of 2 μg of chromosomal DNA into 10–15 kb DNA fragments using restriction enzyme *MboI* (see Note 9). The partial digestion is performed as follows: Transfer 90 μL of the DNA preparation from step 1 to a sterile 1.5-mL tube containing the equivalent of 10 μg chromosomal DNA (may need to dilute stock). Add 10 μL of the appropriate 10× restriction enzyme buffer to the 90 μL DNA solution.
10. Set up five 1.5-mL microfuge tubes (labeled A–E) on ice containing 30 μL of the DNA/buffer mix in tube A, 20 μL of the DNA/buffer mix in tubes B–D, and 10 μL of the DNA/buffer mix in tube E.
11. Add 10 units of restriction enzyme to tube A, mix thoroughly and transfer 10 μL of reaction mix to tube B using a fresh pipette tip. Continue serial transfers of 10 μL restriction enzyme mix from tube B to tubes C–E. All tubes should contain a final volume of 20 μL.
12. Incubate tubes A–E at 37°C for 15 min. At the end of the incubation period, rapidly transfer tubes to ice water bath and add 2.0 μL of 0.5 M EDTA and 5 μL of electrophoresis running dye to stop the restriction enzyme reaction.
13. Resolve chromosomally restricted fragments on a 0.6% agarose gel, excise fragments in 10–15 kb region, and purify using Qiaex kit or other suitable gel purification kit. Chromosomal fragments are concentrated and checked for degradation on a 0.6% agarose gel.
14. Perform ligation reaction using a fivefold molar excess of chromosomal fragment to vector. Ligation reactions should be performed at 16°C or at room temperature (depending on manufacturer's instructions) using T4 DNA ligase (ref. (15); see Note 10; Fig. 1b, c).
15. Transform a high electrotransformation efficiency *E. coli* (DH5 alpha, JM110, XL1) with ligated DNA, select for ampicillin-resistant transformants at 100 μg/mL.
16. Screen visually for white transformants on X-gal + IPTG supplemented agar plates.
17. Restreak white ampicillin-resistant transformants for single colonies on X-gal supplemented agar plates.
18. Pick several single colonies from streaks to inoculate 4-mL test tube cultures containing LB with ampicillin (100 μg/mL).
19. Perform high-purity DNA preparations on overnight (16 h) cultures.

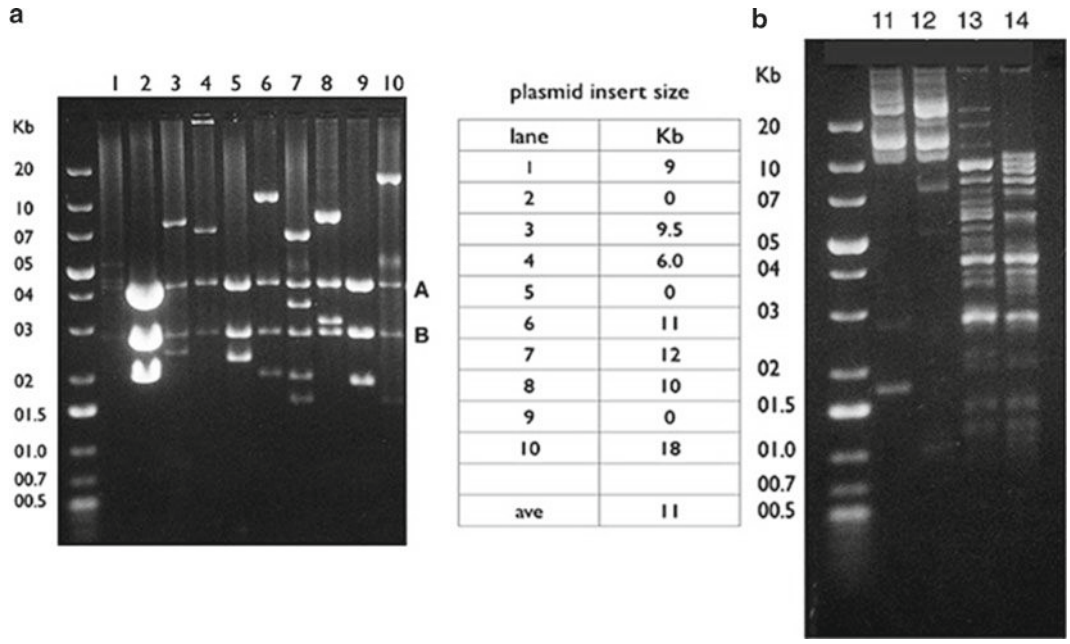


Fig. 2. (**a**) Representative library 1 plasmids subjected to *PstI* digestion and agarose gel electrophoresis. Bands labeled (**a**, **b**) and are vector-related bands; the sizes of the remaining bands in each lane are added together (minus 1.5 kb for the size of the band that was cloned) to determine the size of the cloned DNA. Only plasmids with inserts in the range of 9–15 kb were chosen for Library 2. Deleted plasmids can be seen in lanes 2, 5, and 9. A table showing the calculated size of the plasmid inserts is shown to the right. (**b**) Library 2 DNA, showing lanes 11 and 12 which are uncut Qiagen-kit preparations, lanes 13 and 14 are cut with *Pst*I. DNA size ladders are shown in the left-most lane of each gel.

20. Analyze plasmid DNA using suitable enzymes that cut outside the *Bam*HI site on agarose gels to determine the size of the insert DNA and relative copy number of the plasmid (see Fig. 2a, b; Notes 11 and 12).

3.2. Preparation of Tn Mutagenized Library in E. coli (Library 2)

1. Combine 45–50 plasmids (approximately 500 kb of chromosomal DNA fragments) from the same copy number group from Library 1 (see Note 13).
2. Perform an in vitro transposition reaction on each pool of plasmids from Library 1 combining similar copy number plasmids in each reaction.
3. Prepare highly purified target DNA for the in vitro transposon mutagenesis reaction. In this study pFL2082 containing an average insert size of 10–15 kb was used. Confirm that no chromosomal DNA is present since this will compete with the target DNA for Transposon insertion sites and reduce efficiency.
4. Tn insertion mutagenesis reaction is performed using 0.2 μg of target DNA and equimolar amounts of transposon (see Note 14).

5. Add an equimolar amount of transposon, 1 μL of 10× reaction buffer, 1 μL of transposase, and ddH_2O to a final volume of 10 μL.
6. Incubate at 37°C for 2 h.
7. Stop transposition reaction by adding 1 μL of 10× EZ <Tn> stop buffer.
8. Heat at 70°C for 10 min.
9. Prepare reaction mix for electrotransformation by diluting fivefold with ddH_2O to dilute salt concentration or by purifying the reaction mix with a reaction clean up kit.
10. Transform the transposon-mutagenized plasmid DNA pool into an *E. coli* strain containing the pir^+ protein which is required for expression of the R6Kγ origin of replication (see Note 15).
11. Select for ampicillin and kanamycin-resistant transformants.
12. Assess the randomness of Library 2 by performing DNA sequence analysis from the ends of the transposon (see Notes 16 and 17).
13. Harvest approximately 6,000 or more primary transformant colonies off agar plates from each pool of primary transformants.
14. Perform a high-purity and high-concentration DNA preparation on the harvested colonies (see Note 18).
15. Analyze high-purity DNA preparations by restriction digestion and gel electrophoresis to assess the diversity of DNA fragments in the DNA pool (see Note 19).

3.3. Protoplast Preparation (Adapted from Refs. (16 and 17))

1. Grow 20 μL of a dense glycerol stock of *A. erythreum* B-3381 in test tubes containing 4 mL 2XYT + glucose (0.4%) and incubate overnight at 33°C, 350 rpm (see Note 20).
2. Next day, transfer 2.5 mL of the overnight culture to 25 mL of fresh 2XYT + glucose (0.4%) in a 250-mL shake flask and incubate for an additional 2–4 h.
3. Pellet the cells at 4,270 × *g* for 7 min.
4. Wash the pellet with 20 mL of 0.3 M sucrose, centrifuge as before, and resuspend cells in 5 mL of 1× modified P buffer + 5 mg lysozyme per mL.
5. Incubate cells at 33°C for 1 h with occasional shaking.
6. After 1 h add 6 mL of 1× modified P buffer and incubate another 1 h.
7. Centrifuge at 4,270 × *g* for 10 min.
8. Wash protoplasts with 12.5 mL of 1× modified P buffer and centrifuge again.
9. Resuspend in 750 μL of 1× modified P buffer.
10. Freeze at −80°C or use fresh (see Note 21).

3.4. PEG-Mediated Transformation

1. Mix 50 µL of protoplasts with 5 µL of highly purified and concentrated DNA and 220 µL of 25% PEG (10,000 mw) solution (see Note 22).
2. For integrative transformation use denatured DNA (see Note 23).
3. Spread the protoplast mixture over an R2T2 plate that has been dried to ~95% of its original weight. This can be performed in a sterile laminar flow hood. Incubate at 33°C (see Note 24).
4. Twenty-four h later, overlay the regenerated cells with 1 mL of 800 µg/mL thiostrepton solution or drip proportionately smaller volumes onto sectors.
5. Spread the thiostrepton solution evenly over sector and place in a sterile laminar flow hood until all the liquid is absorbed.
6. Score transformants 3–4 days later (Fig. 3).

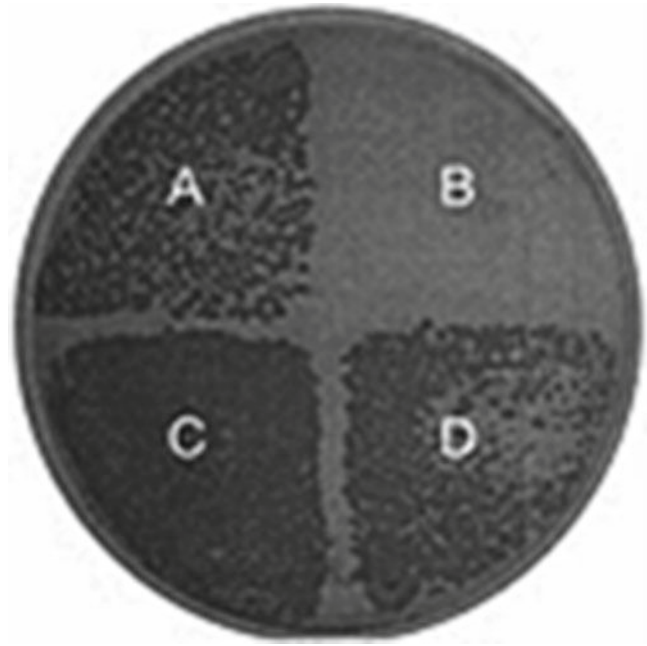


Fig. 3. Primary transformants of *A. erythreum* selected with thiostrepton using the procedure described in Subheading 3.4. *A. erythreum* protoplasts were transformed by the PEG-mediated method using purified, denatured DNA. (**a**). 8 µg denatured plasmid DNA purified by anion-exchange column; (**b**) no DNA added; (**c**) 16 µg denatured plasmid DNA isolated by a rapid alkaline lysis method; (**d**) 8 µg denatured plasmid DNA isolated by a rapid alkaline lysis method. All quadrants containing DNA resulted in confluent or nearly confluent lawns of transformants which were later tested for antibiotic resistance and shown to be recombinants. In this experiment, the amount of denatured plasmid DNA used had the greatest effect on the number of primary transformants obtained.

3.5. Preparation of Host Integration Library (Library 3)

1. Transform *A. erythreum* protoplasts with highly pure and concentrated DNA from each pool from Library 2 (see Note 25).
2. Select primary transformants with thiostrepton (see Subheading 3.4).
3. Plate primary transformants for single colonies on 2XYT with kanamycin (15 μg/mL) plates.
4. Replica plate single colonies to diagnostic media to determine kanamycin and thiostrepton drug resistance phenotypes (see Note 26).
5. Choose colonies with the proper drug resistance phenotypes for −80°C storage and further analyze in fermentation screening process (see Note 27).
6. Perform visual inspection of stained colonies or other assay to distinguish single-crossover integrants from double-crossover (gene replaced) integrants (see Fig. 4; Note 28).
7. Assess randomness of the library by plasmid rescue (see Subheading 3.10; Note 29).
8. Analyze transposon insertion site gene sequences and determine how the insertion might have led to the improved strain's production phenotype. Correlate insertion mutation into existing models of strain improvement (see Note 30).

3.6. Microtiter Plate Screening

These instructions assume the use of the V&P Scientific Alligator magnetic stirrer system (see Note 31).

1. For master stock generation, fill a 96-deep well microtiter plate with 500 μL of MSCM medium.

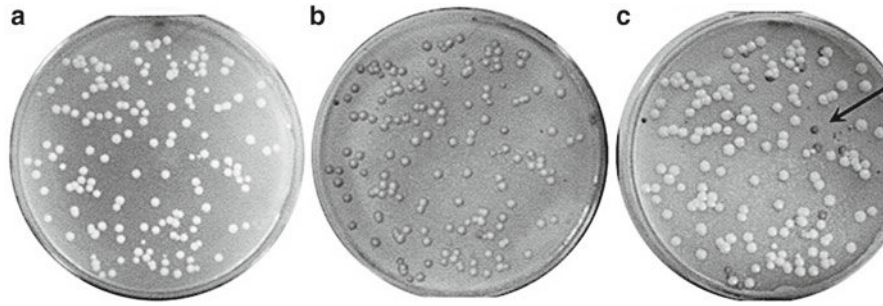


Fig. 4. Colorimetric plate assay to distinguish single-crossover recombinant strains from double-crossover strains using antibiotic selections. (**a**) Dilutions of primary transformants of *A. erythreum* were plated on agar-containing thiostrepton and grown for 4 days at 30°C. In this experiment, the thiostrepton-resistance marker (*tsr*)was carried on the transposon (Fig. 1b), and the kanamycin-resistance marker (*kan*2) was carried on the plasmid (Fig. 1c). (**b**) Same plate 24 h later after being underlayed with an agar slab containing kanamycin (20 μg/mL) and a tetrazolium dye (tetra red, 10 μg/mL). At 24 h incubation, all colonies turned an orange color, indicating uptake of the vital dye. (**c**) The same plate 72 h post underlay. Two different colored colonies appear. The majority of the colonies remain light orange; these are kanamycin-resistant colonies. A smaller percentage of the colonies become dark red. These are the desired double-crossover kanamycin-sensitive colonies (see *arrow*). These colonies can be recovered alive and analyzed in the fermentation screening program.

2. Inoculate wells by dabbing a sterile magnetic stir bar to a single colony from library 3.
3. Place the 96-deep well on the magnetic stirrer and incubate for 72 h at 33°C with lids to prevent evaporation. Maintain relative humidity above 50% to limit evaporation.
4. Add glycerol to 20% and freeze at –80°C until needed for screening.
5. For screening mutants, use 10 μL of the frozen master stock to inoculate 500 mL MSCM medium using a multichannel pipettor.
6. Incubate at 33°C on magnetic stirrer at appropriate power setting for 72 h. Relative humidity should be maintained above 50% to minimize evaporation.

3.7. Shake Flask Screening Method

1. Inoculate seed cultures in unbaffled 250-mL shake flasks with filter closures (0.2 μm pore size), containing 25 mL of MSCM broth, using 250 μL of thawed glycerol stock of culture (see Note 32).
2. Incubate seed cultures at 33°C and 65% humidity on a humidity-controlled incubator-shaker having a 1-in. circular displacement at 380 rpm for 24 h.
3. Use 1.25 mL of the seed culture to inoculate fermentations in 250-mL shake flasks containing 25 mL of MSCM medium.
4. Incubate the fermentations for 3–5 days. At harvest, correct for the loss of volume through the addition of distilled water, if necessary.
5. Perform shake flask screening method on double-crossover recombinants (see Note 33).

3.8. Bioassay of Antibiotic Activity (see Note 34)

1. Pour 100 mL of molten tryptic soy broth agar into the bottom of a large plastic bioassay dish and let harden for 30 min.
2. Pour another 100 mL of molten TSB agar, containing 250 μL of 1% triphenyl-tetrazolium red and a sufficient quantity of *Bacillus subtilis* spores to produce a confluent lawn of growth, onto the first solidified agar layer. Let agar solidify for 1 h.
3. Spot broth samples directly after 1:10 dilution with TE pH 8.0 onto 1/4-in paper bioassay discs, and let dry for 30 min.
4. Place dried bioassay discs onto the surface of the agar plate.
5. Incubate the plate overnight at 37°C.
6. Measure the diameters of the inhibition zones and convert zone diameters to erythromycin concentration using standard curves produced from zone diameters of reference standards on each plate.

3.9. Small-Scale Chromosomal DNA Preparation

1. Resuspend 50 mg of cells wet weight from 300 µL of a 2-day MSCM culture in 500 µL of lysozyme solution and incubate at 37°C for 5 min or until cells become somewhat translucent.
2. Add 250 µL of a 2% SDS and mix on vortex for 1 min; this will reduce viscosity and clear solution noticeably.
3. Add 250 µL of phenol-chloroform solution (pH 8.0), mix on vortex for 30 s, centrifuge for 2 min in a microcentrifuge. Remove 300 µL of the supernatant, leaving the white interface behind.
4. Add 0.1 volume of 3 M sodium acetate, pH 5.2 and mix on vortex, then add 0.6–1 volume of isopropanol, and mix by inverting tube several times; a small clump of DNA should form and visibly fall to the bottom of the tube.
5. Centrifuge for 2 min in a microcentrifuge and a small white spot of DNA appears at the bottom of the tube.
6. Carefully pour off supernatant and wash pellet twice with 500 µL of 70% EtOH, drain liquid off, and air dry for 10 min. Dissolve pellet in 100 µL of TE 8.0 or water. Mixing and pipetting may be required to get the DNA in solution.

3.10. Plasmid Rescue Procedure (see Note 35)

1. Prepare chromosomal DNA from improved strains containing an inserted transposon mutation (see Subheadings 3.1 and 3.9 for method details).
2. Digest the DNA with *Nar*I, *Sal*I, *Apa*I, *Bss*HII, *Stu*I, or *Nde*I (or other frequent cutter enzyme that does not cut within the transposon), using 1 µg of chromosomal DNA (see above procedure) in a 20 µL reaction volume.
3. At end of digestion add 70 µL water, heat at 70°C, 20 min, cool on ice 5 min, return solution to room temperature.
4. Add 10 µL of 10× ligase buffer and 1 µL ligase.
5. Incubate at room temperature for 1 h.
6. Add 1 µL glycogen (20 mg/mL), 10 µl of 3 M sodium acetate (pH 5.2), and 250 µL of cold 100% ethanol.
7. Set on ice for 20 min.
8. Centrifuge at 15,871 × *g* for 15 min at room temperature.
9. Pour off ethanol, revealing small white pellet at bottom of tube.
10. Wash inside of tube with 1 mL of room temperature 70% ethanol. Repeat wash once.
11. Pour off ethanol, let tubes stand inverted for 5–10 min to remove all traces of liquid.
12. Resuspend small white pellets in 25 µL of double distilled water or TE buffer pH 8.0.

13. Add 5 μL of DNA to 20–35 μL of cold electrocompetent EC100D cells in a chilled Eppendorf microcentrifuge tube.
14. Transfer the cell DNA mixture to a cold 0.2 cm cuvette and let sit on ice for 5 min.
15. Electroporate the cells at 2.5 kv, 600 Ω, and 25 microFaradays.
16. Add 1 mL of S.O.C. broth containing the 2× *oriV* inducing agent (Epicentre Biotechnologies; if transposon has the R6Kγori no inducing agent is needed).
17. Incubate the broth-containing cells and the *oriV* inducer at 37°C for 1 h.
18. Plate 20 μL of the cells from the broth directly on LB plates supplemented with kanamycin (40 μg/mL) and the 2× oriV inducer. (This is the "dilute" plating).
19. Centrifuge the broth and cells at 3,220 × *g* for 5 min to gently pellet cells. Decant supernatant and resuspend cells in 50 μL of 20% glycerol.
20. Plate 20 μL of cells in at least a one-quarter to one-eighth sector of a LB plate supplemented with kanamycin (40 μg/mL) and 2× oriV inducer (this is the high-density plating).
21. Incubate at 37°C overnight, colonies should appear by morning. Note that the cells will often appear on the high-density plate at high density and in the dilute plating no colonies may appear.
22. Analyze target site (see Notes 7 and 36).

4. Notes

1. The media used in this method are commonly used for actinomycetes, but any rich medium specifically adapted for growing the host strain can be used. The trace element solution can be made up as a 1,000× stock. Add components in order and mix thoroughly while adding individual components. Protect from light and store at 4°C.
2. Standard molecular biology techniques and reagents were used according to ref. (15).
3. The ratio of target DNA to transposon is important for obtaining a good percentage of mutated plasmids of library 1. When purifying the transposon from an agarose gel limit exposure to UV light as this can damage the 19-bp mosaic ends and significantly reduce efficiency of the in vitro mutagenesis reaction.
4. Potassium phosphate, while important for PEG-mediated transformations of filamentous actinomycetes, was found to have a negative effect with the unicellular *A. erythreum*.

5. R2T2 agar plates need to be dried before PEG-mediated transfer is performed. The degree of dryness for efficient transformation will need to be determined and optimized. For *A. erythreum* 90–95% of the original weight gave the best results, whereas with filamentous actinomycetes plates need to be significantly dry, between 82 and 85%.
6. In this study, we developed a growth and colorimetric assay to distinguish between single- and double-crossover insertion strains since we only wanted to screen strains that had undergone allelic exchange to ensure genetic stability and not use antibiotics for plasmid maintenance which could impact antibiotic titers. The use of this type of assay, which relies on inhibition of growth using a bacteriocidal reagent, kanamycin, requires frequent observation so that gene-replaced colonies can be transferred to fresh nonantibiotic-containing medium shortly after the accumulation of color and cessation of growth to ensure survival.
7. Primers can be purchased directly from Epicentre Technologies or as part of the EZ:Tn in vitro transposon mutagenesis kit. The primers are designed to read off the ends of the Tn5-based transposon to identify the transposon insertion site. They can be used to sequence directly off the transposons from chromosomal templates, or more preferably with rescued plasmid DNA.
8. V&P Scientific has a wide variety of magnetic stirring systems to choose from with different output capacities. A large variety of stirring rods, bars, and discs are available to choose from. We found that the hollow stainless steel stir bars worked best for our application. A series of microfermentation tests was performed and erythromycin titers, well-to-well splashing, and evaporation were checked over a 3-, 4-, and 5-day period. The most consistent data was obtained with the hollow stainless-steel stir rods during a 3-day fermentation period. Longer fermentations worked well but a decrease in culture volume was observed overtime.
9. For library 1 the chromosome of the actinomycete of choice, *A. erythreum*, is cut into 10–15 Kb DNA fragments and cloned into a bifunctional plasmid vector pFL2082 (Fig. 1c). The DNA fragments can be isolated from low percentage agarose gels or by pulsed-field gel electrophoresis (PFGE) to obtain better band resolution.
10. An important consideration for library 1 is the design of the cloning vector. The cloning vector carrying a large insert must replicate stably in *E. coli* and be easily transformed back into the actinomycete of interest for a gene replacement reaction. For *A. erythreum*, plasmid pFL2082 was chosen because it could replicate in high copy number in *E. coli* using the pBR322

replicon from pBS(+) and allows this plasmid to integrate and evict from the chromosome of *A. erythreum* by homologous recombination. The thiostrepton-resistance gene (*tsr*) allows for selection of transformants in *A. erythreum*. The type of vector to be used in Library 1 for each actinomycete will be different and depends on the characteristics of the species of interest. Ideally, the vector itself should be small so that it presents the smallest possible target for the transposon during the in vitro transposition reaction. In the analysis of Library 1 for *A. erythreum*, we found that 10–15 Kb chromosomal inserts could be stably cloned and maintained in pFL2082.

11. Twenty clones from Library 1 were randomly chosen and DNA sequence analysis of the clones showed that they were randomly distributed across the genome. The library 1 clones were visually identified by blue-white screening on X-gal-containing agar plates. Analysis of restriction enzyme digested plasmids on agarose gels revealed that about one-third of the white colonies did not contain inserts but were deleted plasmids (Fig. 2a).
12. It is also possible to observe that the copy number of the plasmids in Library 1 can vary from very low to very high. If that is the case then combine Library 1 plasmids of similar copy number before being used in the in vitro transposition reactions to form Library 2. This helps to maintain a high degree of randomness in the transformants obtained in Library 2.
13. Library 2 is formed by combining plasmids from Library 1 of similar copy number, into pools of approximately 50 plasmids each, representing a combined total of about 500 Kb of chromosomal DNA per pool. The plasmid pool is subjected to an in vitro transposition mutagenesis reaction, and the resulting mutated plasmids are transformed into an *E. coli* strain containing the *pir* protein for generation of individual colonies containing transposon-mutagenized plasmids.
14. Molar amounts of target and transposon DNA are calculated using the following equation: μg target DNA/[target DNA (in bp) × 660]. The average target size of pFL2082 is 19,000 bp, so the amount of target DNA is 0.016 pmoles when starting with 0.2 μg. In this case an equimolar amount of transposon would require only 0.1 μL, since the stock concentration is 0.15 pmoles/μL.
15. In order to select for the transposon insertion and *E. coli* strain containing the pir$^+$ protein is required. Epicentre Technologies has two pir$^+$ *E. coli* strains: EC100D and EC100D pir-116.
16. Each mutated plasmid from Library 2 can be purified and sequenced using the primers that read off the ends of the transposon DNA. In our study, we found that 75% of the plasmids

contained transposon inserts in the cloned chromosomal DNA and the rest in the vector backbone outside of essential functions such as antibiotic markers and replication regions.

17. The reason there were more Tn insertions in the cloned DNA was probably due to the fact that insertions into the vector portion of the plasmid could be non-growers due to knockouts of the antibiotic resistance gene or the origin of replication.
18. The colonies from library 2 are harvested together from primary transformation plates and the plasmid DNA from the mixture is purified using a Qiagen or other suitable plasmid preparation kit. The plasmid preparation is analyzed by restriction enzyme digestion and agarose gel electrophoresis for the presence of a multitude of different sized DNA fragments to indicate randomness of the plasmids in the preparation (Fig. 2b).
19. The plasmids having transposons that missed the insert DNA are still kept with the other plasmids of Library 2 since these plasmids do not affect the generation or quality of Library 3 mutants.
20. Following the *A. erythreum* protoplast transformation procedure carefully leads to reproducible integration of suicide plasmids. Other actinomycetes, for example *Saccharopolyspora erythraea*, are more difficult to transform. In many strains, the age of the cells used for making protoplasts is critical for achieving a high rate of transformation efficiency.
21. Protoplasts suspended in 1×PT buffer can be frozen and thawed multiple times; therefore, once a high-efficiency protoplast preparation is found, it can be useful for several subsequent reactions.
22. If frozen protoplasts are used, spin for 7 min at 845 × *g* in a microcentrifuge to pellet cells and rid supernatant of nucleases. After spin is over, quickly mix in PEG solution and concentrated DNA. Another important factor for obtaining high-efficiency transformations is the use of highly purified and concentrated DNA for the transformation reaction. Qiagen kit (or similar) DNA preparation plasmid kits are recommended. Gentle handling of protoplasts may also be an important factor in achieving high transformation efficiencies which is why low-speed centrifugation is used and pipetting is kept to a minimum. No vortexing of protoplasts is performed.
23. Alkaline denaturation procedure as described in ref. (18). Add 1–5 μg DNA in a small volume into a 1.5-mL microcentrifuge tube. Bring volume up to 9 μL with double-distilled water. Add 2 μL of 1 N NaOH and incubate at 37°C for 10 min. After incubation, quickly place on ice and add 2 μL of 1 N concentrated HCL to neutralize. Use denatured DNA in transformations.

24. The speed of protoplast/PEG puddle absorption into the agar may be an important factor affecting transformation efficiency. To avoid long puddling times, dried regeneration plates are used with level surfaces (created by rotating the plates as they dry in the bio-hood). Also, incubators are kept at low humidity to speed absorption of the protoplast/PEG puddles.
25. Once a lawn of primary transformants is obtained using thiostrepton or kanamycin selection (Fig. 3), the formation of Library 3 gene replacement mutants can begin. Cells from the confluent lawns of primary transformants are passed through one or two rounds of nonselective growth to allow for the spontaneous eviction of the plasmid.
26. The loss of the plasmid can be detected through replica plating of single colonies to diagnostic drug-supplemented agar plates. Single colonies that show loss of the drug resistance marker associated with the plasmid, and the maintenance of the drug resistance marker associated with the transposon, are selected for the fermentation screening process.
27. These double-crossover mutants represent only approximately 5–10% of the population of single colonies when one round of nonselective growth and sporulation is used. These double-crossover mutants will carry the desired transposon insertion without the plasmid sequences, creating essentially the same type of mutation as if the transposon had inserted via an endogenous in vivo transposon event. With two rounds of nonselective sporulation, the frequency of the desired double-crossover mutants is increased significantly.
28. In *A. erythreum* a helpful color reaction was developed to distinguish double-crossover mutants from single-crossover integrants (Fig. 4), and a similar selection can be developed in other systems as appropriate. This visual screening process is significantly easier to perform than the replica-plating process and speeds up the identification of the Library 3 mutants.
29. Using the procedure described here it should be possible to generate thousands of random mutations in the genomes of these erythromycin-producing bacteria. However, it might be expected that the plasmid library approach may not allow saturation mutagenesis of the entire genome because some parts of the chromosome may not be cloned on a high copy plasmid vector in *E. coli*.
30. The transposon could insert into a gene that encodes a hypothetical protein, in which case a new, nonobvious strain improvement target is revealed. Nonobvious targets are the types of mutations that would not normally be created based on our current knowledge of how metabolism can be manipulated to make better strains. Nonobvious targets are therefore

a valuable resource for obtaining new information to broaden our understanding of metabolism.

31. This is the method that was used to find the original *mut*B and *cob*A mutants in the *A. erythreum* screen (8). Because *A. erythreum* is unicellular and not as prone to high-viscosity fermentations, it can be cultured in liquid microtiter fermentations. A rotary tumble stirring apparatus (V&P Scientific Inc., San Diego, CA) can be used. Mycelial organisms are prone to high-viscosity liquid fermentations and are better suited to test tube fermentations at 380 rpm with 1 in. circular displacement. Microtiter fermentations may be a workable alternative, with the disadvantage that there is a larger variation in production level for a particular strain, more false positive hits are generated, and as with any screening program improved strains do not necessarily scale-up to shake flasks or larger scale-up systems without a careful medium design.
32. It is helpful to analyze the new mutant trait under a wide variety of fermentation conditions. And even more can be learned if the mutation is created and studied in different erythromycin-producing organisms. This helps to confirm the significance and validity of the mutant phenotypes found. Many different factors affect antibiotic production levels, and establishing a reliable connection between a mutation and a strain improvement effect takes a comprehensive analysis to be meaningful.
33. Screening procedures for the decoded mutate-and-screen programs are no different than classical programs (2). Our *A. erythreum* shake flask fermentation method has been published (19). Shake flask screening, though reliable, is a laborious and time-consuming process and may not be practical for most small laboratories.
34. One method for measuring erythromycin production is the bioassay (see Subheading 3.8), although colorimetric assays have also been described (18). TLC (8) and HPLC methods (20) are useful for determining the concentration of contaminating coproducts such as erythromycin B and C as well as erythronolide B and mycarosyl-erythronolide B.
35. When the circular DNA is transformed into a pir$^+$ *E. coli* strain, the DNA propagates as a plasmid using the origin of replication (R6Kγ*ori* or *oriV*) contained within the transposon. The plasmid DNA is recovered from *E. coli* and the sequence of the plasmid DNA flanking the transposon is determined using primers that read off both ends of the transposon. Since the procedure is relatively simple and rapid, many mutant hits can be analyzed this way, including mutants that have a decrease in production. Yield-lowering mutants can be just as helpful as

yield-enhancing mutants for the insight they may provide into the secondary metabolic process.

36. Once a genetic target is identified through the location of the transposon insertion site, the reverse engineering is not necessarily complete. If the transposon inserted into an ORF at the end of an operon (i.e., the end of a transcription unit), then the mutagenic effect is most likely due to the loss of function of that gene. However, if the transposon inserted into an ORF that is not at the end of a transcription unit, then the mutant phenotype could be due completely, or in part, to polar effects on downstream genes. A significant amount of follow-up work may be necessary to identify the gene or genes responsible for the mutant phenotype.

Acknowledgments

This work was supported by The National Institutes of Health, Small Business Innovation Research (SBIR) awards R44GM58943 and R44GM063278.

References

1. Queener SW, Lively DH (1986) Screening and selection for strain improvement, pp. 155-169. In: Demain AL and Solomon NA (eds) Manual of Industrial Microbiology and Biotechnology. American Society for Microbiology, Washington, DC.
2. Vinci VA, Byng, G (1999) Strain Improvement by Non-recombinant Methods, p. 103-113. In: Demain AL and Davies JE (eds), Manual of Industrial Microbiology and Biotechnology, 2nd ed. ASM Press, Washington, DC.
3. Goryshin IY, Reznikoff WS (1998) Tn5 *in vitro* transposition. *J Biol Chem* **273**, 7367–7374.
4. Kirby JR (2007) *In vivo* Mutagenesis using EZ-Tn5. *Methods in Enzymology* **421**, 17–21.
5. Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. *Nat Biotechnol* **26**, 1117–1124.
6. 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.
7. Gehring AM, Wang ST, Kearns DB, Storer NY, Losick R (2004) Novel genes that influence development in *Streptomyces coelicolor*. *J Bacteriol* **186**, 3570–3577.
8. Reeves AR, Cernota WH, Brikun IA, Wesley RK, Weber JM (2004) Engineering precursor flow for increased erythromycin production in *Aeromicrobium erythreum*. *Metab Eng* **6**, 300–312.
9. Tannler S, Zamboni N, Kiraly C, Aymerich S, Sauer U (2008) Screening of *Bacillus subtilis* transposon mutants with altered riboflavin production. *Metab Eng* **10**, 216–226.
10. Trötschel C, Kandirali S, Diaz-Achirica P, Meinhardt A, Morbach S, Krämer R, Burkovski A (2003) GltS, the sodium-coupled L-glutamate uptake system of *Corynebacterium* I: identification of the corresponding gene and impact on L-glutamate production. *Appl Microbiol Biotechnol* **60**, 738–742.
11. Ikeda M, Ohnishi J, Hayashi M, Mitsuhashi S (2006) A genome-based approach to create a minimally mutated *Corynebacterium glutamicum* strain for efficient L-lysine production. *J Ind Microbiol Biotechnol* **33**, 610–615.
12. Ikeda M, Mitsuhashi S, Tanaka K, Hayashi M (2009) Re-engineering of an L-Arginine and L-Citrulline Producer of *Corynebacterium glutamicum*. *Appl Environ Microbiol* **75**, 1635–1641.
13. Reeves AR, Brikun IA, Cernota WH, Leach BI, Gonzalez MC, Weber JM (2006) Effects of

methylmalonyl-CoA mutase gene knockouts on erythromycin production in carbohydrate-based and oil-based fermentations of *Saccharopolyspora erythraea*. *J Ind Microbiol Biotechnol* **7**, 600–609.

14. Reeves AR, Brikun IA, Cernota WH, Leach BI, Gonzalez MC, Weber JM (2007) Engineering of the methylmalonyl-CoA metabolite node of *Saccharopolyspora erythraea* for increased erythromycin production. *Metab Eng* **9**, 293–303.
15. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
16. Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton CJ, Kieser HM, Lydiate DJ, Smith CP, Ward JM, Schrempf H (1985) Genetic Manipulation of *Streptomyces*, A laboratory manual. Norwich, UK: John Innes Foundation.
17. Roberts AN, Barnett L, Brenner S (1987) Transformation of *Arthrobacter* and studies on the transcription of the *Arthrobacter ermA* gene in *Streptomyces lividans* and *E. coli*. *Biochem* J **243**, 431–436.
18. Oh SH, Chater KF (1997) Denaturation of circular or linear DNA facilitates targeted integrative transformation of *Streptomyces coelicolor* A3(2): possible relevance to other organisms. *J Bacteriol* **129**, 122–127.
19. Reeves AR, Seshadri R, Brikun IA, Cernota WH, Gonzalez MC, Weber JM (2008) Knockout of the erythromycin biosynthetic cluster gene, *eryBI*, blocks isoflavone glucoside bioconversion during erythromycin fermentations in *Aeromicrobium erythreum* but not in *Saccharopolyspora erythraea*. *Appl Env Microbiol* **74**, 7383–7390.
20. Kanfer I, Skinner MF, Walker RB (1998). Analysis of macrolide antibiotics. *J Chromato* **812**, 255–286.

Chapter 12

Use FACS Sorting in Metabolic Engineering of *Escherichia coli* for Increased Peptide Production

Qiong Cheng, Kristin Ruebling-Jass, Jianzhong Zhang, Qi Chen, and Kevin M. Croker

Abstract

Many proteins and peptides have been used in therapeutic or industrial applications. They are often produced as recombinant forms by microbial fermentation. Targeted metabolic engineering of the production strains has usually been the approach taken to increase protein production, and this approach requires sufficient knowledge about cell metabolism and regulation. Random screening is an alternative approach that could circumvent the knowledge requirement, but is hampered by lack of suitable high-throughput screening methods. We developed a novel fluorescence-activated cell sorting (FACS) method to screen for cells with increased peptide production. Using a model peptide rich in certain amino acids, we showed that increased fluorescence clones sorted from a plasmid expression library contained genes encoding rate-limiting enzymes for amino acid synthesis. These expression clones showed increased peptide production. This demonstrated that FACS could be used as a very powerful tool for metabolic engineering. It can be generally applied to other products or processes if the desired phenotype could be correlated with a fluorescence or light scattering parameter on the FACS.

Key words: Fluorescence-activated cell sorting, Lumio labeling, Peptide production, Inclusion body, Amino acid synthesis, Metabolic engineering

1. Introduction

Flow cytometry is a technique for counting and examining microscopic particles by suspending them in a stream of fluid that passes by a detection apparatus. Fluorescence-activated cell sorting (FACS) is a specialized type of flow cytometry, providing a method for sorting/collecting a heterogeneous mixture of cells based on specific light scattering and/or fluorescent characteristics of each cell (1). Although this technique is still mainly used in mammalian research, there

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_12,

is an increasing interest in biotechnology to utilize it as a powerful tool for library screening and strain development. This technology can have a large impact on metabolic engineering and systems biology (2–4).

Bioactive peptides and proteins are used as curative agents such as insulin (5), interferon (6, 7), and erythropoietin (8) in many therapeutic applications. Other peptides and proteins have found uses in a variety of industrial applications such as the pulp and paper industries and personal care industries (9). Efficient production of the peptides and proteins is the key to the success of the applications. A high-throughput screen is beneficial for isolating strains with enhanced peptide production.

We developed a FACS-based high-throughput screening method for peptide production. The model peptide is a small peptide fused with an inclusion body forming tag so that the fusion peptide can be produced as inclusion bodies (10, 11), which are less prone to protein degradation and easier for downstream recovery. A six-amino acid motif CCPGCC was introduced into the fusion peptide to allow it to be specifically labeled in vivo by fluorescent biarsenical agents. This new system for specific labeling of target proteins in vivo developed in Roger Tsien's laboratory (12) was commercialized by Invitrogen and has seen in many biological applications (13). Using the FACS sorting for high-fluorescence populations, we were able to isolate strains from a genetic library that increased peptide production. Some strains appeared to have relieved the bottleneck (14, 15) for synthesis of the amino acids that are rich in the target peptide.

2. Materials

2.1. DNA Manipulation

1. Wizard Genomic DNA purification kit (Promega, Madison, WI).
2. Isopropanol.
3. Sau3AI and SalI restriction enzymes and reaction buffers.
4. Agarose.
5. 10× Tris-Borate-EDTA buffer (TBE): 1.1 M of Tris, 900 mM of Borate, 25 mM of EDTA, pH 8.0.
6. SYBR Gold Nucleic Acid gel stain (Invitrogen, Carlsbad, CA).
7. Dark Reader Transilluminator (MoBiTec, Gottingen, Germany).
8. DNA polymerase I, Large fragment (Klenow fragment).
9. dGTP/dATP/dCTP/dTTP, 10 mM each.
10. T4 DNA Ligase.
11. QIAquick gel extraction kit (Qiagen, Valencia, CA).

12. ZYMO DNA clean and concentrator (ZymoResearch, Orange, CA).
13. Gene Pulser® Cuvette, 1 mm electrode gap (Bio-Rad Laboratories, Hercules, CA).
14. Qiaprep spin Miniprep kit (Qiagen).

2.2. Cell Culture

1. Luria-Bertani (LB) medium: 10 g/L tryptone, 5.0 g/L yeast extract, and 5.0 g/L NaCl. Add 1.5% agar for agar plates.
2. Ampicillin 100 mg/mL stock in water; Kanamycin 50 mg/mL stock in water.
3. SOC medium: 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, pH 7.0, 20 mM glucose.
4. 50% Glycerol solution.
5. 20% L-arabinose.

2.3. In-Cell Labeling

1. Dulbecco's phosphate buffered saline (PBS) without calcium and magnesium: 0.2 g/L KCl, 0.2 g/L KH_2PO_4, 8 g/L NaCl, 2.16 g/L $Na_2HPO_4{\cdot}4H_2O$.
2. TC-FlAsH II In-Cell Tetracycsteine Tag detection kit (Invitrogen).
3. 10 mM tris-(2-carbonylethylphosphine chloride) (TCEP).

2.4. In-Gel Labeling and SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Lumio™ Green Detection Kit (Invitrogen).
2. NuPage 4–12% Bis-Tris gel.
3. MES SDS Running buffer: 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 50 mM Tris base, 0.1% SDS, 1 mM EDTA, pH 7.3.
4. BenchMark Fluorescent Protein Standard (Invitrogen).
5. Simply Blue™ Safe Stain (Invitrogen).
6. CelLytic Express™ (Sigma, St Louis, MO).
7. 8 M Urea.

2.5. Fluorescence-Activated Cell Sorting

1. FACSVantage™ SE-DiVa Flow Cytometer (BD Biosciences, Franklin Lakes, NJ).
2. ALIGNFLOW488™ 2.5 μm alignment beads (Cat. A7302, Molecular Probes, Inc., Eugene, OR).
3. DiVa software (BD Biosciences).
4. BD Falcon Tube 5 mL, 12×75 mm (BD Biosciences).
5. ACCUDROPS™ (Cat. 345249, BD Biosciences).
6. FACSFlow™ sheath fluid (FACSFlow) (BD Biosciences).
7. Blue Cap Falcon Tube, 35-μm Mesh (BD Biosciences).
8. Methanol.

3 Methods

3.1. Construction of Random Plasmid Expression Library

A random plasmid expression library was constructed on pDCQ601 using 2–4 kb genomic DNA fragments from *E. coli* MG1655 (ATCC47076). Plasmid pDCQ601 contained a multiple cloning site including *Nhe*I and *Sal*I restriction sites introduced at the *Eco*RI site downstream of the Pcat promoter on pBHR1 vector (MoBiTec, Gottingen, Germany).

3.1.1. Isolation of Genomic DNA

The genomic DNA isolation was performed with the Wizard Genomic DNA purification kit as follows:

1. Streak *E. coli* MG1655 strain onto a LB plate, incubate overnight at 37°C.
2. Inoculate a single colony into a culture tube with 3 mL of LB and incubate overnight at 37°C with shaking at 250 rpm.
3. Transfer the culture into 50-mL tube, centrifuge at 6,000×*g* for 10 min.
4. Discard the supernatant; resuspend the pellet in 18 mL of Nuclei lysis solution.
5. Incubate at 80°C for 10 min; cool down at room temperature for 20 min.
6. Add 90 μL of RNase solution and mix; incubate at 37°C for 90 min.
7. Add 6 mL of Protein precipitation solution, vortex and incubate on ice for 15 min.
8. Transfer into 30-mL centrifuge tube, centrifuge at 18,000×*g* for 10 min.
9. Transfer the supernatant into a new 50-mL tube, add 14 mL of isopropanol and invert the tube several times to mix.
10. Spool the DNA precipitant with a pipet tip (wind the DNA around the tip), then transfer into 30 mL of 70% ethanol, mix gently.
11. Spool the DNA again and transfer into a 2-mL tube.
12. Air dry and dissolve the DNA in 1.5 mL of Rehydration solution, determine the quantity and quality with spectrophotometer and store in the refrigerator.

3.1.2. Partial Digest of Genomic DNA

1. Set up *Sau*3AI partial digest reaction as follow:

Genomic DNA:	2 μg
10× *Sau*3AI buffer:	10 μL
100× BSA:	1 μL
*Sau*3AI:	0.3 U
Nuclease-free water:	to make 100 μL

2. Incubate at 37°C for 60 min.
3. Electrophoresis on 1% agarose gel without ethidium bromide.
4. Stain the agarose gel in 100 mL of water with 10 μL of SYBR Gold Nucleic Acid gel stain at room temperature for 10 min.
5. Place the stained gel on the Dark Reader Transilluminator, isolate the gel slice containing DNA fragments with 2–4 kb size.
6. Recover the DNA with QIAquick gel extraction kit described as follows.
7. Transfer the gel slices into a 15-mL tube, measure weight and add threefold volumes of buffer QG, incubate at 50°C water bath for 10 min, invert the tube occasionally.
8. Transfer 750 μL into the column, centrifuge for 1 min at 12,000 rpm in microcentrifuge, and discard the flow-through.
9. Repeat the step 8 until all of the solution were loaded.
10. Add 750 μL of buffer PE into the column, centrifuge for 1 min at 12,000 rpm, and discard the flow-through.
11. Centrifuge 2 min at 12,000 rpm; transfer the column into a new 1.7-mL microfuge tube.
12. Add 100 μL of buffer EB into the column, centrifuge for 1 min at 12,000 rpm, and collect the DNA.

3.1.3. Partial Fill-In of Genomic DNA Fragments Using dGTP and dATP

1. Set up reaction as follow:

*Sau*3A digested DNA fragments:	85 μL
10× NE Buffer 2:	10 μL
dGTP/dATP:	40 μM each
DNA polymerase I, Large (Klenow):	10 U
Nuclease-free water:	to make 100 μL

2. Incubate at 25°C for 15 min, then in 75°C for 20 min to deactivate the enzyme.
3. Purify the DNA with ZYMO DNA clean and concentrator as follows.
4. Add fivefold volumes of DNA binding buffer to DNA samples, mix briefly.
5. Transfer the mix to a Zymo-Spin column, centrifuge at 12,000 rpm in microcentrifuge for 1 min, discard the flow-through.
6. Add 200 μL of washing buffer to the column, centrifuge at 12,000 rpm for 1 min, and discard the flow-through.
7. Repeat the above wash step again.
8. Add 10 μL of nuclease-free water to the column; transfer the column to a new 1.7-mL microfuge tube, centrifuge at 12,000 rpm for 1 min, and collect the pure DNA.

3.1.4. Digest Plasmid pDCQ601

1. Set up digest reaction as follow:

pDCQ601:	2 μg
10× NEBuffer 3:	5 μL
100× BSA:	0.5 μL
SalI:	10 U
Nuclease-free water:	to make 50 μL

2. Incubate at 37°C for 2 h.
3. Purify the vector DNA with ZYMO DNA clean and concentrator, as described above in Subheading 3.1.3, elute in 30 μL nuclease-free water.

3.1.5. Partial Fill-In of the Linearized pDCQ601 Using dCTP and dTTP

1. Set up reaction as follow:

Linearized pDCQ601:	25 μL
10× NEBuffer 2:	10 μL
dCTP/dTTP:	40 μM each
DNA polymerase I, Large (Klenow):	10 U
Nuclease-free water:	to make 100 μL

2. Incubate at 25°C for 15 min, then in 75°C for 20 min to deactivate the enzyme.
3. Purify the DNA with ZYMO DNA clean and concentrator, as described above in Subheading 3.1.3.

3.1.6. Ligation

1. Set up a ligation reaction as follow:

pDCQ601 vector (prepared in Subheading 3.1.5):	1 μL
MG1655 fragments (prepared in Subheading 3.1.3):	1 μL
10× ligation buffer:	1 μL
Ligase:	1 μL
Nuclease-free water:	6 μL

2. And a control ligation reaction using only the vector DNA as follow:

pDCQ601 vector (prepared in Subheading 3.1.5):	1 μL
10× ligation buffer:	1 μL
Ligase:	1 μL
Nuclease-free water:	7 μL

3. Incubate overnight at 16°C.

3.1.7. Transformation into Peptide Production Strain QC1101

The peptide production strain QC1101 was used as the host for the genomic library. The strain QC1101 was derived from *E. coli* strain MG1655 with a knockout of the *slyD* gene and *araBAD* genes in the chromosome and contained a pBAD-based peptide expression plasmid pLR199.

The knockout of the *slyD* gene was engineered to reduce the nonspecific background of Lumio™-based in-cell labeling. The knockout of *araBAD* genes was engineered to increase efficiency of L-arabinose induction. The peptide production plasmid pLR199 expresses a fusion peptide containing a small inclusion body tag IBT139 and a tetracysteine motif CCPGCC followed by a peptide of interest HC124. The fusion peptide contained 6.22% of tryptophan and 10.36% proline, which are several folds higher than the respective compositions for tryptophan or proline in average *E. coli* host proteins.

1. Streak *E. coli* QC1101 strain onto LB containing 100 μg/mL ampicillin (Amp100) plate, incubate overnight at 37°C.
2. Inoculate a single colony into a culture tube with 3 mL of LB/Amp100; incubate overnight at 37°C with shaking at 250 rpm.
3. Inoculate 1 mL overnight culture into a 500-mL flask containing 100 mL fresh LB/Amp100 medium, incubate at 37°C with shaking at 250 rpm for 3–5 h until the OD_{600} reach around 0.6.
4. Transfer the cultures into two 50-mL tubes, chill on ice for 10 min, then centrifuge at 4°C with 6,000 × *g* for 10 min, and disregard the supernatant.
5. Resuspend the cell pellets with 50 mL of prechilled water, leave on ice for 10 min, then centrifuge at 4°C with 6,000 × *g* for 10 min, and discard the supernatant.
6. Repeat step 5 with 25 mL of prechilled water.
7. Repeat step 5 with 5 mL of prechilled 10% glycerol and combine together.
8. Resuspend the cell pellets with 1 mL of prechilled 10% glycerol; make 100 μL aliquots into prechilled 1.7-mL microfuge tubes.
9. Flash freeze the competent cells in liquid nitrogen and store at −80°C.
10. Thaw a tube of QC1101 competent cells on ice and prechill two 1.7-mL microfuge tubes.
11. Transfer 30 μL competent cells into each of the prechilled 1.7-mL tubes.
12. Add each of 2 μL DNA ligation reaction mixture and control mixture from Subheading 3.1.6 into the competent cells separately, tap the tubes to mix.
13. Transfer each of the mixture of DNA and cells into the prechilled 1-mm eletroporation cuvettes, respectively.
14. Set the electroporator at voltage = 1.8 kV, resistance = 200 ohms (Ω) and Capacitance = 25 microFaradays (μF).
15. Electroporate the cells.
16. Immediately after electroporation, add 1 mL SOC medium into the electroporation cuvette, and mix by pipetting.

17. Transfer cells into a culture tube; incubate at 37°C with shaking at 250 rpm for 45 min.
18. Plate the entire library transformation onto LB/Amp100, Kan50 (100 μg/mL ampicillin and 50 μg/mL kanamycin) plates, 50 μL per plate; Plate 100 μL of control transformation onto one LB/Amp100, Kan50 plate.
19. Incubate the plates overnight at 37°C.

3.1.8. Check Quality of the Library

The quality of a library is dependent on two factors: titer and background. Higher titer and lower background means better quality.

$$\text{Titer} = \text{Avg. colonies / plate} \times 1\,\text{mL} / 50\,\mu\text{L} = \text{CFU / plate} \times 20,$$

$$\begin{aligned}\text{Background} &= \text{Avg. colonies / control plate} \times 1\ \text{mL} / 100\ \mu\text{L} \\ &= \text{CFU / control plate} \times 10.\end{aligned}$$

By counting the number of the colonies on plates, the above library, named as QC1300, had a titer of 2×10^4 CFU with around 2% background. The randomness of the library was also checked by sequencing the inserts on the plasmid in 96 colonies. Pool all the colonies in LB/Amp100, Kan50 medium containing 20% glycerol, then store in –80°C freezer.

3.2. Lumio In-Cell Labeling

The library is grown and the induced peptide is labeled in vivo using the TC-FlAsH II In-Cell Tetracycsteine Tag detection kit as follows.

1. Inoculate 20 mL LB/Amp100, Kan50 with 400 μL of library frozen stock, and incubate at 37°C with 250 rpm shaking. Monitor the cell growth at OD_{600}.
2. The cell cultures are induced with 200 μL of 20% L-arabinose (0.2% final concentration) when the cultures reach an OD_{600} of ~0.6 (see Note 1). The cultures are then grown for additional 3 h at 37°C with 250 rpm shaking.
3. The OD_{600} of the culture is determined after the growth and the cells are diluted to $OD_{600} = 1$ in LB/Amp100, Kan50 medium.
4. The labeling mix contains: 144.5 μL of LB media, 0.25 μL of 100 mg/mL Amp, 0.25 μL 50 mg/mL Kan, 2.5 μL 10 μM TCEP, 2.5 μL 2 mM Lumio reagent.
5. Mix 100 μL of the OD_{600} ~1.0 cells with 150 μL of the label mix and incubate at room temperature protected from light for 2 h.
6. The cells are pelleted by microcentrifugation for 5 min at 11,000 rpm; and the supernatant is carefully removed.
7. The cells are washed with 500 μL BAL wash buffer and pelleted by microcentrifugation as in step 6.
8. The cell pellet is resuspended in 300 μL of PBS and is ready for FACS.

3.3. FACS Sorting

The principles of FACS will not be covered in this chapter and you may want to seek further information on flow cytometry before proceeding. The general idea is to sort a diversey library and enrich for a population with a desired phenotype based on fluorescence or light scattering properties. The enrichment is performed by multiple rounds of sorting and regrowth. Due to differences in biological diversity of the libraries, the efficiency of the labeling methods, and the sensitivity and variability of the instrument, the optimal sorting conditions and enrichment cycles need to be determined experimentally. QC1300 library was sorted on the FACS with increased stringency by four rounds of growth and enrichment of the cell population that have higher fluorescence signal from the Lumio labeling of the target peptide. Increase of the mean of the Gaussian population as observed in the FITC channel after the four rounds of sorting is an indication of successful enrichment of cells of higher fluorescence (see Fig. 1).

3.3.1. Set-Up of FACS Instrument

1. Power on all required instrumentation for the Fluorescence-Activated Cell Sorter.
2. The flow cytometer used was a FACSVantage™ SE-DiVa and set-up as detailed (see Note 2).
3. The general alignment was performed using ALIGNFLOW488™ 2.5 μm diameter fluorescent beads. Place three drops of the ALIGNFLOW488™ beads into a BD Falcon Tube and add 1.5 mL of FACSFlow™ sheath fluid to the tube, vortex and place on the FACS system.

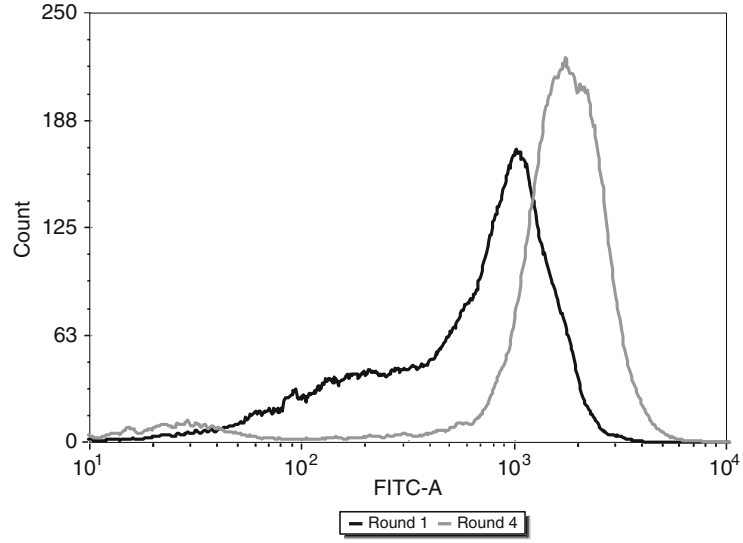


Fig. 1. FITC (Fluorescein isothiocyanate) signal distribution of the library at first and fourth rounds of sorting. After the fourth round, the FITC was significantly improved from that of the first round.

4. Adjust the laser power on the 488 nm argon ion laser to 200 mW or maximum.
5. Adjust the system nozzle and stream adjustments using the X, Y, Z, α, theta, focus lens channel height, and channel height focus to maximize the signals in all channels while minimizing the coefficient of variation (CV) for each channel.
6. Save the instrument settings within the DiVa software.
7. Remove the ALIGNFLOW488™ beads from the FACS, cap and set aside while the system is flushing.
8. Obtain another BD Falcon Tube and place one drop of ACCUDROPS™ and add 1.5 mL of FACSFlow™ sheath fluid and vortex well.
9. Place this sample tube onto the FACS instrument. Energize the high voltage and adjust the differential pressure to move the beads through with an approximate rate of 500 events/s, adjust the FACS stream separation as necessary.
10. Adjust the drop delay and amplitude to enable the best "visually" stable sort spot on the FACS drop stream monitor.
11. Once the adjustments are complete, save the settings and remove the ACCUDROPS™ sample and allow the system to flush to waste for a few minutes.
12. After completing the alignment and sort adjustments, place 1–2 mL of 20% methanol into a BD Falcon Tube onto the FACS system and allow it to flush approximately 1 mL through the sample tubing and nozzle.
13. Remove the 20% methanol tube and replace it with a BD Falcon Tube filled with FACSFlow™ sheath fluid and allow ~1 mL to pass through the system.
14. The FACS is now ready for the Lumio™ samples.
15. Leave the current BD Falcon tube in-place that contains the FACSFlow™ sheath fluid. Obtain a sterile 1.7-mL Microfuge Tube and place it briefly under the nozzle/stream to capture ~1 mL of sheath fluid; this can be used to determine system sterility (see Note 3).

3.3.2. Initial Sample Screens

1. With the FACS aligned and the sample input line clean; set up a new experiment and create a few dot-plots and histograms for the samples to be screened. Consult your FACS data acquisition software manuals for more information if needed.
2. The initial FACS "screening" (see Note 4) of the samples is necessary to confirm the systems operation, labeling efficiency, and sort gates.
3. Prepare the samples to be screened by passing 100 μL of OD_{600} ~ 1.0 through a 35-μm mesh Blue Cap Falcon Tube (see Note 5).

4. Add 100 μL of FACSFlow™ sheath fluid or 2× PBS to dilute the sample and vortex.
5. Place a sample onto the FACS and increase the differential pressure until cells are seen at a rate of approximately 1,000 events/s and uniformly delivered. If the differential pressure must be increased to an excessive amount; please check the threshold (see Note 6) setting or add more cells to the sample.
6. When the sample input rate is approximately 1,000 events/s and appears as an approximate Gaussian distribution on the FSC vs. SSC dot plot within the channel histograms, begin recording 10,000–20,000 events (see Note 7).
7. Once the sample acquisition recording is complete, remove the sample tube and allow the sample line to flush to waste for a few moments.
8. Prepare the next sample to be screened and place it on the FACS and repeat steps 3–7 until all the samples have been screened. Remember to save your data as you proceed.
9. Once the final sample has been screened and removed from the FACS, allow the sample line to drain to waste for several minutes while reviewing the sample screen plots.
10. If a sterile sort will be performed, follow these instructions (see Note 8) to have a clean system before beginning the sort. Caution must also be exercised to prevent any cross-contamination from other samples (see Note 9).

3.3.3. Sort Gate Selection

1. After screening each of the Lumio™ labeled samples, prepare the cells to be sorted by aliquoting 500 μL of labeled Lumio™ cells at OD_{600} ~1 through a 35-μm mesh Blue Cap Falcon Tube, then dilute with approximately 250–500 μL of 2× PBS or FACSFlow™ sheath fluid.
2. Place the sample to be sorted onto the FACS and increase the differential pressure until the cells are moving through the tip at approximately 1,000 events/s.
3. Record 20,000 events for the sample.
4. Create a histogram, if not already present, to observe the Lumio™ labeled cells in the FITC histogram (Em: 530 nm).
5. Select a gate tool within the software and draw a gate to enclose the "Top X %" of the desired cells to be sorted (see Note 10).

3.3.4. Sorting of Lumio™ Labeled Samples

1. Aliquot 200 μL of OD_{600} ~1 cells labeled with Lumio™ and pass it through a 35-μm mesh Blue Cap Falcon Tube followed by approximately 200 μL of 2× PBS or FACSFlow™ sheath fluid.
2. Place the Lumio™ labeled sample on to the sample input holder.

3. Place a capture tube, 1.7-mL microfuge tube, containing a 100 µL of LB into the sorting tube holder or 5-mL culture tube if a very large number of events will be sorted.
4. Increase the differential pressure to approximately 1,000 events/s.
5. Record 20,000 events for the sample.
6. Using the FACS DiVA software, draw a sort gate on the FITC histogram to capture the "Top _%" for the sort as designated in the FACS sort table below.

FACS sort and enrichment table

Sort #	FITC sort gate (Top %)	Events sorted
1	10	100,000
2	5	50,000
3	1	50,000
4	0.50	10,000

7. Enter the number of cells from the selected population to be sorted, for example 100,000 events, into the sort field and begin sorting the cells on the FACS.
8. Vortex the sample gently during the sort and maintain a sort rate of approximately 2,000 events/s during the sort (see Note 11).
9. Set and adjust the sort stream so the sorted cells are deposited into the capture that contains 100 µL of LB medium.
10. After the desired number of events is captured, remove the collection tube, cap, and gently vortex and set it aside, multiple 1.7-mL microfuge tubes can be collected and then combined if necessary.
11. Remove the sort sample and allow the sample line to drain to waste followed by cleaning or disinfecting solution rinse of the sample tubing.
12. Spread the samples on agar plates or place in liquid media for growth (see Note 12).

3.3.5. Recover Cells from Sorting

1. For first three round of sorting, the sorted material from the FACS is plated out on LB/Kan50, Amp100 plates, and grown at 37°C overnight.
2. The colonies are counted on the plates. Typically, the colony counts should be approximately 10–20% of the initial sorted events obtained from the FACS.

3. The colonies are harvested from the plates with 2 mL LB/Amp100, Kan50 medium per plate, and combined into one tube.
4. The materials are then adjusted to OD_{600} of 4.0 and used as seed for growth for the next round. Aliquots of cells harvested from each sorts are also frozen at −80°C.
5. After regrowth, the samples are labeled again with Lumio™ as described in Subheading 3.2 and sorted again as in Subheading 3.3.
6. For the last round of sorting (see Note 13), the sorted material from the FACS is plated out onto LB/Amp100, Kan50 plates, and grown overnight at 37°C.
7. Isolated colonies are picked into 96-well plates with 100 μL LB/Amp100, Kan50 in each well, and grown at 37°C with 800 rpm shaking overnight.
8. The grown colonies are then diluted into Molecular Grade Water for sequencing. Frozen stocks are also made by adding 50 μL of 50% Glycerol to each well, and the plates are stored at −80°C.

3.4. Sequencing of the Isolates from the Last Sorting

The diluted cultures are submitted for sequencing with the primers listed below.

pBHR1F: GCGA TGAAAACGTT TCAGTTT
pBHR1R: CACAAGTTTTATCCGGCCTTT

The sequence results identify the genes included on the DNA fragment of the library clone, as well as orientation of the fragment on the plasmid. Multiple isolates were obtained for the fragments shown on Fig. 2 from the sorting enrichment of QC1300 library.

3.5. Confirmation of Potential Hits for Increased Fluorescence

To verify that the phenotype of the increased fluorescence is due to the expression plasmids in the identified clones, the plasmids are isolated from the identified strains and retransformed into the fresh host; and the retransformed strains are then tested to see if they show the same phenotype as the original isolates (see Note 14).

3.5.1. Retransformation of Plasmids into Fresh Host

1. Overnight cultures of isolates are grown in 3 mL LB/Amp100, Kan50.
2. Plasmid DNA is isolated from the strains using Qiaprep spin Miniprep Kit, as briefed below.
3. Centrifuge the cultures at 14,000 rpm for 3 min in microcentrifuge, discard the supernatant.
4. Resuspend each of the pellets with 250 μL P1 buffer, then transfer into 1.7-mL microfuge tube.
5. Add 250 μL P2 buffer, invert 6–8 times to mix.
6. Add 350 μL N3 buffer, invert 6–8 times to mix.

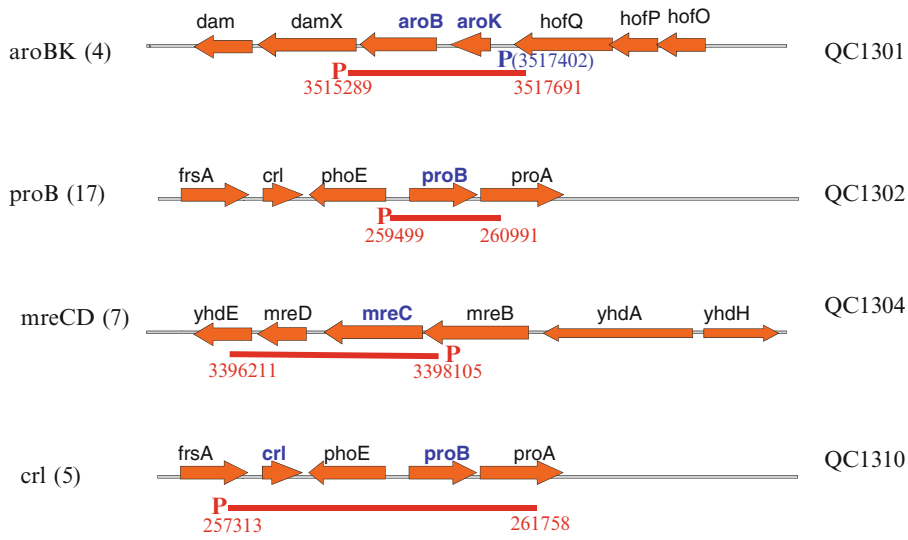


Fig. 2. DNA fragments expressed in higher fluorescent clones selected from FACS sorting. In the *aro*KB clone, the native promoter for *aro*KB (at position 3517402) was in the opposite orientation as the promoter of Cm^r gene on the vector. In the other three cases, the genes on the fragments were expressed in the same orientation as the vector promoter. The numbers in parenthesis showed how many isolates obtained.

7. Centrifuge 10 min at 14,000 rpm in microcentrifuge.
8. Transfer the supernatant into Miniprep column, and then spin for 1 min at 14,000 rpm.
9. Discard the flow-through, add 750 µL PE buffer, and spin for 1 min at 14,000 rpm.
10. Discard the flow-through, repeat 1 min centrifuge at 14,000 rpm.
11. Transfer the column to a new 1.7-mL microfuge tube, add 50 µL EB buffer, then spin for 1 min at 14,000 rpm, and collect the plasmid DNA.
12. The plasmids are then retransformed into the host strain following the same protocol described in the subheading 3.1.7.

3.5.2. FACS Analysis of the Individual Strains

1. Inoculate 3 mL LB/Amp100, Kan50 with loop of frozen stock and grow overnight at 37°C with 250 rpm shaking.
2. In the morning make a 1:100 dilution of the overnight culture into fresh 20 mL LB/Amp100, Kan50, monitor the cell growth at OD600.
3. The cells are grown and labeled with Lumio™ as described in Subheading 3.2.
4. The labeled cells are analyzed on FACS (without sorting) as in Subheadings 3.3.1 and 3.3.2. The FITC signal as well as the FSC and SSC signals of the retransformed strains is recorded and plotted comparing to the control (see Fig. 3).

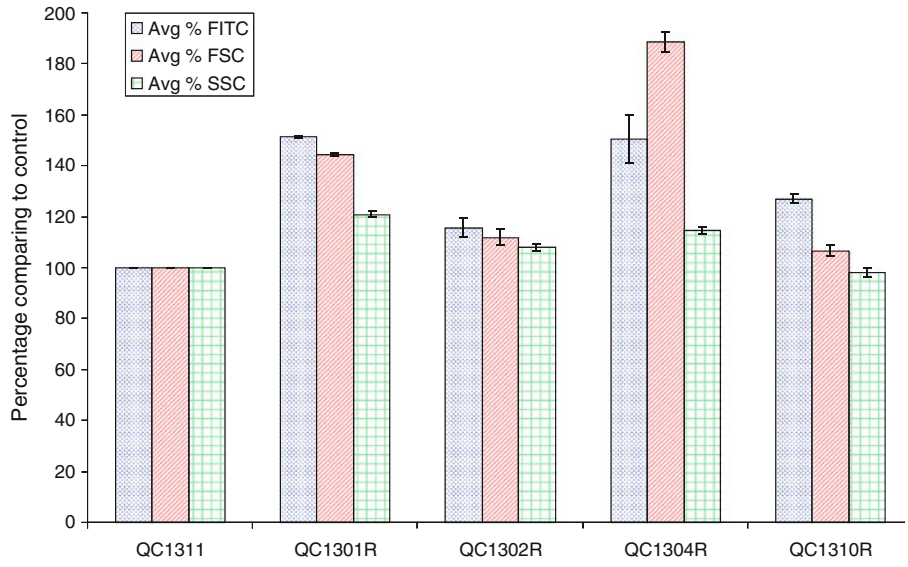


Fig. 3. FACS analysis of *E. coli* hosts containing the isolated expression plasmids. The *error bars* represent standard deviations of three independent cultures. FITC refers to fluorescence signal. FSC refers to forward scattering, which usually correlates with cell size. SSC refers to side scattering, which usually correlates with intracellular granularity.

5. Aliquots of the same labeled cells can also be used to assay for peptide production as described below in Subheading 3.6.

3.6. Assay for Peptide Production

To determine if the increase of fluorescence correlates with increase of peptide production, peptide amount was analyzed on SDS-PAGE and visualized by Lumio™ in-gel labeling and Simply Blue™ staining.

3.6.1. In-Gel Labeling Reaction

1. 100 µL of cells normalized for the same OD are pelleted for 2 min at 13,000 rpm. The supernatants are removed and the pelleted cells may be stored at −80°C or used directly.
2. Add 20 µL 10× CelLytic (see Note 15) to each sample and vortex to mix, incubate at 37°C for at least 30 min, vortex occasionally.
3. Briefly pulse in a microfuge to get all of the liquid condensation down into the tube. If some particulates are visibly floating in lysate, add 350 µL of 8 M urea to solubilize particulates.
4. Make master mix of 2× sample buffer and lumio reagent from the Lumio™ in-gel labeling kit. For 40 reactions add 8 µL of lumio reagent to 400 µL of 2× sample buffer (see Note 16).
5. Add 10.2 µL of the master mix to each new labeled tube.
6. Add 10 µL of lysed sample to each tube of the labeling mix, incubate at 70°C for 10 min.

7. Cool down for 1–2 min and spin briefly to get all condensation into the tube.
8. Add 2 μL of Lumio™ In-Gel detection enhancer. Mix carefully by pipetting and incubate at room temperature for 5 min.

3.6.2. SDS-PAGE

1. Load 10 μL of each sample onto NuPage 4–12% Bis-Tris gel (see Note 17).
2. Run with MES running buffer, at 200 V. When the gel is done, remove it from the gel case and rinse in ddH_2O briefly.
3. Take a picture of the gel using fluorescence setting for ethidium bromide or FITC. Use the best camera available and save the image as a tif file.
4. Stain the PAGE gel with Simply Blue™ stain.
5. Place the gel in ddH_2O, covering it in the microwave and heat on High for 45 s; incubate the gel for 2 min with shaking; pour off the water and replace it with fresh water (see Note 18).
6. Repeat step 5 twice more, do not discard water after third incubation on the shaker.
7. Add about 40 mL of Simply blue™ stain to the container and heat for 45 s in the microwave.
8. Incubate with shaking for 10 min.
9. Remove stain into waste collection. It is possible to see bands at this point.
10. Fill the container with more ddH_2O to destain the gel on the shaker. Depending on the amount of background staining, it might take several hours to overnight to have the gel thoroughly destained. Change the ddH_2O occasionally.

3.6.3. Image Analysis

The data from the image analysis gives a semiquantitative value to the amount of peptide produced in the cells. In the example shown here (see Fig. 4), several strains (QC1301R, QC1302R, and QC1310R) showed more peptide production than the control strain (QC1311).

1. Analyze both the fluorescence image from Lumio™ in gel labeling and the staining image from the Simply Blue™ stain using the Image J program (http://rsbweb.nih.gov/ij/index.html).
2. Open the image file with Image J. Click on the Process tab and go down to "Subtract Background." For the fluorescent gels or Simple blue™ gels, set the rolling ball value to 25 or 12 pixels respectively. Check the light background box for the simply blue gels.
3. Click okay to normalize the background of the picture.

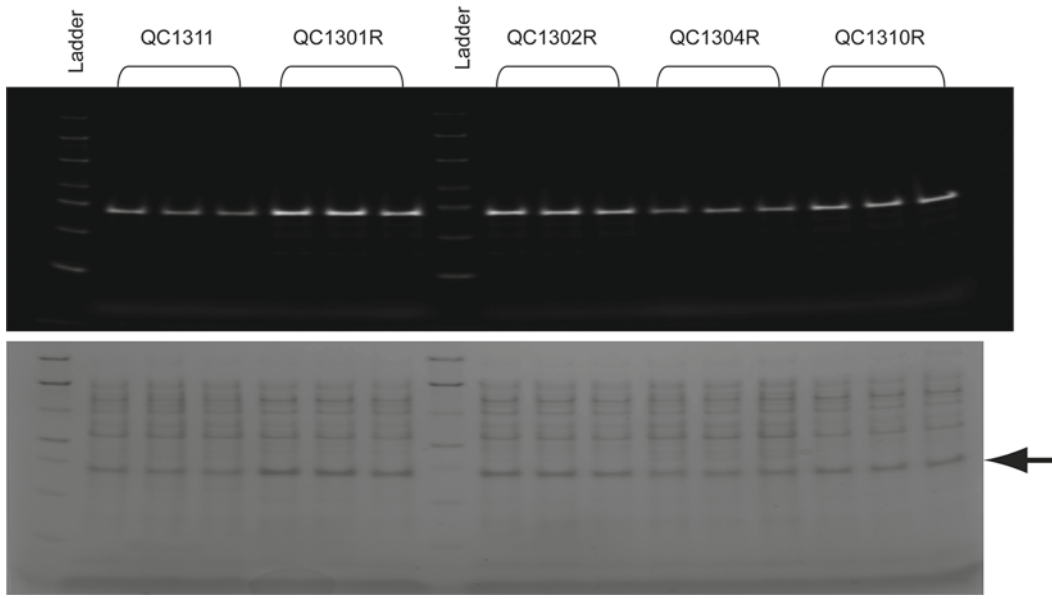


Fig. 4. Gel pictures of the triplicate cultures of the retransformed strains. The final OD_{600} of the cultures were similar. *Top panel* showed the specific in-gel labeling of the fusion peptide viewed under the UV light. The same gel stained with Simply Blue was shown on the *bottom panel*. The *arrow* indicates the fusion peptide.

4. Choose desired measurements such as area, standard deviation, Min and Max Gray values, integrated density, and mean gray value under Set Measurements. Integrated density values were used in this example for relative quantitation.
5. To do the measurements, move the box over the band and either use "ctrl+M" or click on the measurement choice under Analyze.
6. Save the data and complete the analysis with Microsoft Excel software.

4. Notes

1. The OD_{600} of induction is specific to the media used and the bacteria. It may be necessary to test at different OD_{600} for induction to determine what will give the best labeling.
2. The FACS instrument used in this experiment is configured with a single, power adjustable, solid-state 200 mW, 488 nm argon ion laser (Coherent, Santa Clara, CA). The bacterial cell size is in the micron range and would require the PMT detectors (FSC, SSC, FITC) to be set around 200–400 V with the laser set to 200 mW. The side scatter detectors are photomultiplier

tubes (PMT) arranged in an octagon configuration. This experiment uses the side scatter channel [SSC] with a 488 ± 10 nm filter and the fluorescein isothiocyanate [FITC] channel using a 530 ± 15 nm filter. The forward scatter channel [FSC] detector uses a photo-diode detector. The sheath fluid used is FACSFlow™ with an operating system pressure of 28 psi (~193 kPa) and a ceramic 70 μm diameter orifice nozzle.

3. Plate 50–100 μL of the system sheath fluid onto a LB plate and incubate at 37°C for 1–2 days. If any colony growth is observed on the plate, then the system could be potentially contaminated and the sorted samples would need to be disposed. The experiment needs to be repeated.
4. Usually, blank cells are screened first followed by a "positively" stained sample. The FACS system requires that both the laser power and PMT be adjusted at the onset of every experiment in order to place the blanks at or near the bottom in all the channels required. Place the histogram of the stained sample around the middle of the *x*-axis (approximately 10^3 log scale) while keeping the unstained sample ("blank") as low as possible to minimize the background fluorescence. Adjust the laser power and PMT voltages for the FSC, SSC, FITC channels to provide a strong signal for the positive stained sample while minimizing the background fluorescence for the blank sample. Also, the alignment beads can be used to verify or make fine adjustments to the PMT settings if necessary.
5. Using the "Blue Cap" 35-μm Mesh Falcon Tubes has enhanced the ability to quickly and easily prevent samples from plugging the FACS tip and to provide a more uniform sample delivery to the FACS tip.
6. The threshold setting, typically that of the FSC, may need to be adjusted so that real "events" are being detected at a reasonable rate, but not too sensitive that it is erroneously detecting events. Please see your FACS operator's manual for more information if needed.
7. The term "event," as applied in flow cytometry, is used to define a cell; but the event could also be a piece of a cell, double-cells, large foreign debris, or other non-cell entities that may be interpreted as a real cell or event on the instrument. In flow cytometry, a "screen" means to quickly capture a number of events (e.g., 5,000–20,000 events) in order to obtain a view of the population profile. A sample screen on the FACS is necessary for all the blank cells, labeled cells, and alignment beads.
8. The sterility of the FACS system is very important when sorting cells for regrowth and library enrichment. Use of all sterile materials is highly suggested. The instrument sorting chamber and work surfaces can be disinfected and are typically wiped

down at the end of each sorting with a surface disinfectant (RelyOn, Dupont, Wilmington, DE). The sample tubing is flushed using a series of 70% isopropyl alcohol, 20% methanol, and fresh sheath fluid at the end of every experiment. Becton Dickinson implemented an aerosol management option that uses a Whisper (Buffalo Filter, Buffalo, NY) HEPA filtration device connected to the sorting chamber to eliminate any aerosols and prevent environmental air-borne contaminates outside of the sorting chamber.

9. If FACS is done in a core facility, perform library sorting first before other samples and perform only one sort for that day. Other samples could be screened on the FACS but no samples were ever placed onto the system *before* the library to be sorted to prevent any cross contamination.
10. In flow cytometry, the term "gate" refers to one or more user-defined areas that are selected from the sample population which can then be displayed on the plots using the flow cytometry software. The shape and position of the stained-library sample, as displayed in the FITC histogram window of the DiVa Software, enables a capture gate or sort gate to be used to acquire or sort the desired events from the stained cell population into a tube placed on the system. Often, a gate will be used to isolate or provide information to the experimenter on cells or a population of interest. The main function of a sort or collection gate is to encompass the cells of interest, and then enable the FACS hardware to separate or "sort" them out of the population; it is this capability that makes the sorting flow cytometer a very powerful biotechnology tool.
11. Some flow cytometers can sort cells as fast as 60,000 events/s, but this can cause the core stream to increase and can result in decreasing the efficiency and "purity" of the sort. Slowing down the sorting speed has generally enhanced the overall sorting efficiency.
12. For library sorting, regrowth on plates is recommended as this introduces less bias toward enriching the faster growing strains. Regrowth should be carried out as soon as possible.
13. It may take several rounds of FACS sorting to see the increase of FITC signal. The point at which the FITC signal of the sorted fraction does not appear to increase further is the point which the library is fully enriched for the phenotype initially selected.
14. The strains are retransformed and retested on the FACS to verify that the increase in FITC signal is due to the genes expressed on the plasmids, not some other unknown chromosomal mutations in the original isolates that might have occurred during repeated rounds of sorting.

15. CelLytic enzyme is stored as a10× stock. Mix the powder with Molecular Grade Water, vortex well to make sure all of the powder is in solution. Aliquot into single use amounts and store at −20°C. The amount of CelLytic enzyme used here is optimal for less than 6.0 OD × mL cell cultures.
16. To make a working solution of sample buffer, thaw a tube of 4× sample buffer and add 200 µL (equal volume) of Molecular Grade Water. Heat the tube for a few minutes at 70°C to get everything into solution. Vortex to make sure everything is homogenous and pulse in a microfuge to get all the liquid down in the tube.
17. Triplicate samples should be loaded for better quantitation. If there is concern about overloading, serial titrations of each sample, e.g., 10, 5, and 2.5 µL, could be loaded.
18. The advantages of the microwave staining method are faster staining time and more consistency between gels.

References

1. Herzenberg LA, Parks D, Sahaf B et al (2002) The history and future of the fluorescence activated cell sorter and flow cytometry: A view from Stanford. Clinical Chemistry 48:1819–27.
2. Tracy BP, Gaida SM, Papoutsakis ET (2010) Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. Curr Opinion in Biotechnol 21:85–99.
3. Mattanovich D, Borth N (2006) Applications of cell sorting in biotechnology. Review. Microb Cell Fact 5:12–22.
4. Nebe-von-Caron G, Stephens PJ, Hewitt CJ et al (2000) Analysis of bacterial function by multicolour fluorescence flow cytometry and single cell sorting. J Microbiol Methods 42:97–114.
5. Johnson IS (1983) Human insulin from recombinant DNA technology. Science 219:632–7.
6. Brassard DL, Grace MJ, Bordens RW (2002) Interferon-alpha as an immunotherapeutic protein. J Leukoc Biol 71:565–81.
7. Stübgen JP (2009) Recombinant interferon-beta therapy and neuromuscular disorders. J Neuroimmunol 212:132–41.
8. Ng T, Marx G, Littlewood T, Macdougall I (2003) Recombinant erythropoietin in clinical practice. Postgrad Med J 79:367–76.
9. Maurer KH (2004) Detergent proteases. Curr Opin Biotechnol 15:330–4.
10. Misawa S, Kumagai I (1999) Refolding of therapeutic proteins in *Escherichia coli* as inclusion bodies. Biopolymers 51:297–307.
11. Fischer B, Sumner I, Goodenough P (1993) Isolation, renaturation, and formation of disulfide bonds of enkaryotic proteins expressed in *Escherichia coli* as inclusion bodies. Biotechnol Bioeng 41:3–13.
12. Griffin BA, Adams SR, Jones J, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. Science 281: 269–72.
13. Adams SR, Campbell RE et al (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. J AM Chem Soc 124:6063–76.
14. Berry A (1996) Improving production of aromatic compounds in Escherichia coli by metabolic engineering. Trends Biotechnol 14: 250–6.
15. Hayzer DJ, Leisinger T (1980) The gene-enzyme relationships of proline biosynthesis in *Escherichia coli*. J Gen Microbiol 118: 287–93.

Chapter 13

Using Flux Balance Analysis to Guide Microbial Metabolic Engineering

Kathleen A. Curran, Nathan C. Crook, and Hal S. Alper

Abstract

Metabolic engineers modify biological systems through the use of modern molecular biology tools in order to obtain desired phenotypes. However, due to the extreme complexity and interconnectedness of metabolism in all organisms, it is often difficult to a priori predict which changes will yield the optimal results. Flux balance analysis (FBA) is a mathematical approach that uses a genomic-scale metabolic network models to afford in silico prediction and optimization of metabolic changes. In particular, a genome-scale approach can help select gene targets for knockout and overexpression. This approach can be used to help expedite the strain engineering process. Here, we give an introduction to the use of FBA and provide details for its implementation in a microbial metabolic engineering context.

Key words: Flux balance analysis, MOMA, Genomic-scale metabolic model, Metabolic engineering, Strain engineering

1. Introduction

The goal of metabolic engineering is to confer useful properties to a biological system, such as the ability to produce high levels of a biochemical product. In order to do so, it is often necessary to reroute cellular metabolites and cofactors toward the pathway of interest. In this regard, advances in modern molecular biology and genetics have empowered metabolic engineers with an increasingly large toolkit for rewiring metabolic networks. This task is accomplished primarily through the use of gene knockouts, overexpression, and expression of heterologous genes. It is often difficult to identify the most important bottlenecks in a metabolic network a priori due to the complexity and interconnectedness of global cellular metabolism. Moreover, precursor metabolites and cofactors

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_13,

participate in a number of reactions across the cell and it is therefore challenging to identify gene targets by simply looking at a metabolic map. Often, mutations distal from the desired pathway can still exert significant influence. Thus, a robust, global method is required to predict gene targets for improving metabolic pathways.

One approach to systematically identify targets for gene knockouts and overexpressions is through the use of global models of metabolism. Genome-scale modeling enables in silico (on the computer) reconstruction of organism. Using this approach, metabolism and the impact of genetic alterations can be predicted. Genome-scale modeling in its simplest form is a constraint-based approach that relies on the simple fundamental assumption that reactions must be balanced on a mass and energy basis. In the ultimate model of metabolism, kinetic and regulatory control will be incorporated. However, an incomplete dataset exists for both of these important factors and thus only reaction stoichiometry is usually considered (see Note 1). Thus, in its typical form, genome-scale metabolic models allow the user to model all of the metabolic reactions occurring in an organism in silico using only reaction stoichiometry as a required input (although more sophisticated cellular knowledge can result in better, more realistic models). These metabolic models can then be used to predict changes in the metabolic system and thus allow for a rational way to choose gene targets for manipulation and reduce the need for costly large-scale or high-throughput experiments in vivo.

1.1. Fundamental Framework for Flux Balance Analysis

Flux balance analysis (FBA) is the primary method of predicting a metabolic phenotype from a genomic-scale metabolic model. This method is formed around the basic premise that metabolites and reaction fluxes must obey the laws of conservation of mass. FBA takes a genome-scale metabolic model (typically in the form of an under-defined system of linear equations) as its input and uses an objective function to solve for the optimal solution. The output is a potential solution for the flux through every metabolic reaction in the model that satisfies the material balances and maximizes or minimizes the specified objective function. Perturbations to the model can be made (to simulate gene knockouts, overexpression, heterologous gene integration or adjusted environmental conditions) and FBA can be used to search for the optimal conditions for inducing a desired phenotypic change to an organism. Such perturbations can be used to predict gene knockout targets, overexpression targets, flux distributions through critical nodes, or the influence of changed environmental conditions.

The fundamental method of FBA can be broken down into a series of six basic steps (see Fig. 1).

- First, a metabolic model is constructed by defining all of the stoichiometric reactions that can occur within the organism utilizing annotated genomic sequence and literature data.

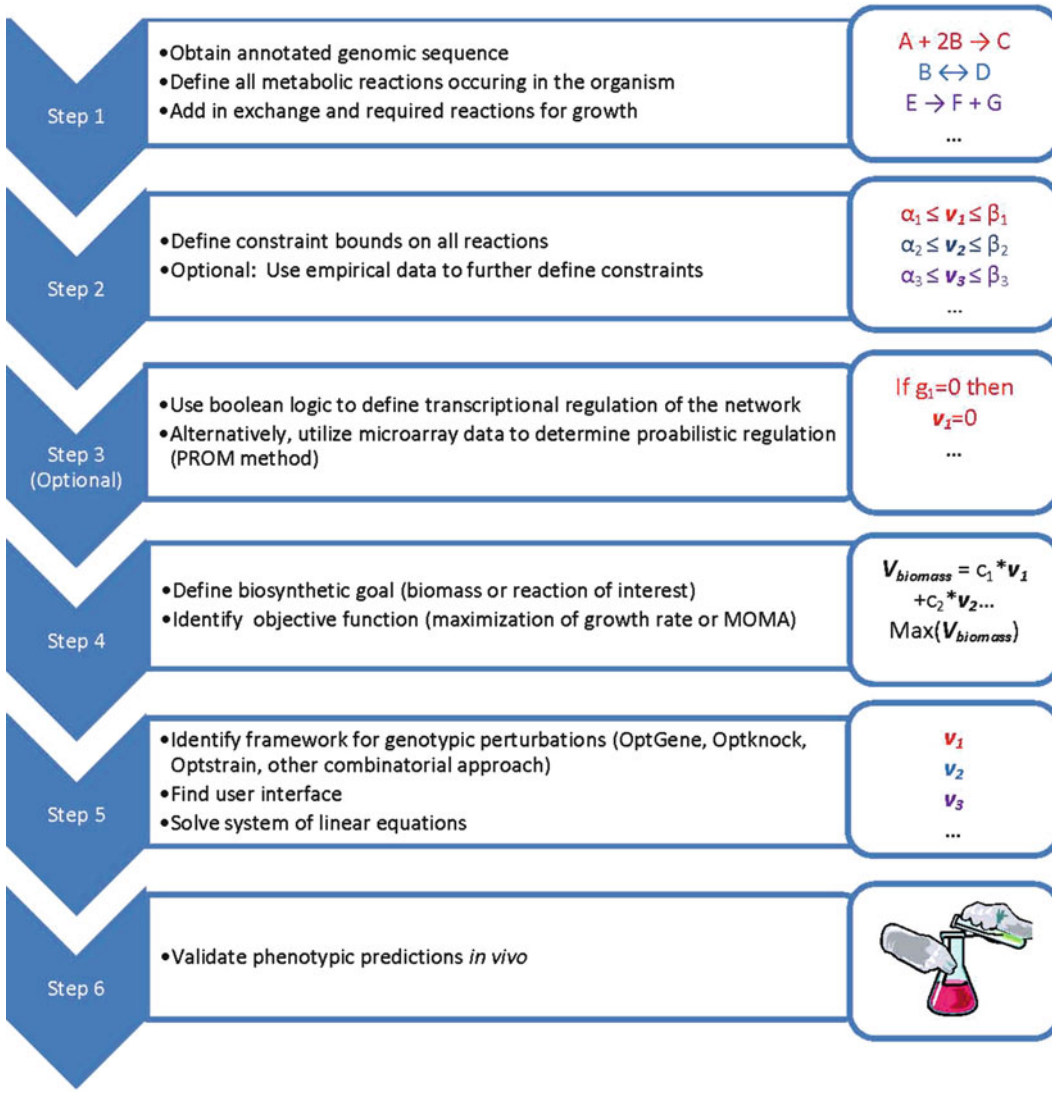


Fig. 1. Flow chart for constructing a genome-scale metabolic model and utilizing FBA.

- Second, each metabolic reaction is bounded by constraints to allow for meaningful solutions of the under-defined system of linear equations.
- Third, optional transcriptional regulation is added to the model through the use of Boolean or probabilistic logic.
- Fourth, the mathematic objective function is defined based on either the biosynthetic objective of the organism, a biotechnological goal, or a hybrid between the two.
- Fifth, the appropriate solution algorithm and software are determined and the system of linear equations is solved by maximizing or minimizing the objective function.
- Sixth, the predicted phenotypic results are validated in vivo.

1.2. Prior Achievements of Flux Balance Analysis

FBA was first introduced in the early 1990s (1, 2) primarily with well-studied organisms, like *Escherichia coli* (3–5). However, with the rapid increase in genome sequence availability, genome-scale metabolic models can be generated for a variety of organisms in an automated fashion (6). As such, there are currently models available for over a hundred different organisms, and the number is growing rapidly (7, 8). Moreover, FBA has been used in a variety of contexts for metabolic engineering tasks and has been successfully used to predict genetic modifications for improved strains.

For example, Alper et al. used FBA to identify gene knockout targets for the production of lycopene in *E. coli* (9). The resulting single, double, and triple knockout strains significantly increased the yield of lycopene, with the highest producing strain increasing yield 37% over the engineered parental strain. Similarly, Lee et al. (10) and Park et al. (11) utilized FBA to identify gene knockouts for increased L-threonine and L-valine production in *E. coli*. The resulting strains produced 0.378 and 0.393 g per gram of glucose of L-threonine and L-valine, respectively.

Significant metabolic engineering has also been accomplished in *Saccharomyces cerevisiae* through the use of FBA. For example, Bro et al. (12) used FBA to identify a strategy to decrease glycerol yield and increase ethanol yield. The resulting strain had a 40% decrease and 3% increase in glycerol and ethanol yields, respectively.

Another unique application of FBA is to use it to identify optimal media conditions for the growth of an organism and production of a desired metabolite. Song et al. (13) used FBA to identify media that would improve the growth of *Mannheimia succiniciproducens* and the production of succinic acid. The identified optimal media increased the yield of succinic acid by 15%. These examples all demonstrate that FBA is a powerful tool for predicting cellular phenotypes. The remainder of this chapter will discuss the steps required for the implementation of this approach.

2. Materials

1. Software and the appropriate hardware (e.g., COBRA toolbox and MATLAB® on PC).
2. Prepared metabolic model OR annotated genome sequence for organism.
3. Empirical phenotype data (optional).

3. Methods

As described above, the basic FBA paradigm can be described by six main steps. These steps include: (1) model construction, (2) defining model constraints, (3) adding in transcriptional regulation, (4) defining an objective function, (5) identifying a solution algorithm and software, and (6) validation in vivo. The detailed procedure for carrying out FBA will depend significantly on the desired phenotype and host cell model that is being used. Moreover, the level of effort involved in each of these steps depends on the ready availability of a working model for the desired host organism. For the case of standard industrial organisms such as *E. coli* and *S. cerevisiae*, these models with constraints are readily available. In this section, the general framework for developing a metabolic model and utilizing it for FBA will be presented. To provide adequate description of this method, each of the six steps is described in two ways in this section. First, this method will be described generically with an emphasis on the fundamental background for the approach as well as various options. Second, in order to more clearly illustrate the methodology, the specific example of identifying gene knockout targets that improve heterologous lycopene production in *E. coli* is selected as a case study (9).

3.1. Step 1: Model Construction

For many commonly used organisms, a genome-scale metabolic model is readily available (and thus, models may be used off the shelf or only slightly updated to include heterologous pathways). However, if this is not the case, the first step in employing a FBA approach is to mathematically reconstruct the metabolic network of the organism in the form of a genome-scale metabolic model. At a minimum, this metabolic network will account for (1) the reactions taking place within the cell and the participating metabolites and cofactors, (2) formulas for biomass, and (3) transport equations between organelles and extracellular space. These components of the metabolic network form the basis for the "stoichiometric matrix"—the mathematical description of cellular reactions. The most common way to do this is to utilize the annotated genome along with published literature as background and determine each metabolic reaction that occurs in the cell. Often, this step can be tedious as quite a bit of hand curation may be necessary for poorly annotated organisms. Thus, a better genome annotation will yield a more accurate (and quicker) reconstruction and prediction. First, it is important to establish all potential reactions in the cell. For this task, these reactions must be elementally balanced, taking into account any cofactor or ATP requirements. These reactions are listed out in equation form. In addition to cellular metabolism, biomass formation and intermembrane transport must also be accounted for. To illustrate the transformation of a metabolic

network to a stoichiometric model, we will consider a simple network,

$$A + 2B \rightarrow C, \tag{1}$$

$$B \rightarrow D, \tag{2}$$

$$B + 2C \rightarrow X_{\text{biomass}}, \tag{3}$$

$$A_{\text{extracellular}} \rightarrow A, \tag{4}$$

$$B_{\text{extracellular}} \rightarrow B, \tag{5}$$

$$D_{\text{extracellular}} \rightarrow D, \tag{6}$$

where A, B, C, and D are metabolites and X_{biomass} is the "metabolite" representing biomass, which is formed from known ratios of intracellular metabolites. These reactions can be rewritten in the form of mass balances as:

$$-A - 2B + C = 0,$$

$$-B + D = 0,$$

$$-B - 2C + X_{\text{biomass}} = 0,$$

$$A - A_{\text{extracellular}} = 0,$$

$$B - B_{\text{extracellular}} = 0,$$

$$D - D_{\text{extracellular}} = 0,$$

where flux into the cell has been defined as positive. The system of equations can be expressed in matrix form as:

$$S = \begin{array}{c} \\ A \\ B \\ C \\ D \\ A_{\text{extracellular}} \\ B_{\text{extracellular}} \\ D_{\text{extracellular}} \\ X_{\text{biomass}} \end{array} \begin{array}{c} \begin{array}{cccccc} (1) & (2) & (3) & (4) & (5) & (6) \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ -2 & -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Also in this framework are reactions driven by enzymes that have not been identified or annotated in the genome and small

molecule reactions that do not require catalysis. With this stoichiometric matrix, we can write a mass balance on the cell:

$$\frac{dx}{dt} = Sv, \tag{7}$$

where x is a column vector of length N representing the concentration of each metabolite, v is a row vector of length M representing the flux through each reaction, and S is the N by M matrix coupling flux through each reaction to the change in concentration of each metabolite. In this simplified model,

$$x = \begin{bmatrix} x_A \\ x_B \\ x_C \\ x_D \\ x_{A_{\text{extracellular}}} \\ x_{B_{\text{extracellular}}} \\ x_{D_{\text{extracellular}}} \\ x_{\text{biomass}} \end{bmatrix}$$

and

$$v = [v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6]$$

Most flux balance analyses treat the organism in pseudo-steady state (see Note 2), so all time derivatives can be set to zero:

$$0 = Sv \tag{8}$$

If the number of reactions equaled the number of metabolites, the state of the cell could be explicitly determined through stoichiometric considerations alone. However, it is most often the case that the number of reactions greatly exceed the number of metabolites, resulting in an underdetermined system. FBA addresses this problem through the use of reaction constraints and an objective function, which is discussed in Subheadings 3.2 and 3.4 respectively.

As outlined above, the task of building an accurate genome-scale model is no simple undertaking; however, new tools are becoming available to make it easier. For example, Henry et al. (8) recently developed an online resource called Model SEED that is designed to simplify the creation of a new model by utilizing an automated framework abbreviated RAST (see Note 3). This resource automates the initial annotation of the genome on the RAST server, performs a preliminary reconstruction of the model, executes auto-completion of missing reactions necessary for

growth, and optimizes the model using phenotyping array and gene essentiality data where available. Using this method, new genome-scale metabolic models can be generated in approximately 48 h (see Note 4).

In many cases, there is no need to create a model from scratch. Instead, there are now models freely available for many commonly studied organisms that can be downloaded and adjusted to fit the needs of a particular project. In addition to the models available at the Model SEED, several research groups have published models for commonly used organisms such as *E. coli*, *B. subtilus*, or *S. cerevisiae,* among others (7). Some of these models are quite robust and represent years of refining. As an example, for *S. cerevisiae*, a "consensus metabolic model" has been created by a consortium of scientists by combining several models and standardizing naming conventions (14) (see Note 5). These preexisting model frameworks provide a significant head start to anyone interested in utilizing them for FBA.

Most freely available models are encoded in a standard format based on the Systems Biology Markup Language (SBML) (15). An SBML formatted model is saved as a .xml file and can be simply imported into most software applications available for FBA.

3.1.1. Lycopene Production Example

In the case of production of lycopene from *E. coli* (9), the *E. coli* genomic-scale metabolic model *i*JE660a GSM (16) was used. This model was the most comprehensive available at the time of the experiments and was downloaded from the deposited website. Lycopene production in this host was heterologously achieved and thus not included in the model. To adjust the model to fit the need of the project, the additional reactions for the lycopene production pathway were added as listed in Table 1.

The resulting model consisted of 965 fluxes (including the exchange/transfer fluxes) and involved 546 metabolite intermediates. This stoichiometric matrix was developed in MATLAB.

3.2. Step 2: Model Constraints

Once all reactions for a metabolic model are defined, it is necessary to place bounds (or constraints) for the value of each reaction flux (v_i). This bounding can be trivial (i.e., between 0 and infinity for nonreversible reactions) or more specific (i.e., a small, finite range of experimentally validated values). As discussed above, the problem of FBA is underdetermined and thus more accurate constraints will yield more meaningful and realistic flux predictions. Constraints can be based on empirical data, such as ^{13}C-based flux analysis (17), or can be based on estimates where data does not exist. Reactions can be set to be reversible by allowing the lower bound of the constraint to be less than zero. For exchange reactions, which are reactions that model the transport of metabolites in and out of the cell (or in and out of organelles in the case of eukaryotes), a negative flux indicates metabolites entering into the cell (or organelle) and

Table 1
Alterations of the iJE606a GSM (9)

Pathway	Gene	Reaction
Isoprenyl-pyrophosphate synthesis pathway	*dxs*	T3P1 + PYR → DXP + CO_2
	ispC	DXP + NADPH ↔ MEP + NADP
	ispD	MEP + CTP → CDPME + PPI
	ispE	CDPME + ATP → CDPMEPP + ADP
	ispF	CDPMEPP → MECPP + CMP
	ispG	MECPP ↔ HMBPP
	ispH	HMBPP → 0.5 IPPP + 0.5 DMPP
Isoprenyl-pyrophosphate isomerase	*idi*	IPPP ↔ DMPP
Farnesyl pyrophosphate synthetase	*ispA*	2IPPP → GPP + PPI
Geranyltranstransferase	*ispA*	GPP + IPPP → FPP + PPI
Octoprenyl pyrophosphate synthase (five reactions)	*ispB*	5 IPPP + FPP → OPP + 5 PPI
Undecaprenyl pyrophosphate synthase (eight reactions)		8 IPPP + FPP → UDPP + 8 PPI
Lycopene pathway	*crtE*	IPPP + FPP → GGPP + PPI
	crtB	2 GGPP → PHYTO + PPI
	crtI	PHYTO + 8 NADP → LYCO + 8 NADPH

a positive flux indicates metabolites exiting. Since knowledge of the precise value for these constraints is often hard to obtain, it is not unusual to simply bound reactions between zero and a high number for irreversible reactions and between a large negative and large positive value for reversible reactions. In this regard, the constraints merely exist as pseudo-thermodynamic constraints.

Once a model is assembled for predicting the wild-type phenotype, defining constraints is the one of the easiest ways to modify a system. First, environmental conditions are easily adjusted by simply changing the constraint bounds for the extracellular exchange reactions. For example, to crudely model anaerobic growth, constraints on the exchange reaction for oxygen are simply set to zero (see Note 6). Similar adjustments can be made to model growth on alternative carbon sources or to change limiting nutrients. Second, constraints for internal metabolic reactions can be changed to model gene knockouts or over- and under-expression. This is accomplished by changing the bounds on the reactions that are catalyzed by the gene product of interest.

3.2.1. Lycopene Production Example

For the production of lycopene in *E. coli* (9), the constraints that were initially included with the *i*JE660a GSM model were used. The constraints for the added reactions were all set such that they were nonreversible with the exception of *idi*, *ispC*, and *ispG*.

Metabolite abbreviations used in Table 1 (9)

Abbreviation	Metabolite
CDPME	4-Diphophocytidyl-2-C-methyl-D-erythritol
CDPMEPP	4-Diphosphocytidyl-2-C-methyl-2-phosphate-D-erythritol
DMPP	Dimethylallyl pyrophosphate
DXP	1-Deoxy-D-xylulose-5-phosphate
FPP	Trans, trans Farnesyl pyrophosphate
GGPP	Geranylgeranyl PP
GPP	Trans Geranyl pyrophosphate
HMBPP	1-Hydroxy-2-methyl-2(E)-butenyl-4-diphsophate
IPPP	Isopentyl pyrophosphate
LYCO	Lycopene
MECPP	2-Methyl-D-erythritol-2,4-cyclodiphosphate
MEP	Polyol 2-C-methyl-D-erythritol-4-phosphate
OPP	Trans Octaprenyl pyrophosphate
PHYTO	Phytoene
PPI	Pyrophosphate
PYR	Pyruvate
T3P1	Glyceraldehyde 3-phosphate
UDPP	Undecaprenyl pyrophosphate

Since glucose was determined to be the limiting nutrient in vivo, this was accomplished by altering the exchange fluxes. As such, glucose, oxygen, and nitrogen uptake rates were set to 5,200 and 1,000, respectively to correlate with the g/L concentrations seen in typical cultures.

3.3. Step 3: Transcriptional Regulation (Optional)

While traditional stoichiometry-based metabolic models are essentially a compilation of mass balance equations, there is a significant interest to move beyond the reaction equation to the regulation influencing these reactions. In some respects, this shift allows these models to shift from feasible flux modes to probable ones. Usually, genes are included in these models as Boolean operators on reactions and can therefore be either on or off. A gene that is turned off sets the flux through all associated reactions to zero by setting the corresponding constraints to zero. This formulation allows for simpler gene knockout simulations since gene products can catalyze

multiple reactions. An extension of this idea can be used to add some transcriptional regulation into the model, known as regulatory flux balance analysis (RFBA) (18). When a particular gene is dependent on a transcriptional factor, the reactions catalyzed by that gene of interest can be listed as dependent on the transcriptional factor. The presence of the transcriptional factor can then be specified using Boolean logic. The common Boolean operators such as AND, OR, and NOT can be utilized as well, allowing for transcriptional dependence on multiple factors. Therefore, when the gene for the transcriptional factor is knocked out all downstream targets are regulated appropriately by adjusting the constraints of the corresponding reactions.

While this method allows some transcriptional regulation to be included in the model, it is unfortunately a highly manual process and lacks finesse due to its on/off nature. A transcriptional factor knockout may therefore be predicted as lethal, when in vivo the transcription factor knockout may only have a partial affect on the transcription of target genes. To address this, Chandrasekaran and Price (19) developed a probabilistic method to predict gene states based on transcription factor expression, called probabilistic regulation of metabolism (PROM). The PROM method utilizes microarray data to determine the probability that a particular gene will be turned on or off when a transcription factor is knocked out. That probability is then used to adjust the constraint bounds for downstream reactions. This allows for a much more accurate and complete regulatory model, and has the additional advantage of being relatively automated rather than requiring manual curation. In the initial application of this method on models for *E. coli* and *Mycobacterium tuberculosis*, knockout phenotypes were able to be predicted with accuracies of 95% (19).

3.3.1. Lycopene Production Example

For the example of production of lycopene from *E. coli* (9), a purely stoichiometric model was used and thus there was no transcriptional regulation added to the model. However, the addition of transcriptional regulation could help suggest further targets for gene knockout since that may influence the flux through a desirable pathway.

3.4. Step 4: Objective Function

The final step in defining the metabolic model is to determine the objective function. As stated above, the system of linear equations that comprises a metabolic model is almost always underdetermined (which means that many possible solutions exist to satisfy the basic material balances). For example, the consensus model for yeast contains several hundred more metabolic reactions than metabolites, implying that this many unknowns must be specified or determined before the complete flux map can be determined. The model constraints discussed in Subheading 3.2 place limits on the values of these unknowns, but these constraints are not enough

to explicitly solve for a singular solution to this problem. The most common method used to solve for an explicit solution involves an objective function which a living organism will maximize (or minimize). Two commonly used objective functions are the maximization of growth (see Note 7) and the minimization of metabolic adjustment (MOMA) (20). Beyond biological objective functions, it is also possible to use biotechnological goals such as maximal product formation.

To maximize growth, the biomass flux, v_{biomass}, is utilized as defined in the model in Subheading 3.1. An important property of v_{biomass} is its linearity; the flux through this reaction depends on a simple linear combination of metabolite levels. The state of the cell is then determined by maximizing v_{biomass} subject to (13.8) and any constraints. Since v_{biomass} is a linear objective function, the solution strategy utilizes a simple Linear Programming (LP) algorithm, which can be solved by any of the software packages described in Subheading 3.5 in a matter of minutes on a typical computer processor (21). FBA using the maximization of growth as an objective function has been shown to be an accurate predictor of growth rate, metabolite secretion, and gene essentiality in wild-type strains (3, 22).

An inherent assumption of using growth maximization is that the organism of interest has had time to adjust metabolic fluxes to maximize growth rates, which is a good assumption for wild-type strains but perhaps less so for knockout strains. Thus, an alternative objective function, the minimization of metabolic adjustment (MOMA), is a method for more accurately solving this problem. In this regard, MOMA solves the problem of an underdetermined system by finding the solution which is most similar to a "reference" pattern of metabolic flux (typically obtained by maximizing the growth in a wild-type strain). MOMA has found its greatest applicability in perturbed systems (e.g., knockouts) where the organism has not had sufficient time to redistribute fluxes so as to maximize growth (9). Hence, the metabolic state of the perturbed organism is assumed to be as similar as possible in flux distributions to its unperturbed state. Thus, the unperturbed state, assuming the organism has had time to acquire an evolutionarily favorable flux distribution, can be determined with FBA. Alternatively, experimental flux values for a wild-type organism can serve as a reference state. If v represents the fluxes in the "reference" state and w the fluxes of interest, then the quantity which MOMA seeks to minimize is:

$$D(w,v) = \sqrt{\sum_{i=1}^{M} (w_i - v_i)^2} \tag{9}$$

which is simply the Euclidean distance between w and v. It is important to note that this objective function is nonlinear,

necessitating the use of quadratic programming (QP) to solve the minimization problem, which is solvable with several of the packages mentioned in Subheading 3.5 (21). Moreover, the computational time for solving a QP problem is higher than a LP problem, however, both are easily solvable in several minutes on a standard computer. MOMA has been found to be more accurate than FBA in predicting growth rates and metabolic fluxes in knockout strains (20). Both FBA and MOMA can be used in isolation to predict the phenotype of a single organism or they can be a part of a larger search strategy to discover perturbations resulting in an improved phenotype.

3.4.1. Lycopene Production Example

In the example of lycopene production from *E. coli* (9), the biomass objective function was maximized to provide the reference fluxes for MOMA calculations. A MOMA algorithm was utilized to predict the flux distributions and lycopene production for the single, double, and triple gene knockout analyses.

3.5. Step 5: Solution Algorithms and Software

Once a metabolic model and solution algorithm has been chosen and validated (at least for the wild-type organism), all that remains is to optimize the phenotype of interest. Many components found in metabolic models can be experimentally perturbed to improve organism performance, such as media concentration and genotype. We discuss methods for modeling these two major perturbations below:

3.5.1. Environmental Perturbations

Media conditions can be altered in silico to establish linkages between nutrient levels and the desired phenotype (often production formation). A simple scheme for exploring these linkages is as follows:

1. Solve the metabolic model under standard environmental conditions.
2. Identify condition of interest (oxygen level, glucose uptake, pH, etc.).
3. Perturb condition of interest by changing the upper or lower bounds on the relevant transport reactions.
4. Solve the metabolic model under the perturbed conditions.
5. Observe a phenotypic trend with respect to the condition of interest and return to step 3, if necessary.

Consequently, this approach can identify any nutrient limitations which may prevent the organism of interest from achieving the desired phenotype. This technique has been used to predict necessary media conditions for recently annotated organisms (8), to improve growth and product formation (13), and to validate new models. In the latter case, experimental nutrient essentiality data was compared with data obtained in silico to identify possible

gaps in a recently curated metabolic model of *E. coli* (23). It should be noted, however, that current FBA models consistently underestimate growth in anaerobic conditions, perhaps reflecting a fundamental shortcoming in their construction (19).

3.5.2. Finding Perturbations with Gene Knockouts

FBA has broad applicability in rational metabolic engineering efforts due to its computational tractability and the scope of biosynthetic problems it can address. Moreover, multiple search strategies can be used to test possible mutants in silico before an expensive experimental test is undertaken. In addition to a manual knockout process, many algorithms have been developed to make this search more efficient, including OptKnock, OptGene, and OptStrain.

Manual Genome Scan

The simplest approach for identifying potential knockouts with FBA is a manual genome scan in which all sets of knockouts are sequentially tested in silico. Knockouts are added to the FBA framework by rewriting the constraints on the reactions occurring in the metabolic model as:

$$\alpha_i \cdot y_i \le v_i \le \beta_i \cdot y_i, \tag{10}$$

where α_i and β_i are the upper and lower bounds on the reaction, and v_i is the flux of the reaction as discussed in Subheading 3.1. The additional variable y_i is constrained to be in the set {0,1} and thus specifies whether or not the reaction is knocked out. The problem is then to find the y which maximizes v_{product} in a growth-maximizing or metabolic adjustment-minimizing organism. An example scheme for the manual process is as follows:

1. Solve the metabolic model for the wild-type organism, usually maximizing growth.
2. Make a set of perturbations to y (i.e., set some elements of y to zero).
3. Solve the metabolic model in the knockout organism (with either maximization of growth or MOMA as on objective) and note any predicted phenotypic changes.
4. Return to step 2 if the set of possible knockouts has not been exhausted, making perturbations either to the wild-type or knockout organism.
5. Identify the set of knockouts resulting in the best predicted phenotype.

OptKnock

Due to the large number of genes present in a typical metabolic model (~1,000), an exhaustive search of knockout sets becomes computationally intractable as the size of the sets increase. Therefore, OptKnock (24) exploits duality theory to recast this problem into an easily solvable form. Duality theory states that for every "primal" LP problem seeking to maximize an objective function,

there is a unique "dual" LP problem whose optimal objective function equals that of the primal. By constructing the dual problem for biomass maximization and simply setting the two objective functions equal to one another as an additional constraint, the problem of maximizing product yield is coupled to maximizing growth as a single mixed-integer linear programming (MILP) problem, so named due to the discreteness of y. OptKnock and other Optx algorithms are usually available as a standalone program or as part of a package, making knockout prediction a one-step process. In this easily tractable form, sets of up to four gene knockouts for improved succinate, lactate, and 1,3-propanediol production in *E. coli* have been predicted (24).

OptGene

One significant drawback of the OptKnock framework is its inability to optimize nonlinear objective functions, such as metabolic adjustment or productivity ($v_{\text{biomass}} \cdot v_{\text{product}}$). Furthermore, computational time for the relatively tractable MILP problem becomes more burdensome as larger sets of knockouts are considered. To address these drawbacks, OptGene (25) searches for an optimal y through a genetic algorithm, described schematically below:

1. Generate a population of random ys.
2. Solve the metabolic model for each member of the population through FBA, MOMA, etc.
3. Assign each member of the population a "score" based upon some function of metabolic state (such as product formation or productivity).
4. Allow the ys with the highest score to "reproduce," either through "crossovers" with other high-scoring ys (akin to sexual reproduction), or through point mutation of a single high-scoring y (akin to asexual reproduction), or both: generating a new population of ys.
5. Return to step 2 until no further improvement in the best score is observed.

Although this algorithm is much faster than MILP and is compatible with nonlinear objective functions, it is possible that the population will converge to a local optimum, unlike for MILP. OptGene has been used to predict sets of up to five gene knockouts resulting in improved vanillin, succinate, and glycerol production in *S. cerevisiae* (25). For both OptKnock and OptGene, the computation time is increased compared to a simple linear programming solution.

OptStrain

Although OptKnock and OptGene are very powerful in their ability to predict knockouts, the space of possible mutants is ultimately constrained to alterations of the reactions present in the metabolic model. OptStrain (26) addresses this concern by searching a universal

database of biotransformations for sets of heterologous pathways predicted to improve the yield of a compound of interest. OptStrain performs this search according to the scheme below:

1. Determine the maximum possible product formation from among the reactions present in the Universal database using a LP approach. This maximum becomes the "baseline" yield.
2. Use MILP to identify the minimum number of heterologous genes (from the Universal database) required to augment a given host so that it meets or exceeds the baseline yield. It is important to note that in this calculation, product yield is calculated assuming the organism of interest is maximizing product yield, not maximizing growth or minimizing metabolic adjustment. This simplification allows promising pathways to be selected while still keeping the problem tractable.
3. The genes identified in step 2 are added to the stoichiometric model, and an OptKnock-like calculation is undertaken to further optimize the organism. Thus, a set of insertions and knockouts is found which are predicted to improve product formation.

Using this method, an in silico optimization of hydrogen production was undertaken in the context of *E. coli*, *C. acetobutylicum*, and *M. extorquens*, and a similar process was undertaken for vanillin production in *E coli* (26).

3.5.3. Finding a User Interface

There are several software options that are now freely available for researchers wishing to pursue FBA. As mentioned before, most software packages employ SBML files, but some can also use flat (text) files or create their own file format. The four software packages described here were chosen for their ease of use and flexibility to aid in FBA and metabolic engineering specifically. However, it may be beneficial to explore other options as well depending on the scope of the project (see Note 8).

- COBRA Toolbox
  The Constraint-Based Reconstruction and Analysis (COBRA) toolbox (27) is a software package that utilizes MATLAB® to perform maximization of growth and MOMA analyses. For researchers familiar with the MATLAB® environment, it is simple to use and allows a large amount of flexibility in project scope as each function (i.e., solution algorithms, objective functions, and tools to adjust the model) is defined in a separate program file. This allows the user to utilize the different functions in the way best suited for the project. The COBRA Toolbox is available at http://gcrg.ucsd.edu/Downloads/Cobra_Toolbox. There are step-by-step instructions available and several good tutorials for the COBRA Toolbox have been published (28).

- OptFlux
  OptFlux (29) is similar to the COBRA toolbox in its capacity to perform both maximization of growth and MOMA optimization. However, instead of requiring and interfacing with MATLAB®, it is a stand-alone Java™ script with a simple user-friendly interface. Additionally, OptFlux has the capability to utilize the OptKnock algorithm and to add in Boolean logic regulation. Due to its Java™-script interface, OptFlux isn't as flexible as the COBRA toolbox for simple one-time changes. However, OptFlux does have the advantage of being an open-source software package that allows contributions from more advanced users. Furthermore, the plug-in architecture will allow development of new functionalities within OptFlux in the future. OptFlux is available at http://www.optflux.org/.
- CellNetAnalyzer
  CellNetAnalyzer (30), like the COBRA Toolbox, utilizes MATLAB®. However, it is presented in a graphical user interface that is simpler than the traditional MATLAB® command window. In addition to metabolic fluxes, CellNetAnalyzer has the capability to allow for significant Boolean regulation of the model, and further eases this task by making a variety of interactive tools and visualizations available. However, for the metabolic network solution, CellNetAnalyzer does not have the capability to utilize MOMA or any other advanced solving techniques other than maximization of growth. CellNetAnalyzer is available to academic researchers at http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html.
- SBRT
  The Systems Biology Research Tool (SBRT) (31) is a broad scope software package. It was designed to facilitate many different aspects of systems biology in addition to FBA. SBRT also runs in Java™ script and is open source and plug-in ready. SBRT is available at http://www.bioc.uzh.ch/wagner/software/SBRT/.

3.5.4. Lycopene Production Example

In the case of lycopene production in *E. coli* (9), a sequential knockout approach was undertaken to identify promising knockout combinations. First, wild-type lycopene production was determined assuming maximization of growth. Second, lycopene production was determined under every single gene knockout assuming minimization of metabolic adjustment (MOMA) from wild-type. The second step was then repeated with the most promising single knockout until the optimal single, double, and triple gene knockouts were identified. To confirm the efficacy of this sequential approach, an exhaustive search of all double knockouts was undertaken. Interestingly, the best-performing mutant from

this exhaustive search was the most promising double-knockout isolated from the sequential approach.

This example was completed prior to the publication of the above software packages. As such, it was done by utilizing a PERL script interface provided by Dr. Daniel Segre to enable the MOMA code. However, for the purpose of doing multiple gene knockout in series and utilizing MOMA, the COBRA toolbox would now be an appropriate selection.

3.6. Step 6: Validation In Vivo

The final step in any FBA problem is to validate the predicted solution in vivo. Obviously, this step is highly project specific; however, there are some general guidelines one should keep in mind. First, by using maximization of biomass as an objective function, the underlying assumption is that a cell is always growing and the product is growth-associated. This assumption may not always be the case, especially when utilizing nonoptimal conditions or knockout strains. Ibarra et al. (32) found that in the case of growth of *E. coli* on glycerol, evolution of the lab strain by selective pressure was required to meet the predicted in silico growth rate. Second, one must keep in mind that the purely stoichiometric framework of FBA does not take into account a variety of variables that occur in vivo, like thermodynamics, kinetics, or genomic regulation (in cases where this is not added to the model). However, with careful selection of constraints and objective functions it is possible to generate accurate and highly novel results.

3.6.1. Lycopene Production Example

Once a series of single, double, and triple gene knockouts had been identified for increasing production of lycopene in our example (9), they were created in vivo in individual strains of *E. coli* harboring a plasmid with the necessary carotenoid genes. Each strain was then cultured and tested for lycopene production by measuring absorbance of the fractionated cellular supernatant at 475 nm. This validation confirmed the overall predicted results, and yielded a triple knockout with a nearly 40% increase in lycopene production over the parental strain.

4. Notes

1. Transcription regulation and other similar considerations (thermodynamics, kinetics, etc.) are beginning to be incorporated into some models (see Subheading 3.3); however, most models still only take into account stoichiometric considerations.
2. Pseudo-steady state is a good assumption when there is little net accumulation of metabolites in cells (e.g., exponential phase). However, it has been shown that transient models are necessary, for example, when doing ^{13}C labeling experiments on CO_2-fixing organisms (33).

3. The Model SEED can be found at http://www.theseed.org/models/. RAST is Rapid Annotation using Subsystem Technology (http://rast.nmpdr.org/).
4. Of the 130 models that were generated for the study, 22 were validated and optimized using phenotype and gene essentiality data. The optimized models demonstrated 66% accuracy in predicting phenotype before optimization and 87% after optimization (8).
5. The current version of the consensus model (ver. 4, available at http://www.comp-sys-bio.org/yeastnet/) contains 2,342 reactions (of which 2,183 are metabolic) and 2,657 chemical species (of which 1,494 are metabolites and 1,163 are proteins or complexes).
6. Setting the exchange rate of oxygen to zero only crudely models anaerobic growth because the switch from aerobic growth to anaerobic growth also induces a large number of regulatory changes that are not incorporated into a purely stoichiometric model.
7. This is often somewhat confusingly referred to as flux balance analysis or FBA. Here we will use FBA to mean the overall technique rather than this specific maximization.
8. A good source to learn about additional software packages is the SBML website, http://sbml.org.

References

1. Savinell, J. M., and Palsson, B. O. (1992) Optimal Selection of Metabolic Fluxes for *in vivo* Measurement .1. Development of Mathematical Methods, *J. Theor. Biol. 155*, 201–214.
2. Savinell, J. M., and Palsson, B. O. (1992) Optimal Selection of Metabolic Fluxes for *in vivo* Measurement .2. Application to *Escherichia coli* and Hybrdoma Cell Metabolism, *J. Theor. Biol. 155*, 215–242.
3. Varma, A., and Palsson, B. O. (1994) Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110, *Appl. Environ. Microbiol. 60*, 3724–3731.
4. Varma, A., and Palsson, B. O. (1993) Metabolic Capabilities of *Escherichia coli* .1. Synthesis of Biosynthetic Precursors and Cofactors, *J. Theor. Biol. 165*, 477–502.
5. Varma, A., and Palsson, B. O. (1993) Metabolic Capabilities of *Escherichia coli* .2. Optimal Growth Patterns, *J. Theor. Biol. 165*, 503–522.
6. Edwards, J. S., and Palsson, B. O. (1999) Systems properties of the Haemophilus influenzae Rd metabolic genotype, *J. Biol. Chem. 274*, 17410–17416.
7. Oberhardt, M. A., Palsson, B. O., and Papin, J. A. (2009) Applications of genome-scale metabolic reconstructions, *Mol. Syst. Biol. 5*.
8. Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models, *Nat. Biotechnol. 28*, 977–U922.
9. Alper, H., Jin, Y. S., Moxley, J. F., and Stephanopoulos, G. (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli, *Metab. Eng. 7*, 155–164.
10. Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U., and Lee, S. Y. (2007) Systems metabolic engineering of Escherichia coli for L-threonine production, *Mol. Syst. Biol. 3*.
11. Park, J. H., Lee, K. H., Kim, T. Y., and Lee, S. Y. (2007) Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation, *Proc. Natl. Acad. Sci. USA 104*, 7797–7802.
12. Bro, C., Regenberg, B., Forster, J., and Nielsen, J. (2006) In silico aided metabolic engineering of

Saccharomyces cerevisiae for improved bioethanol production, *Metab. Eng. 8*, 102–111.

13. Song, H., Kim, T. Y., Choi, B. K., Choi, S. J., Nielsen, L. K., Chang, H. N., and Lee, S. Y. (2008) Development of chemically defined medium for Mannheimia succiniciproducens based on its genome sequence, *Appl. Microbiol. Biotechnol. 79*, 263–272.
14. Herrgard, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Bluthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novere, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasic, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kirdar, B., Penttila, M., Klipp, E., Palsson, B. O., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J., and Kell, D. B. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology, *Nat. Biotechnol. 26*, 1155–1160.
15. Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, II, Hedley, W. J., Hodgman, T. C., Hofmeyr, J. H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novere, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., and Forum, S. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics 19*, 524–531.
16. Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR), *Genome Biol. 4.*
17. Kim, T. Y., and Lee, S. Y. (2006) Accurate metabolic flux analysis through data reconciliation of isotope balance-based data, *J. Microbiol. Biotechnol. 16*, 1139–1143.
18. Covert, M. W., Schilling, C. H., and Palsson, B. (2001) Regulation of gene expression in flux balance models of metabolism, *J. Theor. Biol. 213*, 73–88.
19. Chandrasekaran, S., and Price, N. D. (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences 107*, 17845–17850.
20. Segre, D., Vitkup, D., and Church, G. M. (2002) Analysis of optimality in natural and perturbed metabolic networks, *Proc. Natl. Acad. Sci. USA. 99*, 15112–15117.
21. Vaderbei, R. J. (2001) *Linear Programming: Foundations and Extensions*, 2 ed., Kluwer Academic Publishers, Boston.
22. Edwards, J. S., Ibarra, R. U., and Palsson, B. O. (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data, *Nat. Biotechnol. 19*, 125–130.
23. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. O. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information, *Mol. Syst. Biol. 3.*
24. Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization, *Biotechnol. Bioeng. 84*, 647–657.
25. Patil, K. R., Rocha, I., Forster, J., and Nielsen, J. (2005) Evolutionary programming as a platform for in silico metabolic engineering, *BMC Bioinformatics 6.*
26. Pharkya, P., Burgard, A. P., and Maranas, C. D. (2004) OptStrain: A computational framework for redesign of microbial production systems, *Genome Res. 14*, 2367–2376.
27. Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O., and Herrgard, M. J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox, *Nat. Protoc. 2*, 727–738.
28. Orth, J. D., Thiele, I., and Palsson, B. O. (2010) What is flux balance analysis?, *Nat. Biotechnol. 28*, 245–248.
29. Rocha, I., Maia, P., Evangelista, P., Vilaca, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C., and Rocha, M. (2010) OptFlux: an open-source software platform for in silico metabolic engineering, *BMC Syst. Biol. 4.*
30. Klamt, S., Saez-Rodriguez, J., and Gilles, E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer, *BMC Syst. Biol. 1.*
31. Wright, J., and Wagner, A. (2008) The Systems Biology Research Tool: evolvable open-source software, *BMC Syst. Biol. 2.*
32. Ibarra, R. U., Edwards, J. S., and Palsson, B. O. (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth, *Nature 420*, 186–189.
33. Shastri, A. A., and Morgan, J. A. (2007) A transient isotopic labeling methodology for C-13 metabolic flux analysis of photo auto trophic microorganisms, *Phytochemistry 68*, 2302–2312.

Chapter 14

Using an Advanced Microfermentor System for Strain Screening and Fermentation Optimization

Dongming Xie

Abstract

Industrial biotechnology employs microorganisms (strains) for manufacture of certain food or industrial products to meet the increasing need of the world. To develop a bioproduction process, the first step is to screen out a production strain from isolated, mutated, or genetically engineered strain candidates. To maximize the bioproduction of a selected strain, bioreaction (fermentation) conditions need to be optimized. Fermentation experiments in shake flasks, bench-scale fermentors, or a combination of both are the conventional methods for both strain screening and fermentation optimization. Shake-flask experiments are easy to handle and cost-effective compared to experiments in fermentors, but the lower controllability makes the shake-flask data less informative for fermentation scale-up. Bench-scale fermentor experiments (>0.5 L) are well controlled under designed conditions and provide high-quality data, but they are also very time- and cost-consuming. The novel microfermentor system (typically <100 mL), or mentioned as microbioreactor, mini-fermentor, mini-bioreactor, or miniature bioreactor, combines the advantages of both shake-flask's easy handling and bench-scale fermentor's controllability, thus can achieve comparable results from fermentors at much higher efficiency and lower cost. This chapter introduces an example of how to use a microfermentor system for strain screening and fermentation optimization.

Key words: Industrial biotechnology, Fermentation, Microfermentor, Microbioreactor, Strain screening, Optimization

1. Introduction

Industrial biotechnology, or industrial microbiology, employs microorganisms (strains) to produce a wide range of products, including enzymes, foods, beverages, chemical feedstocks, fuels, and pharmaceuticals, and is showing more and more economical, environmental, and social importance throughout the world (1). With the help of modern biotechnology, microorganisms are isolated, mutated, and/or genetically engineered to make the needed

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_14,

products using renewable resources. With all the strain candidates, biotechnology process development is to screen strains and then optimize fermentation conditions to increase yields and productivity. Currently, shake flasks and/or conventional microtiter plates are the main culture tools employed for high-throughput screening of potential strain candidates and optimization of fermentation conditions (2). However, many of the strains selected by these methods fail to perform well in bench-scale fermentors or bioreactors. The fermentation conditions determined by these methods often need to be adjusted when applied to fermentor experiments. The low data density (usually only end-point measurements) and the lack of control over cultivation conditions in standard shake flasks and microtiter plates make it difficult to guide the further strain evaluation and process optimization in bench-scale fermentors.

Recently, a novel microfermentor or microbioreactor technology has been developed. A microfermentor system integrates multiple reactors with a working volume of only 0.1–100 mL (2–6). Different from shake flasks, each reactor's temperature, dissolved oxygen level, and in some cases pH values can be monitored and controlled. With such a high controllability, the system becomes a scale-down version of bench-scale fermentors with a capacity to acquire real-time experimental data via cheap and high-throughput experimentation. It combines the advantages of both shake-flask's easy handling and bench-scale fermentor's controllability, thus can achieve comparable results from fermentors at much lower costs (3).

With the help of a microfermentor system, the process flow for strain screening and evaluation includes the following three typical steps (Fig. 1):

1. A large-scale screening for thousands of strain candidates in micro plates or test tubes.
2. A continued screening for hundreds of strain candidates in shake flasks, microfermentors, or a combination of both.
3. A more detailed screening for dozens of strain candidates in bench-scale fermentors.

After all the three screening steps, the selected strain would be then tested and further evaluated in pilot- or commercial-scale fermentors. For fermentation optimization, the strain selected from the steps above needs to be further tested in microfermentors (or microbioreactors) under different medium and process (e.g., T, pO_2, pH, etc.) conditions to improve production titer, rate, and/or yield.

Currently, there are many microfermentor (or mentioned as microbioreactor, mini-bioreactor, mini-fermentor, miniature bioreactor, etc.) systems available on the market (2–6). Some typical

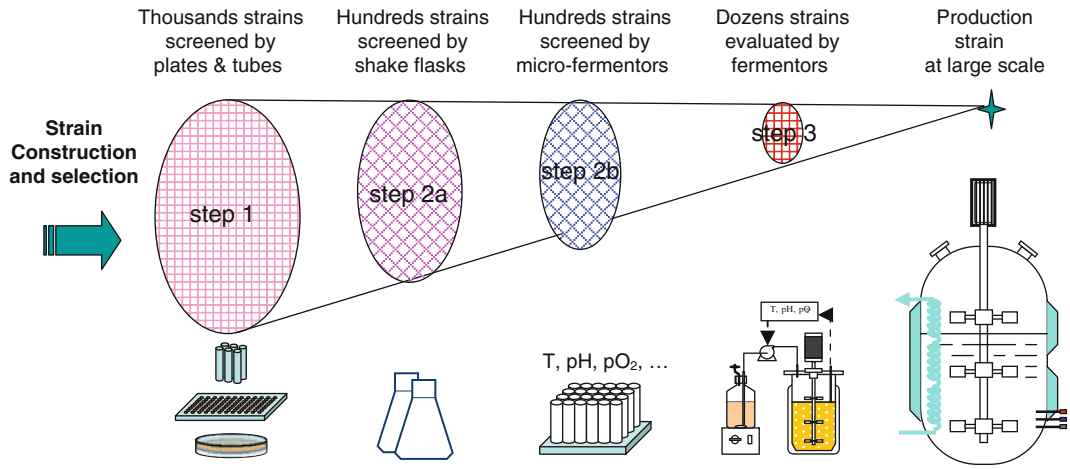


Fig. 1. The process flow for strain screening and evaluation.

mini-scale fermentation systems (<100 mL) suitable for microbial fermentation are summarized in Table 1, which include shake flasks (7–9), test tubes (10, 11), microtiter plates (12, 13), mini-bubble columns (14, 15), and mini-stirred vessels (16, 17). This chapter introduces an example of using a shaken mini-bubble column system, Applikon Micro-24 Bioreactor (14), for both strain screening and fermentation optimization. Details of the protocols include setting up a microfermentor system, starting a microfermentor run, maintaining a run, shutting down a run, and post-fermentation analysis. Although the experimental methods for microfermentor setup and maintenance are described by using the example of the Micro-24 Bioreactor system, the purpose is to show the importance of using any type of micro-/mini-fermentors for high-throughput strain screening and fermentation optimization. The described protocols for medium and feed preparations, seed culture, some sampling or maintenance operations of the reactor system, feed of buffer and sugar for both pH and carbon-source controls, and post-fermentation analysis can also be directly applied to or modified and used for many other micro-/mini-fermentor systems.

2. Materials

2.1. Microfermentor System and Setup Materials

1. The Applikon Biotechnology® Micro-24 Bioreactor (14) is used as an example of the microfermentor system in this chapter. It includes the following:
   (a) A microreactor console with cover.

Table 1
Some typical mini-scale fermentation systems (<100 mL)

Fermentor (references)	Main features	Remaining issues
• Shake flask (7–9)	• Easy handling and cost-effective • Typically dozens of flasks in a shaker • Size ≤ 100 mL • On-line pH and pO_2 measurements are available for some advanced flask system (8, 9) • Widely used for microbial fermentation and cell culture	• Low OTR/mass transfer capacity • On-line pH and pO_2 control usually not available • Still a challenge for fed-batch process
• Test tube (10, 11)	• Easy handling and cost-effective • Typically hundreds of test tubes in a shaker • Typical size ≤ 10 mL	• Low OTR/mass transfer capacity • On-line pH and pO_2 measurement and control usually not available • Still a challenge for fed-batch process
• Microtiter plate (12, 13) • Examples—BioLector microbioreactor (19) and SimCell microbioreactor (20)	• High-throughput screening with dozens to thousands wells/reactors • Typically working volume <1.0 mL • On-line biomass, *T*, pH, and pO_2 measurement and control	• Too small size to provide enough samples for further titer, rate, and yield analysis for microbial system • Still a challenge for individual fed-batch process • Risk of cross-contamination
• Mini-bubble column, including shaken and fixed bubble columns (15) • Example of shaken bubble column—Applikon micro-24 bioreactor (14)	• Up to 24 reactors in parallel • 3–7 mL working volume • Air/O_2 ventilation for pO_2 control • Usually low mass transfer rate, but can be significantly improved by sparging O_2 and increasing shaking speed (Micro-24 Bioreactor) • On-line *T*, pH, and pO_2 measurement • Individual *T*, pH, and pO_2 control • For cell culture and microbial fermentation	• Using NH_3 as base for pH control (Micro-24 Bioreactor), still a challenge for processes requiring nitrogen limitation • Still a challenge for fed-batch process
• Mini-stirred vessel (16, 17)	• 4–16 vessels in parallel • 10–100 mL working volume • Air/O_2 ventilation and agitation for pO_2 control	• Relatively complicated operations compared to other micro- or mini-fermentor system

(continued)

Table 1 (continued)

Fermentor (references)	Main features	Remaining issues
• Examples—Cellstation high-throughput bioreactors (21) and DAS GIP parallel bioreactor system (22)	• On-line *T*, pH, pO_2, off-gas O_2, and CO_2 measurement • Individual *T*, pH, and pO_2 control • For cell culture and microbial fermentation	• Relatively high capital cost

(b) A laptop PC with a USB connection cable and the preinstalled microreactor application software.

(c) An optional vessel for generating ammonia gas for pH control (see Note 1).

2. The consumable materials:

(a) The bioreactor cassettes (24 regular reactors/cassette, 3–7 mL working volume for each reactor).

(b) The cassette closure (type A caps for $pO_2 < 20\%$ and type D caps for $pO_2 > 20\%$).

3. House gases or gas cylinders for experimental setup:

(a) Air, 90 psi or above.

(b) Air, downregulated to 10–30 psi.

(c) O_2, downregulated to 10–30 psi.

(d) CO_2, downregulated to 10–30 psi.

(e) N_2, downregulated to 10–30 psi.

2.2. Microbial Strains and Culture Medium

1. Microbial cells (yeast, bacteria, etc.) prepared as glycerol vials (0.5–2 mL) and stored in a –80°C freezer (see Note 2).

2. Sterilized cultured medium required for growing the microbial cells, including carbon source (glucose, etc.), nitrogen source (yeast extract, corn steep liquor, $(NH_4)_2SO_4$, urea, and/or any other nitrogen sources), phosphate buffer ($H_2PO_4^-$ and/or HPO_4^{2-}), and/or any other necessary nutrients (SO_4^{2-}, Mg^{2+}, trace metals, etc.). For the culture process using any gas flows (air, O_2, N_2, CO_2, etc.), autoclaved antifoam 204 (Sigma) is also added to the medium to 100–1,000 ppm to avoid foaming.

(a) For example, a typical minimal *E. coli* fermentation medium (1 L): 7.5 g K_2HPO_4, 2.1 g citric acid, 0.3 g ammonium iron (III) citrate, 0.5 g antifoam 204 (Sigma), and 1.2 mL

Table 2
Feed solutions (500 mL) for controls of glucose concentrations and pH values

Solution no.	KH_2PO_4	K_2HPO_4	$NaHCO_3$	Glucose (g)	Volume (mL) & pH value
1	–	–	–	50–300	500 & 5.9
2	13.6 g	–	–	–	500 & 4.5
3	–	17.4 g	–	–	500 & 8.8
4			42.0		500 & 8.5
5	13.6 g	–	–	50–300	500 & 4.5
6	–	17.4 g	–	50–300	500 & 8.8
7	–	–	16.8 g	50–300	500 & 8.5
8	1.9	6.3	8.4	50–300	500 & 7.5

concentrated H_2SO_4. The pH is adjusted to 7.0 using concentrated NH_4OH prior to autoclaving. Before inoculation, the following supplements are added to the fermentation medium (1 L): 30 g D-glucose, 0.24 g $MgSO_4$, and trace minerals including 0.0037 g $(NH_4)_6(Mo_7O_{24})\cdot 4H_2O$, 0.0029 g $ZnSO_4\cdot 7H_2O$, 0.0247 g H_3BO_4, 0.0025 g $CuSO_4\cdot 5H_2O$, and 0.0158 g $MnCl_2\cdot 4H_2O$ (18).

(b) A typical yeast culture medium (1 L) for microfermentation: 20–50 g D-glucose, 5 g yeast extract, 0–1.7 g yeast nitrogen base without amino acids and $(NH_4)_2SO_4$, 1 g urea, 6 g KH_2PO_4, 2 g Na_2HPO_4, 0–1.5 g $MgSO_4$, and 0.5 g antifoam 204 (Sigma) (see Note 3).

2.3. Feed Solutions for Controls of Carbon-Source Concentrations and pH Values

Here, we use glucose as an example since it is the most used carbon source. Some typical feed solutions for controls of glucose concentrations and pH values are shown in Table 2. All solutions are sterilized by filtering through 0.22 μm membranes (see Notes 4–6).

1. Control of only glucose concentrations: feed solution 1 (Table 2).
2. Control of only pH values: buffer solutions 2, 3, 4 (Table 2).
3. Control of both glucose concentrations and pH values: feed buffers 5, 6, 7, and 8 (Table 2).

3. Methods

3.1. Setup the Microfermentor System

1. Prepare house gas lines or gas cylinders with pressure meters and regulation valves (Air—90 psi, Air—10–30 psi, O_2—10–30 psi, N_2—10–30 psi, and CO_2—10–30 psi) (see Note 7).

2. For the micro-24 bioreactor processes requiring strict pH control, prepare the liquid NH_3 vessel for generating NH_3 gas as base.
3. Choose the right gas input for different pO_2 and pH control requirements and connect the right gas lines based on the chose gas configuration (see Note 8).
4. Connect the microbioreactor to the computer, through which the temperature, pO_2, and pH values in each individual reactor are monitored and controlled.
5. Calibrate pH offsets for a standard reactor cassette by following the manual provided by the manufacturer. A buffer solution or medium with a target pH control value within 6–8 should be prepared and loaded to all reactors for the calibration (see Note 9).

3.2. Prepare Glycerol Stock Vials (see Note 2)

1. Pick a single colony of a strain grown on a culture plate to inoculate a flask culture.
2. Mix the culture with sterilized glycerol solution (20–50%, v/v) at a ratio of 1:1 when cells in the culture are in log phase and reaches an OD600 of 1–5.
3. Split the culture-glycerol mixture into small portions (1–2 mL) and put in glycerol stock vials.
4. Freeze all the vials immediately in liquid nitrogen and store them in a −80°C freezer.

3.3. Prepare Seed Cultures (see Note 2)

1. Prepare 10-mL sterile culture tubes with appropriate caps for each strain to be tested in microreactor system.
2. Add 2–5 mL culture medium to each tube and cover each tube with an appropriate cap suitable for corresponding oxygen requirement.
3. If the strains to be run in microfermentors are stored as glycerol vials in −80°C freezer, take the vials out and thaw them at room temperature for about 15 min.
4. Take 5–10 μL thawed strain solution to inoculate each culture tube, then move to step 6.
5. If the strains are ready as single colonies on plates, pick a single colony to inoculate each culture tube.
6. Grow the microbial cells in culture tubes for 12–48 h until the cell OD600 reaches 1–5 or any other target levels required by specific strains and fermentation processes. The tube cultures are used to inoculate microfermentors.

3.4. Start a Microfermentor Run

1. Turn on the micro-24 bioreactor system including the computer connected.
2. In a biological laminar flow hood, preload 3–5 mL sterilized culture medium for each micro-24 reactor.

3. Cover all reactors with appropriate caps based on requirements for dissolved oxygen control levels, and then keep the reactor cassette in hood at room temperature for about 30 min.
4. Inoculate each reactor with 0.05–0.5 mL seed culture from Subheading 3.3, step 6.
5. Start the microreactor control software from the connected computer.
6. Open the configuration panel; input the data log rate (1–15 min) for data recording frequency.
7. Set the target temperature control value for each individual reactor, set the environment temperature 2–5°C below the lowest target temperature of the individual reactors (see Note 10).
8. For aerobic fermentation process, set rotation speed = 500–800 rpm; for anaerobic process, set rotation speed at a lower rate (e.g., 400–600 rpm) but enough for a good mixing in all reactors.
9. Choose the right gas configuration based on the requirements for dissolved oxygen and pH controls (see Note 8).
10. Choose the right dissolved oxygen (DO) control mode. For aerobic fermentation process, turn on DO control and set the target pO_2 value for each individual reactor, or just choose the flow mode and set the target air/O_2 flow rate for each individual reactor (see Note 11).
11. If there is no requirement for pH control or the pH is controlled only by feeding phosphate and/or bicarbonate buffer solution, then turn off the pH control (see Notes 12 and 13). If the pH is required for a control, turn on the control mode and set the target pH value for each individual reactor (see Note 14).
12. Clamp the reactor cassette into the microreactor console.
13. Save the configuration and exit the panel.
14. Start a run and input the running information (e.g., title of the run, strain names, fermentation medium and process information, etc.).

3.5. Maintain a Microfermentor Run

1. After starting a new run in the computer, click T, pH, and DO in the active control window to monitor microreactor status. Also, click any of the individual reactors to view all *T*, pH, and pO_2 process curves.
2. Every a few hours, take 0.01–0.10 mL sample from each reactor to measure growth OD and carbon-source (e.g., glucose) concentrations or for any other purpose. To take samples:
   (a) Pause the run through the control panel in the computer.
   (b) Unclamp the reactor cassette and put it in a laminar-flow bio-hood.

(c) Pipette 0.01–0.10 mL sample from each reactor and put into a 1.5-mL micro-vial in order.

(d) Reclamp the cassette into the microreactor console.

(e) Resume the microfermentor run in the computer.

3. After taking samples, add dd (distilled and deionized) H_2O to each sample to achieve 0.5–1.0 mL total volume. The diluted samples can then be measured for growth OD, carbon-source concentrations, and/or any other off-line measurements using standard methods (UV-Vis, YSI 2700, HPLC, GC, etc., see Note 15).

4. Every few hours or days, feed carbon-source solution, pH buffer solution, or a mixture of both (see Subheading 2.3) based on on-line pH and pO_2 process curves, off-line measurements from step 3, or empirical data (see Note 16). To feed carbon source and/or pH control buffer:

(a) Pause the run through the control panel in the computer.

(b) Unclamp the reactor cassette and put it in a laminar-flow bio-hood.

(c) Add 0.1–2.0 carbon-source feed or buffer solution (see Table 2).

(d) Reclamp the cassette into the microreactor console.

(e) Resume the microfermentor run in the computer.

3.6. Shutdown a Run and Analyze Fermentation Results

1. To shutdown a microfermentor run:

(a) Pause the run through the control panel in computer.

(b) Unclamp the reactor cassette and put it in a laminar-flow bio-hood.

(c) Shutdown the run in the computer and input the run information.

2. In a laminar-flow bio-hood, using 5–10 mL pipette carefully measure the final fermentation volume of each reactor and transfer all fermentation broth to a 15-mL sterile centrifuge tube.

3. Record the final volume, all the sample volume, and feed volume for each reactor during the run for mass balance and yield analysis.

4. For intracellular product analyses, take 1 mL or appropriate amount of each reactor's sample and save it in a refrigerator or freezer for further analyses.

5. For residual carbon source, medium components, extracellular product, and/or by-product analyses, take 1 mL or appropriate amount of each reactor's sample, spin down the cells by

centrifugation for 10 min at 4,000–8,000 × *g*, and save the supernatant in a refrigerator or freezer for further analyses.

6. For dry cell weight analysis:
   (a) Take all the rest samples in the centrifuge tubes and record volumes.
   (b) Harvest cells by centrifugation for 10 min at 4,000–8,000 × *g*.
   (c) Pour off supernatant and resuspend in fresh dd water.
   (d) Repeat the centrifugation and wash cells twice.
   (e) Resuspend cells in dd water.
   (f) Prepare aluminum weighing boats and preweigh them.
   (g) Pour the washed cells into a preweighed aluminum weighing boat.
   (h) Rinse the tube twice with 1 mL of dd water and pour the washings into the weigh boat.
   (i) Dry the boat and the cells at approximately 70–80°C under vacuum conditions for more than 12 h or until a constant weight is achieved.
   (j) Weigh the boat plus dry cells and record the combined weight.
   (k) Dry cell weight (g/L) will be calculated by:
   DCW (g/L) = [weight of dried cells plus boat (g)–predetermined weight of boat (g)]/volume of the rest sample in a tube from step 1 (L).
7. Collect all the measured OD, dry cell weight, residual carbon-source concentrations, final volumes, total samples and feed volumes, product/by-product titers, rates, and yields, and the process data recorded in the computer to evaluate the strains and process conditions.

4. Notes

1. There are many microfermentors available on the markets (see Table 1). The capacity/size of each reactor ranges from 5 μL to 100 mL. Generally, the smaller the reactor size, the more reactors a system can have, and the more experiments the system can handle at one time. However, some fermentation processes need more samples to analyze information such as OD600, DCW (dry cell weight), residual glucose, intracellular product titer, extracellular product titer, etc. This requires a microfermentor system with bigger reactor size. For example, to measure DCW, 2–5 mL broth is required for an accurate

measurement. Therefore, a microfermentor system with about 5-mL or more working volume for each reactor should be considered for this purpose.

2. The experimental steps for glycerol stock preparation are dependent on strains used and slightly vary in labs with different experience and requirements.
3. The maximum cell density is dependent on the available nitrogen in fermentation medium. The yeast culture medium (1 L) containing 5 g yeast extract and 1 g urea is good to make a final DCW of 10–20 g/L, which is in a right range for the microfermentor system to maintain the dissolved oxygen (pO_2) level. Urea is better than ammonium salts (e.g., $(NH_4)_2SO_4$) for the microfermentor system to maintain the pH value within 5.5–7.5, especially in the growth phase.
4. The feed solution for controls of glucose concentrations contains 100–600 g/L glucose. This can be adjusted based on how much and how frequently the feeding is required in the fermentation process. More frequent feeding operation requires more concentrated glucose feed solutions.
5. The feed solution for controls of pH values contains phosphate and bicarbonate. The concentration of each can be adjusted based on the need of the fermentation, i.e., how much organic acid is accumulated and need to be neutralized.
6. In general, phosphate buffer is only efficient for pH control of processes with low cell density (e.g., dry cell weight < 3 g/L in typical yeast fermentation), very short fermentation time (e.g., <48 h), or low organic acid production. With bicarbonate in the buffer, the pH control can be applied to processes with much higher cell density (e.g., 20 g/L dry cell weight in typical yeast fermentation), longer fermentation time (up to a week), or higher organic acid production.
7. House gases or gas cylinders for microfermentor setup: (1) 90 psi Air is to create vacuum to fix the reactor cassette; (2) 10–30 psi Air/O_2 is to supply O_2 for pO_2 control; (3) 10–30 psi CO_2 is used as acid for pH control; and (4) 10–30 psi N_2 is used for anaerobic processes and/or to flush out unnecessary O_2, CO_2, or NH_3 existing in reactors.
8. There are three gas input options (see Table 3): (1) standard mode for aerobic fermentation processes with Air/O_2 for pO_2 control and CO_2 as acid and NH_3 as base for pH control, (2) microaerobic mode for low O_2-requirement processes with Air, O_2, and N_2 for various pO_2 adjustments, and (3) constant flow mode for aerobic fermentation processes with only CO_2 gas as acid for pH control.
9. For standard reactor cassette, the temperature and dissolved oxygen sensors do not need calibrations to get accurate

Table 3
Three gas input modes for different pO_2 and pH control requirements

Gas input	Standard mode	Microaerobic	Constant flow
Gas 1	CO_2	N_2/CO_2	CO_2
Gas 2	NH_3	NH_3/N_2	N_2
Gas 3	Air/O_2	Air/O_2	Air/O_2

measurements. However, the slope of all pH sensors needs to be adjusted based on cassette lot number and the target pH control levels. The pH offset of each reactor should also be individually calibrated by using buffer or medium with known pH values predetermined by any other portable pH meters.

10. The Micro-24 Bioreactor system uses both heating and air cooling to maintain each reactor's temperature. Each reactor has an individual heating unit while all the reactors share the same environmental air flow for cooling purpose. In general, the reactors located in edge zones or near the entrance of the environmental air flow suffer more air cooling and require more heating to maintain the same target temperature. For strain screening process, temperature for all individual reactors is set the same. For temperature optimization process, the difference of the set points of any adjacent reactors/wells should not be more than 2°C to avoid affecting each other. To make the best balance of cooling and heating for all reactors, higher temperatures should be set for the reactors in center zones. The lowest temperature should be set for those in the first column (Well A1, B1, C1, and D1) facing the environmental air-blowing entrance. Also, the environmental air's temperature should be at least 2°C below the lowest reactor temperature. See Fig. 2 for an example of a temperature optimization experiment using the micro-24 bioreactor system.
11. For strain screening process, pO_2 value for all individual reactors is set the same. For pO_2 optimization process, the pO_2 value for each individual reactor can be set and controlled independently.
12. When the pH control is off, for strain screening process use always the same phosphate and bicarbonate buffer for all reactors. For pH optimization process, each reactor's phosphate and bicarbonate buffer needs to be individually prepared with target pH values and fed independently (see also Subheading 2.3).
13. There is not a perfect pH control available for most current microfermentor systems, which use optical sensors based on

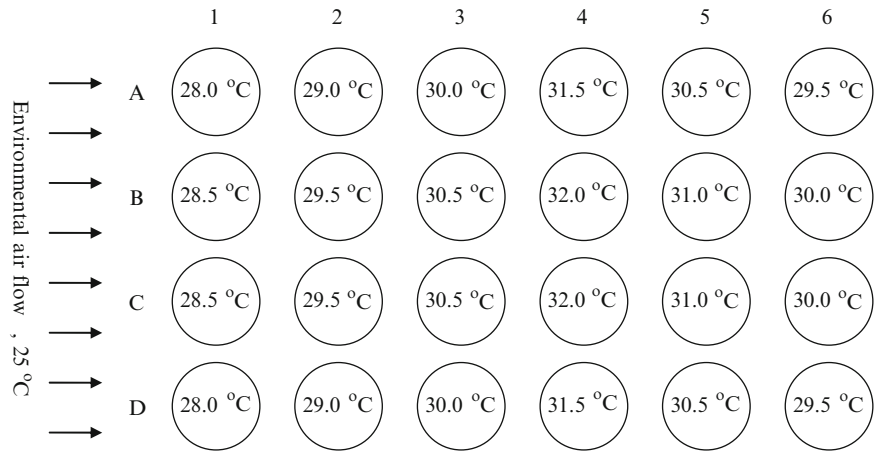


Fig. 2. An example of a temperature optimization experiment using the micro-24 bioreactor system.

fluorescence sensor spots or "optodes" and solid-state, ion-sensitive, field-effect transistor (ISFET) pH sensor chips (2). Unlike most pH probes for bench-scale fermentor with a wide range of pH measurements (between 2 and 10), the pH sensors for microfermentors measure pH values in a much narrower range (e.g., pH 6–8). In addition, it may take hours for the system to get stable and precise pH readings, which make it a big challenge for a good pH control. For this reason, using phosphate and bicarbonate buffer solution may also be a good option for processes without a good pH feedback control or processes with some acid/base restrictions (e.g., processes requiring for ammonium limitation).

14. When the pH control is on, for strain screening process the pH value for all individual reactors is set the same. For pH optimization process, the pH value for each individual reactor can be set and controlled independently.
15. When samples are taken to measure growth OD and/or any other off-line parameters, each sample volume is about 0.10 mL for early growth-stage (e.g., OD600 < 10) and 0.01–0.10 mL for late growth-stage. Do not discard the samples after the OD measurements since they can be reused for measuring the concentrations of the carbon source, any other detectable residual medium components, or accumulated products.
16. The continuous liquid feed operation (e.g., for glucose, acid, and base) is still difficult or not available for most microfermentor systems on the market; this makes a challenge for carbon-source feed and/or pH control. The Micro-24 Bioreactor system uses CO_2 and NH_3 gas for pH control, but this is not appropriate for the fermentation processes either sensitive to

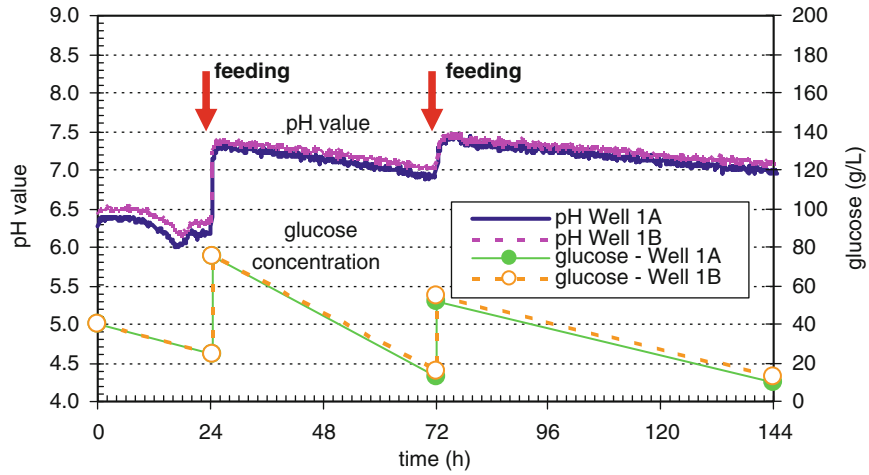


Fig. 3. Time course of the pH values and residual glucose concentrations of two duplicate yeast fermentation runs in the Micro-24 Bioreactor. The fermentation achieved a cell density of 10 g/L. The target pH values were 6.2 for 0–24 h and 7.2 for 24–144 h. The pH value in the first 24 h was maintained within 6.0–6.5 by the phosphate contained in the fermentation medium. The residual glucose concentrations were controlled within 10–80 g/L and the pH value was maintained within 7.0–7.5 after 24 h by feeding two shots (1.5 mL/shot) of the No. 8 solution containing 3.8 g/L KH_2PO_4, 12.6 g/L K_2HPO_4, 16.8 g/L $NaHCO_3$, and 160 g/L glucose, as shown in Table 2.

CO_2 gas or requiring nitrogen limitation. Here, we introduce a multi-shot feed approach to resolve the glucose feed and pH control issues. For glucose feeding, the prepared solution is fed into a reactor in multiple shots; and the feeding timing and amount are dependent on off-line measured glucose concentrations. For the pH control, bicarbonate and/or phosphate buffer is also fed into a reactor in multiple shots; and the feeding timing and amount are dependent on on-line pH values. In many fermentation processes, glucose feeding and pH control can be combined to one feeding operation if the feed solution contains appropriate amount of glucose, bicarbonate, and phosphate (e.g., feed solution 5–8 in Table 2). By multi-shot feeding certain amount of a well-designed glucose-bicarbonate-phosphate solution, the glucose concentration and pH value can be maintained within acceptable ranges. Figure 3 shows well-controlled pH and glucose process curves in a yeast fermentation with a two-shot feeding strategy.

Acknowledgments

The author would like to thank Dr. Bjorn D. Tyreus from DuPont Central Research and Development for his review and suggestions.

Reference

1. Michael J. Waites, Neil L. Morgan, John S. Rockey, and Gary Higton (2001) Industrial microbiology: an introduction. Wiley-Blackwell.
2. Schäpper D., Alam M.N.H.Z., Szita N., and Lantz A.E., and Gernaey K.V. (2009) Application of microbioreactors in fermentation process development: a review. *Anal Bioanal Chem* **395**, 679–695.
3. Weuster-Botz D., Hekmat D., Puskeiler R., and Franco-Lara E. (2007) Enabling Technologies: Fermentation and Downstream Processing. *Adv Biochem Eng Biotechnol* **105**, 205–247.
4. Kumar S., Wittmann C., and Heinzle E. (2004) Minibioreactors. *Biotechnology Letters* **26,** 1–10.
5. Jonathan I Betts and Frank Baganz (2006) Miniature bioreactors: current practices and future opportunities. *Microbial Cell Factories* 5, 21.
6. Fernandes P. and Cabral J.M.S. (2006) Microlitre/millilitre shaken bioreactors in fermentative and biotransformation processes—A review. *Biocatal Biotransformation* **24**, 237–252.
7. Büchs J. (2001) Introduction to advantages and problems of shaken cultures. *Biochem Eng J* **7**, 91–98.
8. Anderlei T., Büchs J. (2001) Device for sterile online measurement of the oxygen transfer rate in shaking flasks. *Biochem Eng J* **7**, 157–162.
9. Wittmann C., Kim H.M., John G., Heinzle E. (2003) Characterisation and application of an optical sensor for quantification of dissolved oxygen in shake-flasks. *Biotechnol Lett* **25**, 377–380.
10. Danielson P.B., Büchs J., Stockmann C., Fogleman J.C. (2004) Maximizing cell densities in miniprep-scale cultures with H15 medium and improved oxygen transfer. *Biochem Eng J* **17**, 175–180.
11. Stockmann C., Losen M., Dahlems U., Knocke C., Gellissen G., Büchs J. (2003) Effect of oxygen supply on passaging, stabilizing and screening of recombinant Hansenula polymorpha production strains in test tube cultures. *FEMS Yeast Res* **4**, 195–205.
12. Houston J.G., Banks M. (1997) The chemical-biological interface: developments in automated and miniaturised screening technology. *Curr Opin Biotechnol* **8**, 734–740.
13. Duetz W.A., Ruedi L., Hermann R., O'Connor K., Büchs J., Witholt B. (2000) Methods for intense aeration, growth, storage and replication of bacterial strains in microtiter plates. *Appl Environ Microbiol* **66**, 2641–2646.
14. Micro-24 MicroReactor System. See http://www.pall.com/biopharm_52961.asp.
15. Doig S.D., Diep A., Baganz F. (2005) Characterisation of a novel miniaturized bubble column bioreactor for high throughput cell cultivation. *Biochem Eng J* **23**, 97–105.
16. Betts J.I., Doig S.D., Baganz F. (2006) The characterization and application of a miniature 10 ml stirred-tank bioreactor, showing scale-down equivalence with a conventional 7L reactor. *Biotechnol Prog* **22**, 681–688.
17. Gilla N.K., Appletonb M., Baganza F., and Lye G.J. (2008) Design and characterisation of a miniature stirred bioreactor system for parallel microbial fermentations. *Biochem Eng J* **39**, 164–176.
18. Xie D., Shao Z., Achkar J., Zha W., Frost J.W., and Zhao H. (2006) Microbial synthesis of Triacetic Acid Lactone. *Biotechnol Bioeng* **93**, 727–736.
19. Funke M., Buchenauer A., Schnakenberg U., Mokwa W., Diederichs S., Mertens A., Muller C., Kensy F., Buchs J. (2010) Microfluidic BioLector—Microfluidic Bioprocess Control in Microtiter Plates. *Biotechnol Bioeng* 107, 497–505.
20. Amanullah A., Otero J.M., Mikola M., Hsu A., Zhang J., Aunins J., Schreyer HB., Hope J.A., Russo A.P. (2010) Novel micro-bioreactor high throughput technology for cell culture process development: Reproducibility and scalability assessment of fed-batch CHO cultures. *Biotechnol Bioeng* **106**, 57–67.
21. Cellstation high throughput bioreactors. See http://www.fluorometrix.com.
22. DAS GIP parallel system for microbial fermentation in process development. See http://www.dasgip.com.

Chapter 15

Rapid Strain Evaluation Using Dynamic DO-Stat Fed-Batch Fermentation Under Scale-Down Conditions

Jun Sun

Abstract

While large amount of strains can be quickly generated via metabolic engineering, the speed/efficiency of evaluating each strain becomes the bottleneck in the process from strain development to final production. In this chapter, a method is introduced to rapidly evaluate strain performance in fed-batch fermentation mode by using dynamic dissolved oxygen stat feed back control with no additional advanced online measurement. In addition, a scale-down feature is integrated in the method to mimic the limitation of oxygen transfer in large-scale vessels, so that strains can be evaluated under the conditions close to that in large-scale bioreactors. The method has been implemented in several commercial standard benchtop scale fermentation systems with different fermentation control software.

Key words: Strain evaluation, Fed-batch fermentation, Dynamic DO-Stat control, Scale-down, Process development

1. Introduction

The advances in metabolic engineering have made it possible to generate significant amount of genetically engineered microbial strains and screen for a specific trait within a short period of time (1, 2). Thus, effective evaluation of the performance of strains under conditions close to the large-scale fermentor is a challenge task for bioprocess engineer. Today most of microbiology labs still use shake flasks or microwell plates to evaluate the stains preliminarily to get end point data on strain performance. The recent advances in minibioreactors have provided the tools to evaluate strains with dynamic dataset due to the integration of online sensors/controllers into minibioreactors, such as dissolved oxygen, pH and optical cell density, etc (3, 4). These can improve throughput significantly to

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_15,

reduce the number of strains for comprehensive evaluation in well controlled bioreactors (5, 6). However, due to the limitations of shake flasks/microwell plates and minibioreactors, a considerable amount of strains will have to be selected from preliminary evaluation step to be further tested in fully instrumented fermentors.

When evaluating strains in benchtop fermentors, the general approach is to perform fermentations for all strains under same conditions to compare fermentation performance. However, this approach may result in missing of potential good strains because the cultivation conditions used for fermentation might not be optimal for certain strains. In addition, if the experimental design at this stage doesn't take into account the conditions in large-scale fermentors, strains selected at this stage might not be good candidates for scale-up. Thus, it is very critical to design the fermentation experiment to evaluate strains at the conditions not only to enable each strain to deliver its near optimal performance but also to keep the major process parameters close to that in large-scale fermentors, such as oxygen transfer rate (OTR) limit and oscillation of pH, dissolved oxygen (DO) and substrate concentrations, etc. Paradoxically at this stage there is not much prior knowledge available for each strain to help design the fermentation experiments to deliver best performance for comparison. For batch fermentation process, the lack of prior knowledge is not as bad as for fed-batch fermentation process in which the substrate concentration has to be controlled at certain level to promote cell growth and production formation. In another hand, the fed-batch fermentation is still the most widely used fermentation mode to achieve high product titer/yield even though some studies have demonstrated the possibility of replacing fed-batch fermentation with batch fermentation after genetic modification of microbial strains (7, 8).

When designing a fed-batch fermentation, the most intriguing variable needs to be defined is the feed profile. Various methods have been described to determine and optimize the feed profile for specific strain and process as summarized in (9). A commonly used strategy is to use a predefined feed profile based on a designated specific growth rate for all strains. However, because of the lack of prior knowledge of the strains, the designated specific growth rate is not always optimal for each strain. Consequently, some good candidates may get excluded because the predefined feed profile may not be suitable for certain strains. Although some control strategies have been implemented to calibrate the predefined feed profile on-fly by probing the by-product formation based on DO response to feed pulse (10, 11), it is preferable to define feed profile during fermentation process for each strain to be evaluated.

In this chapter, a protocol for substrate limited fed-batch fermentation is introduced. This protocol is robust, universal (independent on strains, products, media, etc.), and easy to implement in standard commercial fermentation systems for fully automation

without additional advanced online measurements. A feature called oxygen uptake rate (OUR) clipping to mimic the limit of OTR in large-scale vessels is also implemented in this protocol, which makes it a powerful tool for rapid strain evaluation and process development under scale-down conditions.

2. Materials

2.1. Fermentation Systems

Most commercially available benchtop fermentation systems for microbial fermentation can be used for this method. The fermentation control unit should be able to be controlled remotely by fermentation supervisory control and data acquisition (SACDA) software. The method has been validated in commercial fermentation systems such as Bioflo series from New Brunswick Scientific (Edison, New Jersey, USA), laboratory fermentation systems from Applikon (Schiedam, Netherland) and BioStat series from Sartorius Stedim Biotech Gmbh (Goettingen, Germany), etc. For all the case study data shown in this chapter, the Biostat C 10L sterilize-in-place (SIP) fermentation system (Sartorius Stedim Biotech Gmbh) is used.

2.2. Fermentation Supervisory Control and Data Acquisition Software

The commercial fermentation supervisory control and data acquisition (SCADA) software with Visual Basic or other script languages programming capability can be used to implement this method. The method has been implemented in commercial fermentation SCADA software such as BioCommand from New Brunswick Scientific, BioXpert XP from Applikon and MFCS/win from Sartorius Stedim Biotech. For the case study data shown in this chapter, the Wonderware InTouch from Wonderware (Lake Forest, CA, USA) is used to implement the control algorithm.

2.3. Process Measurements

1. Dissolved oxygen (DO) level in the broth was measured by InPro6800 DO probe from Mettler Toledo (Bedford, MA, USA).
2. Glucose and acetate acid level in the fermentation broth were measured by HPLC method using Shodex Sugar SH1011 column in a Agilent 1200 HPLC instrument.
3. The component of exhausted gas from fermentors was analyzed by VG Prima dB process mass spectrometer from Thermo Fisher Scientific (Sugar Land, TX, USA).

2.4. Microorganisms

This method is suitable for most of bacteria and yeast under aerobic cultivation conditions. For all the case study data, *E. coli* MG1655 strain harboring plasmid for proprietary recombinant protein expression was used.

3. Methods

3.1. Dynamic DO-Stat Fed-Batch Control Algorithm Development

Many fed-batch fermentation experiments need to keep substrate limited at fed-batch stage to minimize by-product formation. For example, the excess glucose level can divert the carbon flux into acetate production, thus inhibit the cell growth and protein expression when cultivating *E. coli* for recombinant protein expression (12, 13). When choosing a substrate limited feeding strategy for strain evaluation, the feeding strategy should be (1) universal (independent on strains, products, media, etc.); (2) robust; (3) requiring as less online measurements as possible; (4) easy to scale-up; (5) easy to implement and use, preferably with fully automation; and (6) requiring very little prior information regarding cell growth properties. Based on the above criteria, DO-stat and pH-stat are among the best choices. DO and pH probes are cheap and relatively reliable. Most of modern fermentation systems have the DO/pH controlled substrate feed program built in the control unit. However, the pH-stat has several disadvantages such as slow response to the feed addition, and being medium dependent. So, the DO-stat is a great choice for initial substrate limited fed-batch fermentation for strain evaluation.

The commonly used DO-stat control feeding strategy for fed-batch fermentations involves adding the substrate when the dissolved oxygen (DO) level rises above the setpoint due to depletion of the substrate. When this strategy is used for certain fermentations, such as recombinant protein expression or when the product is toxic to the cells, it can cause overfeeding of substrate as demonstrated by the example in Fig. 1.

One approach to resolve the overfeeding problem associated with DO-stat is to feed the substrate based on the patterns of DO change. As shown in Fig. 2, the increase of DO can be caused by (a) substrate depletion; (b) cell metabolic slowing down or product toxicity; and (c) DO probe noise. The latter two cases will lead to substrate overfeeding when using standard DO-stat control strategy. Fortunately, it is possible to distinguish the DO change patterns and design software controller to feed substrate only in response to substrate depletion.

To design a control algorithm to distinguish the DO change patterns, the following approach is taken:

1. Initialize the DO setpoint (DO_{set}) and a 10-element DO stack [DO1, DO2, DO3, …, DO10] to initial DO value before inoculation which is generally 100% for most benchtop fermentors if no pressure is set in the vessel (see Note 1). Set the final DO setpoint (DO_{fset}) to be the lowest acceptable DO value (This is the DO setpoint in conventional DO-stat control strategy). Set the high DO threshold (DO_{ht}) and the low DO threshold (DO_{lt}) which should be larger than the noise of DO probe (see Note 2).

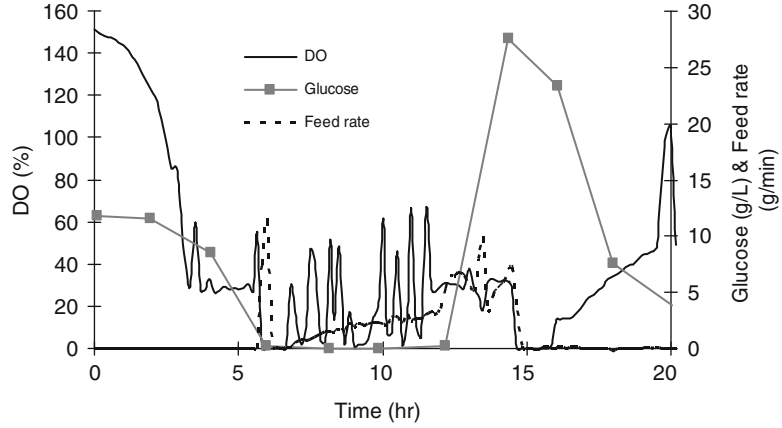


Fig. 1. An example of fed-batch fermentation using DO-stat to control the feed rate. In this fermentation, a recombinant protein was expressed in *E. coli* cell. The feed solution is 50% (w/w) glucose. The protein expression was induced at 8.5 h. The DO setpoint is 30%. A built-in DO controlled feed program in a Sartorius Biostat C control unit (Sartorius, Germany) was used to control feed rate of the glucose feed solution by keeping the DO at 30%. However, after induction at 8.5 h, the DO stayed higher than the setpoint due to the cell metabolic slowing down, subsequently caused overfeed of glucose. The glucose feed had to be shut down manually to prevent further detrimental impact of high glucose level on the protein expression.

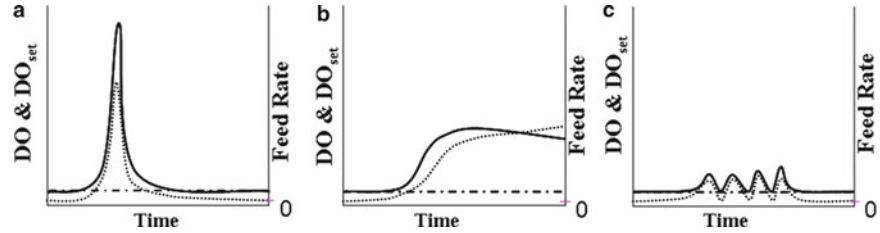


Fig. 2. DO patterns in response to (**a**) substrate depletion which leads to substrate feeding; (**b**) cell metabolic slowing down or product toxicity which leads to substrate overfeeding; (**c**) DO probe signal noise which also leads to substrate overfeeding. *Solid line*: DO profile. *Dotted line*: feed rate. *Dash-dotted line*: DO setpoint.

2. Every 10 s, insert the current DO value into the DO stack and drop the oldest DO value from the stack. Change DO_{set} to the minimal value in the DO stack.
3. If the minimal value of DO stack $< DO_{fset}$, then set $DO_{set} = DO_{fset}$.
4. If $DO > (DO_{set} + DO_{ht})$, initiate substrate feeding, with the feed rate controlled by a P-I-D control algorithm based on $(DO - DO_{set})$. A proportional control algorithm is usually enough to control the feed rate.
5. If $DO < (DO_{fset} - DO_{lt})$, increase DO through some other mechanisms, such as increasing the agitation speed, the air flow rate, or the oxygen enrichment.

The above strategy is illustrated in Fig. 3. The feed will be triggered if DO is higher than the DO setpoint plus the DO high threshold value. When substrate is exhausted, DO will increase dramatically as shown in Fig. 3a and the feed will start since

$OD > (OD_{set} + OD_{ht})$. In case of cell metabolic slowing down or product toxicity as shown in Fig. 3b, DO will increase slowly and the DO setpoint will increase accordingly to avoid substrate overfeeding. For the DO probe noisy issue, the parameter of DO_{ht} which is higher that probe signal noise will prevent substrate overfeeding as shown in Fig. 3c.

A real-time example is shown in Fig. 4. From this example, it can be observed that the feed pump started only when a DO spike occurred. After 17.8 h, the cell metabolism slowed down as indicated by the baseline increase of DO profile. However, the control strategy was still able to recognize the pattern of DO changes and initiated the feed pump only in the period of DO spike.

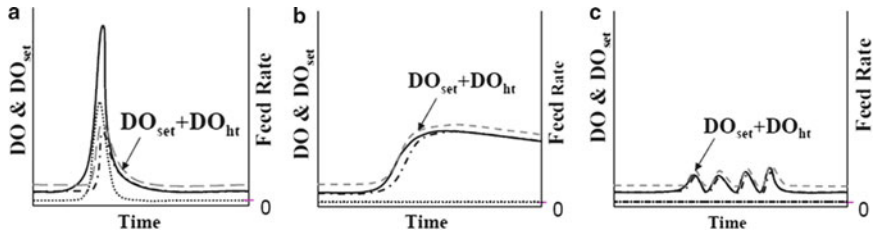


Fig. 3. Dynamic DO-stat to avoid substrate overfeeding. The DO setpoint (DO_{set}) will follow the DO profile with 100 s of delay. When the DO is higher than DO setpoint plus a DO high threshold value (DO_{ht}), the feed will be initiated. (**a**) Substrate depletion leads to dramatic increase of DO, subsequently initiates feeding as $DO > (DO_{set}+DO_{ht})$. (**b**) Cell metabolism slowing down or product toxicity leads to gradual DO increase, but overfeeding is avoided because the DO_{set} follows the DO profile and $DO < (DO_{set}+DO_{ht})$. (**c**) DO probe signal noise will not cause overfeeding because DO_{ht} is selected to be higher than signal noise to ensure $(DO_{set}+DO_{ht}) > DO$ for a noisy probe. *Solid line*: DO profile. *Dotted line*: feed rate. *Dash-dotted line*: DO setpoint. *Dashed line*: DO setpoint plus a DO high threshold value.

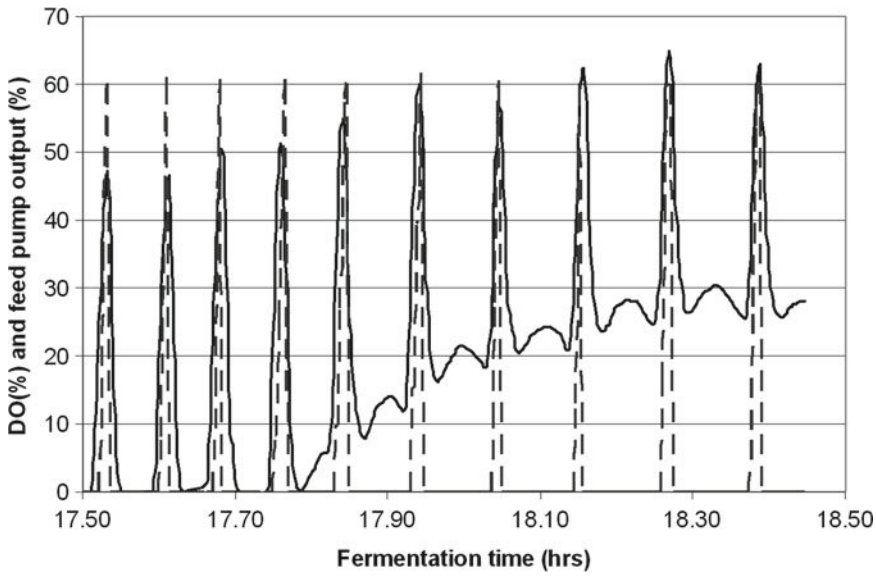


Fig. 4. A real-time example of dynamic DO-stat feed strategy. *Solid line*: DO (%). *Dashed line*: feed pump output (%). The glucose level was kept at zero at fed-batch stage (data not shown).

In order to fully utilize the dynamic DO profile to define the substrate feed, the DO control by agitation or air flow is designed as following to minimize the disturbance of DO profile by DO control loop:

1. If $DO < (DO_{fset} - DO_{lt})$, control DO by first ramping up agitation speed until the maximal agitation setpoint is reached, then ramping up air flow rate.
2. If $DO > (DO_{fset} + DO_{ht})$ for more than 10 min as an indication of cell metabolic slowing down or product toxicity, control DO by decreasing the air flow rate first until the minimal air flow rate setpoint is reached, then tuning down agitation speed (see Note 3).
3. If OUR clipping feature is enabled, the maximal setpoints for agitation and air flow will be dynamically reset to their current process values as soon as the OUR reaches its setpoint. The OUR can be calculated online based on off-gas component measurement detected by a gas analyzer (see Note 4).

3.2. Dynamic DO-Stat Fed-Batch Control Algorithm Implementation and Application

The above dynamic DO-stat control strategy has been successfully implemented using Visual Basic programming language or other scripting languages in most commercial fermentation SCADA software as described in Subheading 2. One example of the control program interface implemented using Wonderware InTouch software is shown in Fig. 5. To use this specific feed control program, follow the instruction as below:

1. Set the fermentor per experimental needs.
2. Set the DO control to "off" mode in the fermentation control unit. The DO level will be controlled by the control program.
3. Set stir, air flow, and substrate feed pump control to "auto" mode in the fermentation control unit.
4. Set the fermentation control unit to "remote" mode. This enables the control unit to accept the setpoints from the control program.
5. Determine proper values for coefficients of stir, air flow, and substrate pump in term of $(DO - DO_{set})$. These values are equipment and scale dependent (see Note 5).
6. Set the minimal DO setpoint, initial DO setpoint, DO_{lt}, DO_{ht}, minimal and maximal setpoints for stir and air flow, coefficients for stir, air flow, and substrate pump accordingly (see Note 6).
7. Check OUR clipping option and set OUR setpoint if OUR clipping experiment is needed.
8. If a pH-stat control feed is needed, check the option of "activate feeding controlled by pH" (see Note 7).
9. Click "Start/Reset" button to start the program after inoculating the fermentor. No human intervention is needed during whole fermentation process.

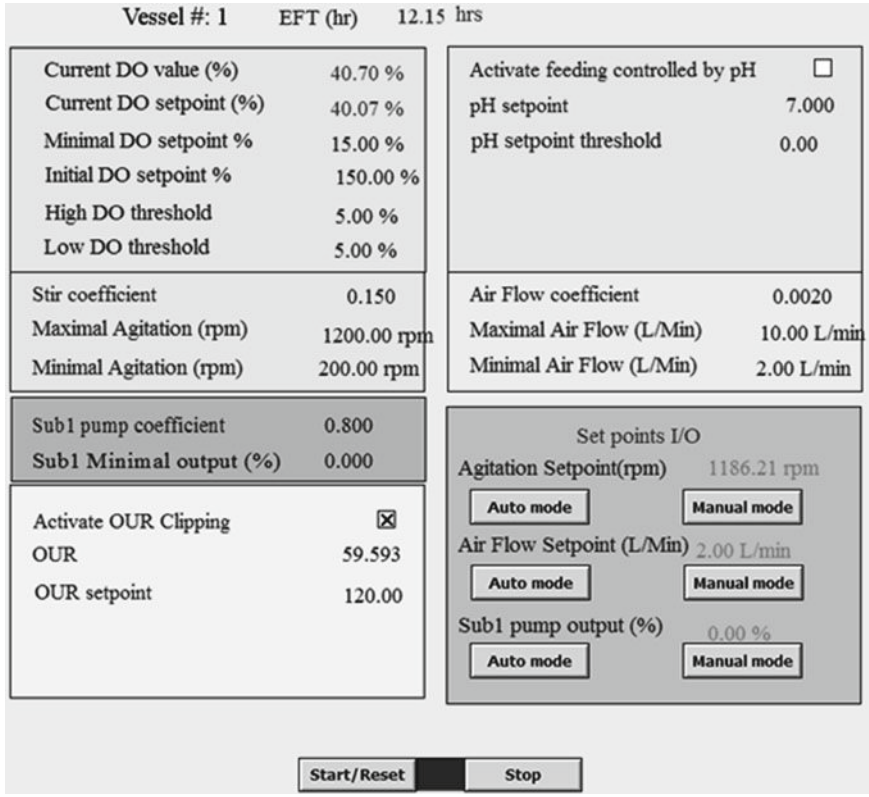


Fig. 5. Dynamic DO-Stat substrate feeding control program interface. The interface is developed in Wonderware InTouch Human Machine Interface.

10. Click the "Stop" button to end the program after the fermentation is finished.

One example of using above dynamic DO-stat for recombinant protein production in *E. coli* fermentation is shown in Fig. 6. After induction at 11 h, the cell metabolism slowed down and caused DO increase to above minimal DO setpoint (10%). But the control program was able to distinguish the DO increase pattern and feed the glucose accordingly to keep the glucose level limited. Another advantage of using this dynamic DO-stat is that the by-product can be controlled to a limited level for *E. coli* fermentation. As shown in Fig. 6, acetate accumulated during batch culture stage was assimilated by cells at fed-batch stage and the final acetate level was below 0.3 g/L.

When scaling-up an aerobic process, the limit of OTR in large-scale vessels is a key factor because the fermentation performance generally will be related to oxygen transfer capacity of fermentors. In small-scale fermentors, OUR can reach to as high as 300 mM/h. But in large-scale vessels, OUR is limited to about 100–160 mM/h with compressed air supply. So, it is very critical to include the OUR limited condition in experimental design at strain evaluation stage to select strains with best performance at conditions close to

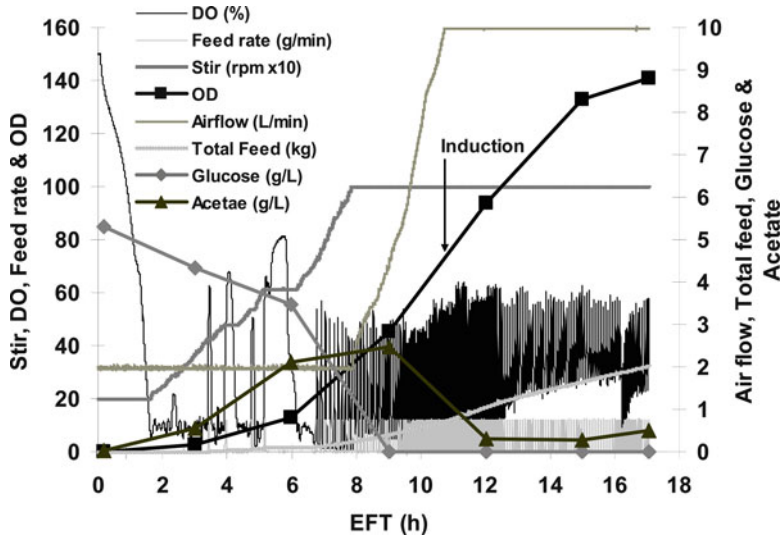


Fig. 6. An example of fed-batch fermentation using dynamic DO-stat to control the glucose feed rate. In this fermentation, a recombinant protein was expressed in *E. coli* cell. The feed solution is 50% (w/w) glucose. The minimal DO setpoint is 10%. The protein expression was induced at 11 h.

that in large-scale vessels. The dynamic DO-stat control strategy can easily integrate with OUR clipping strategy to meet the scale-down experiment. One example is shown in Fig. 7. Without OUR clipping, the *E. coli* culture can grow up to OD_{600} above 290 within 28 h with maximal OUR reaches to about 300 mM/h as illustrated in Fig. 7a. However, at OUR clipping conditions with maximal OUR set to 165 mM/h, the cells grow slower and reach to $OD_{600} \approx 220$ within 30 h as shown in Fig. 7b. Using dynamic DO-stat, the feed rate is adjusted automatically to adapt to the different OUR setpoints.

4. Notes

1. For pressurized vessel, the initial DO value in DO stack depends on the pressure setpoint. If the back pressure of the vessel is 0.5 bar as usually used for most of SIP fermentors, the initial DO value should be set to 150%.
2. For most DO probes, 5% is a reasonable number for both DO_{lt} and DO_{ht}. To prevent occasional overfeeding, DO_{ht} can be set to as high as 10%.
3. For most of processes, 10 min is long enough to distinguish the cause of DO level increase due to metabolism slowing down. But for certain microorganisms with slow growth rate, the time can be extended to 30 min without adverse impact on the performance.

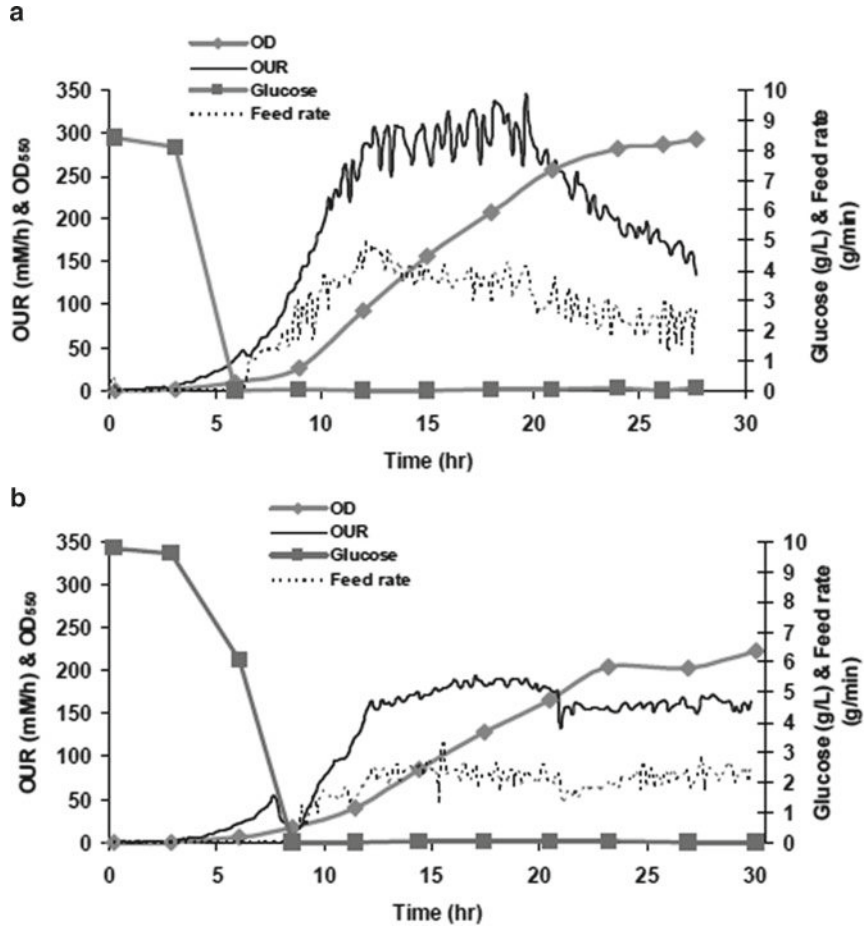


Fig. 7. Comparison of regular fermentation process with OUR clipping process. (**a**) Fed-batch fermentation without OUR clipping. (**b**) Fed-batch fermentation with OUR clipping. The OUR was set at 165 mM/h. *Diamond symbol*: OD_{550}; *square symbol*: glucose concentration (g/L); *solid line*: OUR (mM/h); *dashed line*: average feed rate with 10 min of interval (g/min).

4. If off-gas analyzer is available, OUR (mM/h) can be calculated based on online measurement as following:

$$\mathrm{OUR} = \frac{26.8 \times \mathrm{Airflow} \times (O_{2\mathrm{in}} - (N_{2\mathrm{out}} / N_{2\mathrm{in}}) \times O_{2\mathrm{out}})}{V_{\mathrm{broth}}},$$

where *Airflow* is inlet air flow rate (L/min) generally measured online by mass flow meter or controller, $O_{2\mathrm{in}}$ and $O_{2\mathrm{out}}$ are oxygen level (%) in inlet air and outlet gas, respectively; $N_{2\mathrm{in}}$ and $N_{2\mathrm{out}}$ are nitrogen level (%) in inlet air and outlet gas, respectively; V_{broth} is total broth volume which can be measured by loading cell if equipped or can be calculated by mass balance.

5. These coefficients are the *P* value of P-I-D control loop for stir speed, air flow rate, and feed pump in term of $(\mathrm{DO} - \mathrm{DO}_{\mathrm{sct}})$.

The values shown in Fig. 5 are suitable for 10 L Sartorius Biostat C fermentation system. When choosing the coefficient for feed pump, this value should be capped so that the maximal dosing of feed at each DO spike period (as shown in Fig. 4) will not cause the substrate accumulation. For example, the glucose level will be less than 1 g/L after each pulse of feeding.

6. The minimal feed pump rate is usually set to 0 for substrate limited fed-batch fermentation. However, this variable can be set to a minimal feed rate to prevent cell starvation for certain microorganisms.
7. It is not necessary to use both dynamic DO-stat and pH-stat in same run. However, the pH-stat controller can be used as a backup in case that the dynamic DO-stat cannot work because a DO probe failure occurs during a run. To use the pH-stat feed control, the pH setpoint threshold needs to be set to (a) a positive valve if substrate starvation will cause pH increase such as using sugar as substrate, (b) a negative value if substrate starvation will cause pH decrease such as using organic acid as substrate. The feed pump will start (a) when current pH is higher than pH setpoint plus the setpoint threshold if the threshold is positive, (b) or when current pH is less than pH setpoint plus the setpoint threshold if the threshold is negative.
8. Using this control strategy the substrate is fed in pulse mode instead of continuous feeding mode, subsequently cells will experience the oscillation of substrate concentration. This is another scale-down feature to simulate fluctuation of substrate concentration in large-scale vessels. The program has been successfully used to control fed-batch fermentations for various recombinant protein expression, enzyme production, and biomass production by bacteria and yeast. In conclusion, this dynamic DO-stat control strategy is a powerful tool for rapid strain evaluation and process development without prior knowledge of the strains. The feed profile generated by the program can be used as the baseline for further process optimization.

References

1. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., and Gill, R. T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides, *Nat Biotechnol* 28, 856–862.
2. Gill, R. T., Wildt, S., Yang, Y. T., Ziesman, S., and Stephanopoulos, G. (2002) Genome-wide screening for trait conferring genes using DNA microarrays, *Proc Natl Acad Sci USA* 99, 7033–7038.
3. Isett, K., George, H., Herber, W., and Amanullah, A. (2007) Twenty-four-well plate miniature bioreactor high-throughput system: Assessment for microbial cultivations, *Biotechnol Bioeng* 98, 1017–1028.
4. Huber, R., Ritter, D., Hering, T., Hillmer, A. K., Kensy, F., Müller, C., Wang, L., and Büchs, J. (2009) Robo-Lector – a novel platform for automated high-throughput cultivations in microtiter plates with high information content, *Microb Cell Fact* 8, 42.
5. Tang, Y. J., Laidlaw, D., Gani, K., and Keasling, J. D. (2006) Evaluation of the effects of various culture conditions on Cr(VI) reduction by

Shewanella oneidensis MR-1 in a novel high-throughput mini-bioreactor, *Biotechnol Bioeng* 95, 176–184.

6. Kensy, F., Engelbrecht, C., and Büchs, J. (2009) Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in *Escherichia coli* and *Hansenula polymorpha* fermentations, *Microb Cell Fact* 8, 68.
7. Ingo Knabben, Regestein, L., Marquering, F., Steinbusch, S., Lara, A. R., and Büchs, J. (2010) High cell-density processes in batch mode of a genetically engineered *Escherichia coli* strain with minimized overflow metabolism using a pressurized bioreactor, *J Biotech* 150, 73–79.
8. Lara, A. R., Caspeta, L., Gosset, G., Bolívar, F., and Ramírez, O. T. (2008) Utility of an *Escherichia coli* strain engineered in the substrate uptake system for improved culture performance at high glucose and cell concentrations: An alternative to fed-batch cultures, *Biotechnol Bioeng* 99, 893–901.
9. Lee, S. Y. (1996) High cell-density culture of *Escherichia coli*, in *Trends in Biotechnology*, pp 98–105. Elsevier Science Ltd.
10. Åkesson, M., Hagander, P., and Axelsson, J. P. (2001) Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding, *Biotechnol Bioeng* 73**,** 223–230.
11. Åkesson, M., Karlsson, E. N., Hagander, P., Axelsson, J. P., and Tocaj, A. (1999) On-line detection of acetate formation in *Escherichia coli* cultures using dissolved oxygen responses to feed transients, *Biotechnol Bioeng* 64**,** 590–598.
12. Luli, G. W., and Strohl, W. R. (1990) Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations., *App Environ Microbiol* 56, 1004.
13. Mey, M., Maeseneire, S., Soetaert, W., and Vandamme, E. (2007) Minimizing acetate formation in *E. coli* fermentations, *J Ind Microbiol Biotechnol* 34, 689–700.

Chapter 16

Preparation and Evaluation of Lignocellulosic Biomass Hydrolysates for Growth by Ethanologenic Yeasts

Ying Zha, Ronald Slomp, Johan van Groenestijn, and Peter J. Punt

Abstract

Lignocellulosic biomass is a potential feedstock for bioethanol production. Biomass hydrolysates, prepared with a procedure including pretreatment and hydrolysis, are considered to be used as fermentation media for microorganisms, such as yeast. During the hydrolysate preparation procedure, toxic compounds are released or formed which may inhibit the growth of the microorganism and thus the product formation. To study the effects of these compounds on fermentation performance, the production of various hydrolysates with diverse inhibitory effects is of importance. A platform of methods that generates hydrolysates through four different ways and tests their inhibitory effects using Bioscreen C Analyzer growth tests is described here. The four methods, based on concentrated acid, dilute acid, mild alkaline and alkaline/oxidative conditions, were used to prepare hydrolysates from six different biomass sources. The resulting 24 hydrolysates showed great diversity on growth rate in Bioscreen C Analyzer growth tests. The approach allows the prediction of a specific hydrolysate's performance and helps to select biomass type and hydrolysate preparation method for a specific production strain, or vice versa.

Key words: Lignocellulosic biomass hydrolysate, Bioscreen C analyzer, Ethanologenic yeast, Inhibitory effect, Growth rate

1. Introduction

Lignocellulosic biomass, such as sugar cane bagasse, wheat straw and willow wood, are potential feedstocks for bioethanol production. They are inexpensive, abundant and not competitive with world food resources. Generally, lignocellulosic biomass is composed of cellulose, hemicellulose and lignin. To use lignocellulosic biomass as fermentation feed, a pretreatment and a hydrolysis step need to be carried out, which breaks the structure of the biomass and releases the sugar monomers, respectively. The resulting mixture after removing the lignin fraction, the lignocellulosic biomass hydrolysate, is then used as fermentation medium.

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_16,

One of the drawbacks of the lignocellulosic biomass hydrolysate, when used as feedstock for bioethanol production, is its inhibitory effect on the growth of microorganisms (1, 2). The inhibitory effects are caused by the release of toxic compounds during the hydrolysate preparation process, since conditions like high temperature, high pressure, and acid/alkaline environment are used for the pretreatment. Growth inhibition mainly results in a longer lag-phase, a reduced growth rate and a lower ethanol productivity; and was found in both laboratory and industrial strains (3–5).

To restrict the inhibitory effect of the hydrolysate, identifying inhibitory compounds is a crucial step. In the last several decades, studies on hydrolysate inhibitory compounds have been conducted; the results showed that the potential inhibitors are aliphatic and aromatic acids, aldehydes and ketones from the degradation products of cellulose and hemicellulose, as well as phenolic compounds from lignin (6, 7). Among all the potential inhibitors, formic acid, acetic acid, furfural and hydroxymethylfurfural (HMF) were considered as the representatives, and were used to test the inhibition resistance of different microorganisms (8, 9). Heer et al reported the identification of furfural as a key inhibitor in the wheat and barley straw hydrolysates, using ethylacetate extraction GC-MS (10).

Due to the complexity and the variability of the lignocellulosic biomass hydrolysate composition, and the requirement of advanced tools, such as LC-MS, GC-MS and solid-phase microextraction (SPME), for chemical analysis, the identification of inhibitory compounds has not yet been carried out systematically. In our research, we aim to use a non-targeted metabolomics approach to elucidate the substrate inhibitory compounds. The first requirement of this approach is a high-quality experimental design (11, 12). Such an experimental design should include a variety of hydrolysates that hold diverse inhibitory effects on fermentation. It is expected that some inhibitory compounds are universal, while the others are more specific to a particular hydrolysate.

In our study, six types of biomass are selected and prepared with four different methods to test the inhibitory effects of the resulting 24 hydrolysates on different microorganisms. The procedures of the four hydrolysate preparation methods, namely, concentrated acid, dilute acid, mild alkaline and alkaline/oxidative method, are described here. In addition, we describe the procedure of the growth test using Bioscreen C Analyzer (13). The results show that there is a great diversity of inhibitory effects among the 24 hydrolysates, which is contributed by both biomass type and hydrolysate preparation method.

The method provides a platform for testing lignocellulosic biomass hydrolysates on various microorganisms, especially bacteria and yeasts. The results obtained with this analysis platform will help to select biomass type and hydrolysate preparation method for a specific production strain, or vice versa.

2. Materials

2.1. Lignocellulosic Biomass

1. Sugar cane bagasse is a kind gift from Zilor, Brazil.
2. Corn stover is from the University of Cape Town, South Africa.
3. Wheat straw, barley straw and willow wood are purchased from Oostwaardshoeve, the Netherlands.
4. Oak sawdust is from ESCO, the Netherlands, a wood-flooring supplier.
   All biomass is pre-dried at 80°C for 5 h when received, and stored at room temperature. All biomass, except oak sawdust, is ground to pieces with average length of 3 mm.

2.2. Hydrolysate Preparation

1. H_2SO_4, purity 95–97%, stored at room temperature.
2. NaOH, purity ≥98%, stored at room temperature.
3. $Ca(OH)_2$ purity ≥96%, stored at room temperature.
4. Acetic acid, purity 100%, stored at room temperature.
5. 30% H_2O_2, stored at 4°C.
6. Penicillin:streptomycin (pen:strep) solution: 10,000 unit penicillin and 10 mg streptomycin/mL. Used at 1:100 (v:v) dosage. Stored at –20°.
7. Enzyme cocktail Accellerase 1500 (a gift from GENENCOR): endoglucanase activity 2,200–2,800 CMC U/g; beta-glucosidase activity 525–775 pNPG U/g.
8. Rotation is carried out on a two-deck modular roller apparatus (348971-C) from Wheaton Roller.
9. All filtrations, unless otherwise stated, use Grade GF/D filter purchased from Whatman.
10. HACH LANGE SO_4^{2-} testing kit (LCK353).
11. Anion exchange is carried out with a MP-Cell from ElectroCell A/S (Tarm, Denmark), which includes two PVDF fluid distribution frames, four EPDM gaskets and two PVDF end frames. Anion exchange membranes are Neosepta AFN membranes from Eurodia Industrie SA, France. The ElectroCell equipment is assembled with stainless steel end plates and hardware, indicated as "EC" in Fig. 1. Liquid in both donor and acceptor compartment are pumped in a circuit over a glass bottle, a pump and the cell and back to the bottle again. The liquid volume of the circuit is 0.45 L. The flow direction over the two cells was counter current at a rate of 0.5 L/min.

2.3. Yeast Strains

1. *Saccharomyces cerevisiae* CEN.PK 113-7D is purchased from CBS Yeast Collection (CBS 8340).
2. Ethanol Red yeast is a gift from Fermentis, France.
3. Both strains are stored as 1 mL glycerol stocks at –80°C.

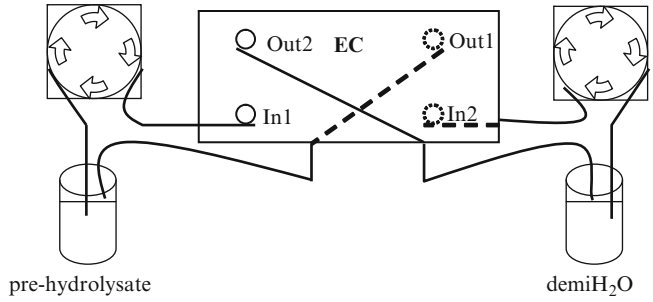


Fig. 1. Ion exchange membrane facility for neutralization.

2.4. Bioscreen C Analyzer Growth Study

1. Bioscreen C Analyzer, Labsystems OY, Helsinki, Finland.
2. In the Bioscreen C Analyzer, two honeycomb plates, each contains 100 wells, can be incubated at a preset temperature, and the growth of the microorganisms can be monitored by measuring optical density ($OD_{420–580}$) at a preset time interval, in our case, 15 min. The honeycomb plates can be shaken linearly at a preset speed.
3. The Bioscreen C Analyzer is connected to a computer, which records all the measured optical density values.
4. Mineral medium is prepared according to the composition as described by van Hoek et al. (14). Per liter: 5.0 g $(NH_4)_2SO_4$, 3.0 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 1 mL 1,000× trace elements solution, 1 mL 1,000× vitamin solution and 80 mL 250 g/L glucose solution.
   a. Per liter 1,000× trace element: 15.0 g EDTA, 4.5 g $ZnSO_4 \cdot 7H_2O$, 0.3 g $CoCl_2 \cdot 6H_2O$, 1.0 g $MnCl_2 \cdot 4H_2O$, 0.3 g $CuSO_4 \cdot 5H_2O$, 4.5 g $CaCl_2 \cdot 2H_2O$, 3.0 g $FeSO_4 \cdot 7H_2O$, 0.44 g $Na_2MoO_4 \cdot 2H_2O$, 1.0 g H_3BO_4 and 0.1 g KI.
   b. Per liter 1,000× vitamin solution: 0.05 g biotin, 1.0 g calcium pantothenate, 1.0 g nicotinic acid, 25.0 g inositol, 1.0 g thiamine·HCl, 1.0 g pyridoxine·HCl, and 0.2 g *para*-aminobenzoic acid.

      250 g/L glucose solution is sterilized by autoclaving at 110°C, 1,000× vitamin solution is sterilized by filtering through 0.2 µm filter and 1,000× trace element solution and the $(NH_4)_2SO_4$, KH_2PO_4, $MgSO_4 \cdot 7H_2O$ solution are sterilized by autoclaving at 121°C.

3. Methods

To investigate the inhibitory effects of the lignocellulosic biomass hydrolysates using exo-metabolomics, generating a diverse collection of hydrolysates is crucial (11, 12). Biomass type and hydrolysate

preparation method are the sources for creating such diversity. In particular, the hydrolysate preparation method is a target of optimization. Therefore, establishing various hydrolysate preparation methods and applying them on different biomass types is of importance to allow a meaningful experimental design for process improvement.

The four hydrolysate preparation methods selected in our study are (1) concentrated acid, (2) dilute acid, (3) mild alkaline and (4) alkaline/oxidative. Each method composes of a pretreatment step, which breaks down the structure of the biomass, and a hydrolysis step, which hydrolyzes the cellulose and hemicellulose into sugar monomers. Dilute acid, mild alkaline and alkaline/oxidative are relatively mild methods; they employ acidification, alkalization and oxidation for pretreatment, respectively. Dilute acid pretreatment takes place under high pressure for several minutes; alkaline/oxidative pretreatment is carried out at high temperature for many hours; while mild alkaline pretreatment undergoes high temperature for multiple days. All three methods apply enzymatic hydrolysis, which can be conducted in a fed-batch manner to reach high glucose concentration in the hydrolysate. Concentrated acid is more aggressive, it uses concentrated sulfuric acid for both pretreatment and hydrolysis. This method is more suitable for preparing hard wood hydrolysate than the three mild methods.

The detailed procedure of these four hydrolysate preparation methods are described in the following two sections, separating pretreatment and hydrolysis.

3.1. Pretreatment

To prepare 1 L lignocellulosic biomass hydrolysate, 500 g dried biomass is used for concentrated acid method, while 300 g is used for the other three methods.

The pre-dried biomass contains 5–10% free water, indicating that the initial biomass amount before drying should be above 550 and 330 g. The drying takes place in an oven at 80°C, for at least 16 h.

3.1.1. Concentrated Acid

1. 500 g dried biomass is impregnated with 72% H_2SO_4 in a 5 L Duran laboratory glass bottle. The biomass dry matter to H_2SO_4 ratio is 1:2 (w:w), the concentration of 72% H_2SO_4 is 1,155 g/L, so the amount of 72% H_2SO_4 added is 866 mL.
2. The biomass is impregnated by rotating the bottle on the two-deck modular roller at room temperature for 24–48 h, rotating speed 10 rpm.
3. The pretreated biomass is stored at 4°C before hydrolysis.

3.1.2. Dilute Acid

1. 300 g dried biomass is soaked in 2% (20.28 g/L) H_2SO_4, at a ratio of 8 g dried biomass/100 mL 2% H_2SO_4, in a 5 L polypropylene (PP) bucket, i.e. in total 3,750 g 2% H_2SO_4 is used. The biomass is impregnated in an oven at 60°C for minimum 3 h, with the bucket lid closed.

2. The impregnated biomass is cooled down to room temperature, and the free liquid in the wet biomass is partially removed by vacuum filtration.
3. The wet biomass is transferred onto an aluminum dish, which is then autoclaved at 160°C for 3.5 min.
4. After the wet biomass is cooled down to room temperature, its liquid content is determined by measuring the total wet biomass weight. If the liquid content in the biomass exceeds 1,000 g, i.e. if the biomass dry matter content is below 23%, the wet biomass is air dried at room temperature (20°C) until the liquid content drops below 1,000 g (see Note 1).
5. The wet biomass is stored at 4°C before enzymatic hydrolysis.

3.1.3. Mild Alkaline

1. A $Ca(OH)_2$ suspension is added to the dried biomass at 0.15 g $Ca(OH)_2$/g biomass and at a 20% biomass dry matter content in a 5 L PP bucket: 300 g biomass is mixed with 48 g $Ca(OH)_2$ suspended in 1,500 g demineralized water.
2. The biomass is impregnated by placing the bucket in an oven at 80°C for 3–5 days, with the bucket lid closed.
3. After the wet biomass is cooled down to room temperature, its liquid content is determined by measuring the total wet biomass weight.
4. The wet biomass is air dried at room temperature (20°C) to reduce its liquid content to below 1,000 g, i.e. biomass dry matter content increases to 23% (see Note 1).
5. Store the biomass at 4°C before enzymatic hydrolysis.

3.1.4. Alkaline/Oxidative

This pretreatment method is adapted from a method described by Xuebing Zhao et al (15).

1. 300 g dried biomass is mixed with 3.3% NaOH at a ratio of 1:3 (w:v) in a 5 L Duran laboratorial glass bottle. The biomass is impregnated by placing the bottle in a water bath at 90°C for 90 min (see Note 2).
2. The impregnated biomass is cooled down to room temperature and washed with demineralized water, using vacuum filtration, to remove the detached and dissolved lignin (see Note 3). Vacuum filtration is continued at the end of the washing to increase the dry matter content of the wet biomass.
3. Prepare peracetic acid (PAA) by adding acetic acid, 30% H_2O_2 and 96% H_2SO_4 at a ratio of 60:40:1 (v:v:v) together, and allowing to react at room temperature for 72–96 h before use.
4. As soon as the PAA solution is ready for use, add it to the washed biomass at a ratio of 1:1 (w:w) with respect to the initial dried biomass amount, i.e. 300 g, in a 2 L Duran laboratory

glass bottle. The mixture is impregnated in an oven at 70°C for 2 h, while the bottle is rotated at a speed of 10 rpm on the two-deck modular roller apparatus.

5. After the impregnation, the biomass is washed with demineralized water to remove PAA. The washing is facilitated by vacuum filtration, and PAA is considered as removed when washing water pH reaches 3.5.
6. The washed biomass is mixed with 3–4 L demineralized water, and the slurry pH is adjusted to 5.0 using 6 M NaOH.
7. The free liquid in the mixture is then partially removed by vacuum filtration. The liquid content of the vacuum filtrated biomass is determined by measuring the total wet biomass weight. If liquid content of the wet biomass exceeds 1,000 g, i.e. if the biomass dry matter content is below 23%, the wet biomass is air dried at room temperature (20°C) till the liquid content drops below 1,000 g (see Note 1). Store the biomass at 4°C before enzymatic hydrolysis.

3.2. Hydrolysis

Two different hydrolysis methods are applied to the pretreated biomass, namely acid hydrolysis and enzymatic hydrolysis. The biomass pretreated with concentrated acid is hydrolyzed with acid, while the biomass pretreated with either dilute acid, mild alkaline, or peracetic acid method is hydrolyzed enzymatically.

3.2.1. Acid Hydrolysis

1. At the end of the impregnation, 900 mL demineralized water is added to the impregnated biomass to reduce SO_4^{2-} concentration to around 42% (570 g/L). The slurry is then pre-hydrolyzed in the bottle in an incubator at 60°C for 8 h, with shaking speed 100 rpm (see Note 2).
2. The pre-hydrolyzed slurry is cooled down to room temperature and vacuum filtrated to separate the liquid and the solid phase. The filter cake, which has around 50% biomass dry matter content, is collected and stored at 4°C, while the filtrate, the pre-hydrolysate, is neutralized by using anion exchange membranes to reduce SO_4^{2-} concentration to a required level (see Note 4).
3. The anion exchange process is facilitated with ElectroCell (EC) equipment and two pumps, as shown in Fig. 1. The donor compartment is filled with pre-hydrolysate, while the acceptor compartment contains 5 L demineralized water. In about 5 days, the SO_4^{2-} concentration of the pre-hydrolysate is reduced to the required level (see Note 4); during which the acceptor, demineralized water, is refreshed daily, and the SO_4^{2-} concentration is checked daily with HACH LANGE LCK353 kit (see Note 5).
4. The neutralized pre-hydrolysate is then combined with the filter cake from the pre-hydrolysis in a 2 L Duran laboratory glass

bottle. The slurry is hydrolyzed in a second round by placing the bottle in a water-bath at 95°C for 3 h (see Note 2).

5. After the hydrolysis, the slurry is cooled down to room temperature and vacuum filtrated. The filter cake, mainly composed of lignin, is discarded.
6. The filtrate, the hydrolysate, is again neutralized with the anion exchange membranes facilitated with EC equipment, as described before. In about 5 days, the SO_4^{2-} concentration is reduced to below 15 g/L (1.5%); during which the acceptor, demineralized water, is refreshed daily, and the SO_4^{2-} concentration is checked every other day (see Note 5).
7. After anion exchange neutralization, the hydrolysate pH is adjusted to 5.0 using $Ca(OH)_2$ (see Note 6).
8. The hydrolysate is sterilized by using 0.2 μm filters (see Note 7) and stored at −20°C before use.

3.2.2. Enzymatic Hydrolysis

1. The free liquid content of the wet biomass after pretreatment, W_{FL}, is determined by measuring the total wet biomass weight (see (1)).

$$W_{FL}(g) = W_{WB} - W_{DB} \quad (1)$$

Where, W_{WB} is the total wet biomass weight (g)

W_{DB} is the initial dried biomass weight (300 g)

2. W_{H_2O} amount of demineralized water is added to a 2 L Duran laboratorial glass bottle (see Note 8 and (2)).

$$W_{H_2O}(g) = 1200g - W_{FL} \quad (2)$$

3. Wet biomass is added into the same bottle until a mixable slurry is formed (see Note 9). Record the amount of wet biomass added as W_{PWB} and calculate the added amount of initial dried biomass, W_{PDB}, using (3).

$$W_{PDB}(g) = \frac{300g}{W_{PWB}} \times W_{PWB} \quad (3)$$

4. Enzyme Accellerase 1500 and pen:strep are then added into the slurry according to the mixing ratio, 0.5 mL Accellerase 1500/g initial dried biomass, and 1 mL pen:strep/100 mL slurry.
5. Estimate the H_2SO_4 or the $Ca(OH)_2$ content, $W_{PH_2SO_4}$ or $W_{PCa(OH)_2}$, in the slurry if dilute acid or mild alkaline pretreated biomass is used, with (4a) and (5a). Pre-neutralize the slurry by adding either $Ca(OH)_2$ powder or pure H_2SO_4, $W_{ACa(OH)_2}$ or $W_{AH_2SO_4}$, according to (4b) and (5b).

$$W_{PH_2SO_4}(g) = (W_{PWB} - W_{PDB}) \times 0.0208 g / g \quad (4a)$$

$$W_{ACa(OH)_2}(g) = 95\% \times \frac{W_{PH_2SO_4}}{98 g / mol} \times 74 g / mol \quad (4b)$$

$$W_{PCa(OH)_2}(g) = W_{PDB} \times 0.15 g / g \quad (5a)$$

$$W_{AH_2SO_4}(g) = 105\% \times \frac{W_{PCa(OH)_2}}{74 g / mol} \times 98 g / mol \quad (5b)$$

Where, 0.0208 g/g is the concentration of 2% H_2SO_4
0.15 g/g is the adding ratio of $Ca(OH)_2$ in mild alkaline pretreatment
98 g/mol is the molar mass of H_2SO_4
74 g/mol is the molar mass of $Ca(OH)_2$
95 and 105% are rough compensation factors for reaching pH 5.0 instead of pH 7.0

6. The hydrolysis is conducted in an incubator at 50°C with shaking speed 150 rpm. After about 1 h, the slurry pH is measured with pH electrode and re-adjusted to 5.0 using either 6 M NaOH or 3 M H_2SO_4. The hydrolysis is then continued under the same condition till the slurry is liquefied (see Note 10).
7. The pH of the liquefied slurry is adjusted with either 6 M NaOH or 3 M H_2SO_4 to 5.0, before the next round of wet biomass addition.
8. Step 3–7 are repeated until all the wet biomass is used, the hydrolysis slurry is then incubated at 50°C with shaking speed 150 rpm for another 48–72 h.
9. After the enzymatic hydrolysis, the slurry is cooled down to room temperature and vacuum filtrated or centrifuged to separate the hydrolysate from the remaining solids.
10. The hydrolysate is stored at −20°C before use.

3.3. Bioscreen C Analyzer Evaluation

The inhibitory effects of the generated hydrolysates are examined with growth tests, measuring growth rates. The growth tests were conducted in 100-well honeycomb plates in a Bioscreen C Analyzer (13), which monitors the growth curves by measuring optical density in a preset time interval.

3.3.1. Preculture Preparation

1. The preculture of the growth test is conducted in a 500 mL shake flask with 100 mL mineral medium. The preculture is inoculated with 1 mL glycerol stock and cultivated at 30°C for about 24 h with shaking speed 200 rpm.
2. After the incubation, the pre-culture optical density (OD_{600}) is measured, and if it reaches approximately 3.5, 5 mL pre-culture is centrifuged at 15,000 rpm, 4°C, for 20 min to separate the cells from the medium.

3. After the centrifugation, the medium is discarded and the cells are resuspended in sterilized demineralized water to reach OD_{600} 1.5–2.0, which is used as inoculum for Bioscreen C incubation.

3.3.2. Bioscreen C Incubation

1. The growth test is conducted in triplicate in the 100-well honeycomb plates. The Bioscreen C Analyzer can cultivate two plates at the same time. To minimize medium evaporation, the wells that are on the outside border are filled with 400 μL demineralized water each, leaving 72 wells to be used for growth tests on one plate.
2. Before filling the wells, Bioscreen C Analyzer is set at a temperature of 30°C, a detecting wavelength of 420–580 nm, a measuring interval of 15 min and no shaking (see Note 11) (13). The experimental duration is set long enough to allow the growth in all wells to complete, e.g. 5 days.
3. Fill each of the 72 wells with either 400 μL medium or hydrolysate according to the experimental design, and inoculate with 4 μL prepared inoculum.
4. Place the inoculated plate(s) in the Bioscreen C Analyzer and start the experimental run.
5. When all growth curves are complete, stop the experimental run and collect the measured optical density data from the computer. Calculate the $\ln(OD_{\text{measure}} - OD_{\text{blank}})$ value for each measurement point and plot these values against time (hours). Determine the slope of the linear part of the plot and use it as measure for growth rate.

3.4. Growth Test Results of 24 Hydrolysates Prepared from Six Biomass Types

The hydrolysate preparation methods were applied to 6 different biomass, namely, sugar cane bagasse, corn stover, wheat straw, barley straw, oak sawdust and willow wood chips. The resulting 24 hydrolysates were used as media of the Bioscreen C Analyzer incubation to test their inhibitory effects on two model strains, *S. cerevisiae* CEN.PK 113-7D and Ethanol Red Yeast. The results are presented as growth rates in Figs. 2 and 3. The reproducibility of the Bioscreen C Analyzer growth tests is high. The experiments were carried out in triplicate, of which the standard deviations of calculated growth rates were between 0.5 and 15%, with an average of 2.6%. An example of the Bioscreen C growth test results is shown in Fig. 4. In general it can be seen that the optical density readings are stable for all growth curves. Occasional unstable readings are discarded for further growth rate calculations.

The results show that the growth rate among different hydrolysates was diverse. This diversity was caused by both biomass type and hydrolysate preparation method, though the latter seems to be the major influencer, indicating that the forming or/and releasing of the inhibitory compounds tightly relates to the pretreatment

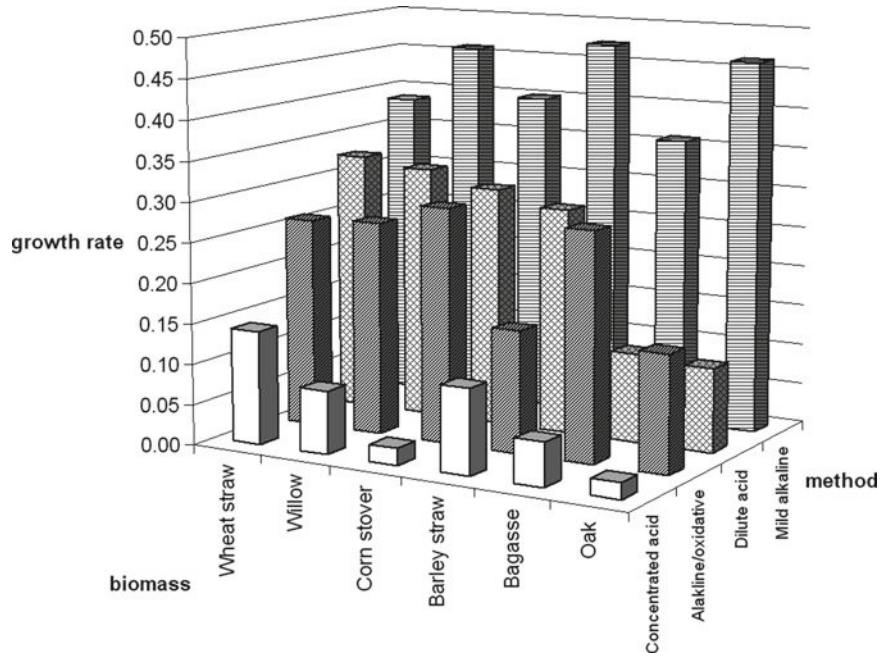


Fig. 2. Growth rates of *S. cerevisiae* CEN.PK 113-7D in 24 different hydrolysates.

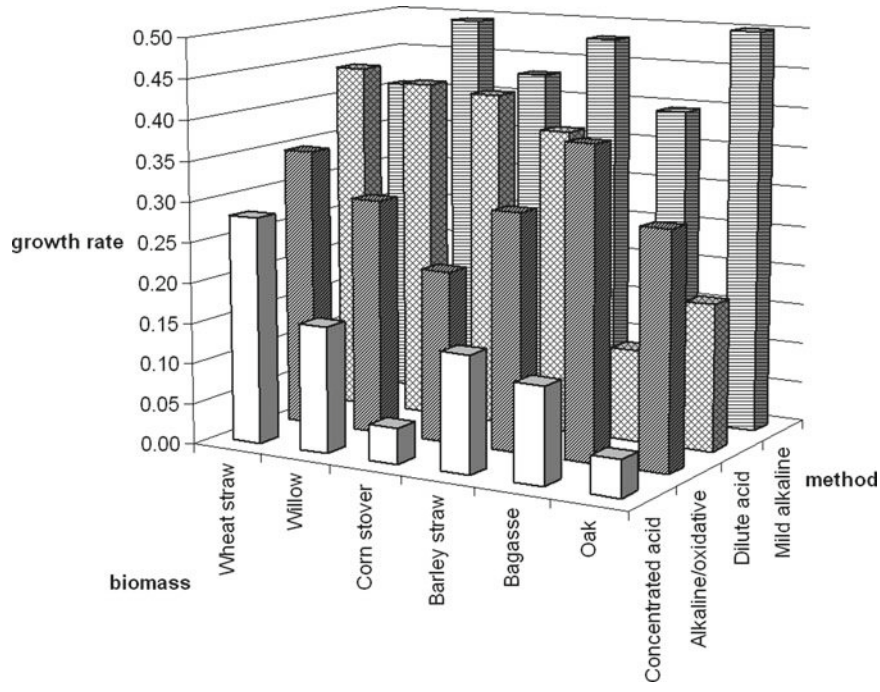


Fig. 3. Growth rates of Ethanol Red Yeast in 24 different hydrolysates.

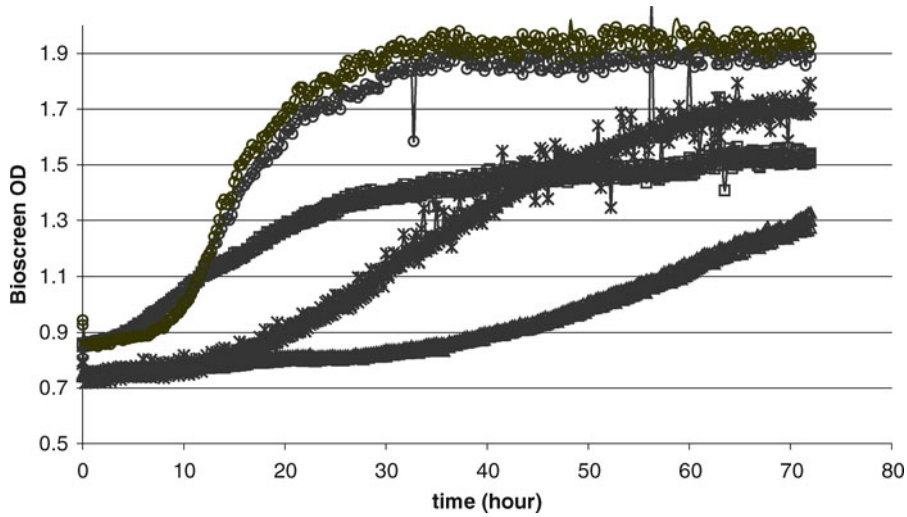


Fig. 4. An example of Bioscreen C growth test results: the growth of *S. cerevisiae* CEN.PK 113-7D in hydrolysates prepared from bagasse. *Circle*, mild alkaline hydrolysate; *star*, dilute acid hydrolysate; *square*, alkaline/oxidative hydrolysate; *triangle*, concentrated acid hydrolysate.

method. As expected, the growth rates in all hydrolysates were lower than those in mineral medium (MM).

It can be seen from Figs. 2 and 3 that both yeasts performed best in mild alkaline prepared hydrolysates, independent of the biomass type, suggesting that least inhibitory compounds were released or formed when mild alkaline was used as pretreatment method. However, the hydrolysis efficiency of the alkaline pretreated biomass was lower compare to the other 3 methods, especially when wood is used (results not shown). This makes mild alkaline a less attractive pretreatment method on industrial scale. On the contrary, the hydrolysates prepared with concentrated acid method had much stronger inhibitory effect on both yeasts. Particularly, when corn stover or oak sawdust was used, little growth was observed. Concentrated acid was a much severer method, which resulted in hydrolysates with high glucose concentration (80–95 g/L) but also strong inhibitory effects. This makes this method, because of its high efficiency, interesting for industrial use, valuable for studying the effects of inhibitory compounds in the hydrolysate.

As far as biomass type is concerned, the highest average growth rate among all four hydrolysate preparation methods was in wheat straw hydrolysates, considering concentrated acid and dilute acid methods particularly. The performance of wheat straw and barley straw hydrolysates were quite similar (see Figs. 2 and 3). This observation is different from the result published by João RM Almeida et al. (1), which indicated that barley straw hydrolysate is much more inhibitory than wheat straw hydrolysate. Bagasse and

oak hydrolysates had the lowest average growth rate, especially in dilute acid prepared hydrolysates (see Figs. 2 and 3).

The two microorganisms used in this experiment, *S. cerevisiae* CEN.PK 113-7D and Ethanol Red yeast, had similar performance in different hydrolysates, indicated by the comparable pattern of Figs 2 and 3. This may due to the close phenotype of the two microorganisms, both are ethanol production yeast strains. An overall comparison of the two strains shows that the average growth rate of Ethanol Red yeast is higher than CEN.PK, suggesting that the industrial strain is more resistant to the inhibitory compounds than the laboratorial one.

In conclusion, the method described here provides a platform for, on one hand, a quick and simple analysis of hydrolysate performance, which helps select the biomass type and the hydrolysate preparation method when a specific microorganism is used as production host. On the other hand, for screening of different microorganism strains in a particular hydrolysate to test their resistance on inhibitory compounds as well as other phenotypes, such as optimum growth pH and temperature.

4. Notes

1. To ensure the starting of enzymatic hydrolysis, a certain amount of water needs to be added to the pretreated biomass to form a mixable slurry. Limiting the free liquid content in the wet biomass to below 1,000 mL allows the addition of minimum 200 mL demineralized water at the beginning of the hydrolysis.
2. To avoid pressure difference, don't completely close the bottle lid.
3. The washing process may cost 20–30 L demineralized water, depending on the biomass type. The detached and dissolved lignin is considered removed when the color of the washing water turns from dark black to light yellow.
4. The SO_4^{2-} concentration in the pre-hydrolysate should decrease to a specific level so that when it is mixed with the filter cake, the slurry has a SO_4^{2-} concentration of 150–200 g/L. This concentration range is optimum for the second round hydrolysis, which assists the releasing of the remaining sugars in the filter cake, while causes little sugar degradation in the hydrolysate.
5. SO_4^{2-} concentration testing kit, HACH LANGE LCK353, has a testing range of 150–900 mg/L. Therefore, the samples need to be diluted 100–1,000 times before testing.

6. To avoid dilution, $Ca(OH)_2$ is added as dry powder into the hydrolysate, so pH is only measured after complete $Ca(OH)_2$ dissolution.
7. The filter sterilization is only suitable for small scale experiments. For industrial scale, contaminations may be eliminated by using antimicrobial compounds, such as antibiotics or antimicrobial plant (hop) extracts.
8. To make 1 L hydrolysate, there should be between 1,300 and 1,400 mL free liquid in the enzymatic hydrolysis slurry. Since, per 300 g dried biomass, in total 150 mL enzyme Accellerase 1500 will be added, $1{,}200\text{g} - W_{\text{FL}}$ demineralized water is added here to make up the volume.
9. The slurry should be mixable to allow pH adjustment and to ensure the starting of enzymatic hydrolysis. For most biomass types, if 200 mL demineralized water was added firstly, the wet biomass adding amount is around 200 g.
10. It takes about 4 h to liquefy the slurry, which allows the proceeding to the next hydrolysis step.
11. Choosing no shaking is to avoid uneven cell distribution and to obtain steady optical density reading during the whole growth stage (13).

Acknowledgment

This project was co-financed by the Netherlands Metabolomics (NMC) which is part of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research.

References

1. Almeida JRM, Karhumaa K, Bengtsson O, Gorwa-Grauslund MF (2009) Screening of *Saccharomyces cerevisiae* strains with respect to anaerobic growth in non-detoxified lignocellulose hydrolysate. Bioresource Technology 100(14):3674–7
2. Modig T, Almeida JRM, Gorwa-Grauslund MF, Lidén G (2008) Variability of the response of *Saccharomyces cerevisiae* strains to lignocellulose hydrolysate. Biotechnology and Bioengineering 100(3):423–9
3. Tofighi A, Azin M, Mazaheri Assadi M, Assadirad MHA, Nejadsattari T, Fallahian MR (2010) Inhibitory effect of high concentrations of furfural on industrial strain of *saccharomyces cerevisiae*. International Journal of Environmental Research 4(1):137–42
4. Albers E, Larsson C (2009) A comparison of stress tolerance in YPD and industrial lignocellulose-based medium among industrial and laboratory yeast strains. Journal of Industrial Microbiology and Biotechnology 36(8):1085–91
5. Martín C, Jönsson LJ (2003) Comparison of the resistance of industrial and laboratory strains of *Saccharomyces* and *Zygosaccharomyces* to lignocellulose-derived fermentation inhibitors. Enzyme and Microbial Technology 32(3-4): 386–95
6. Klinke HB, Thomsen AB, Ahring BK (2004) Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Applied Microbiology and Biotechnology 66(1):10–26

7. Palmqvist E, Hahn-Hägerdal B (2000) Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification. Bioresource Technology 74(1):17–24
8. Almeida JRM, Modig T, Petersson A, Hahn-Hägerdal B, Lidén G, Gorwa-Grauslund MF (2007) Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. Journal of Chemical Technology and Biotechnology 82(4):340–9
9. Nichols NN, Sharma LN, Mowery RA, Chambliss CK, van Walsum GP, Dien BS, Iten LB (2008) Fungal metabolism of fermentation inhibitors present in corn stover dilute acid hydrolysate. Enzyme and Microbial Technology 5;42(7):624–30
10. Heer D, Sauer U (2008) Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. Microb Biotechnol 1:497–506
11. Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: A web server for metabolomic data analysis and interpretation. Nucleic Acids Research 37
12. Braaksma M, van den Berg RA, van der Werf MJ, Punt PJ (2010) A top-down systems biology approach for the identification of targets for fungal strain and process development. In: Borkovich KA, Ebbole DJ (ed) Cellular and molecular biology of filamentous fungi. ASM Press, Washington, DC
13. Franden MA, Pienkos PT, Zhang M (2009) Development of a high-throughput method to evaluate the impact of inhibitory compounds from lignocellulosic hydrolysates on the growth of *Zymomonas mobilis*. Journal of Biotechnology 144(4):259–67
14. van Hoek P, van Dijken JP, Pronk JT (1998) Effect of specific growth rate on fermentative capacity of baker's yeast. Appl Environ Microbiol. 64(11):4226–33
15. Zhao X, Peng F, Cheng K, Liu D (2009) Enhancement of the enzymatic digestibility of sugarcane bagasse by alkali-peracetic acid pretreatment. Enzyme and Microbial Technology 44(1):17–23

Chapter 17

Engineering Whole-Cell Biosensors with No Antibiotic Markers for Monitoring Aromatic Compounds in the Environment

Aitor de las Heras and Víctor de Lorenzo

Abstract

A cornerstone of Synthetic Biology is the engineering of gene regulatory networks. Construction of such biological circuits has been used not only to elucidate the dynamics of gene expression but also for designing whole-cell biosensors that translate environmental signals into quantifiable outputs. To this end, distinct components of given regulatory systems are rationally rewired in a way that translates an external stimulus (for instance, the presence of one chemical species) into a measurable readout typically fluorescence or luminescence. Various biosensors for BTEX (a mixture of benzene, toluene, ethylbenzene and xylenes) are based on XylR, the main transcriptional regulator of the TOL pathway of *Pseudomonas putida* mt-2. In the presence of its natural effectors (e.g., *m*-xylene, toluene or 3-methylbenzylalcohol), XylR triggers expression of the *upper* pathway genes by means of the *Pu* promoter. Available biosensors combine the *xylR* gene and a direct fusion between the cognate *Pu* promoter and the *luxCDABE* operon, all components stably integrated in the chromosome of *P. putida*. A versatile development of the same biosensing concept is described, aimed at increasing the sensitivity of the genetic circuit toward XylR inducers. The new platform utilizes mini-transposon vectors tailored for engineering an artificial expression cascade that operates as an amplifier of the signal/response ratio of the biosensor. This strategy was applied to the construction of a strain that carries a transcriptional fusion between the *Pu* promoter and T7 RNA polymerase (which becomes under the control of XylR and its effectors), along with a T7 promoter controlling expression of the *luxCDABE* operon. This simple regulatory architecture produced a dramatic increase of bioluminescence emission in respect to the strain that carries only the direct fusion between the *Pu* promoter and the *luxCDABE* reporter.

Key words: Synthetic biology, Bacterial biosensors, Non-antibiotic selection, Genetic stability, Mini-transposons, Orthogonal circuits.

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_17, © Springer Science+Business Media, LLC 2012

1. Introduction

Engineering bacteria with novel regulatory networks to produce singular outputs has been one of the topics most addressed by Synthetic Biology (1–4). Construction and implantation of such biological circuits in bacterial hosts have been used both to elucidate fundamental questions of gene expression control and for designing whole-cell biosensors for physical or chemical signals. One biotechnological application of considerable interest lies in the construction of bacteria for monitoring specific chemicals in soil (5, 6). In the cases where such sensors are intended for extensive release, it is necessary to maintain the functionality of the engineered genetic circuit and the corresponding reporter output in the absence of external selective pressure e.g., antibiotic selection, which are commonplace in laboratory setups. This places specific constraints on the genetic tools utilizable in bacteria destined for environmental release, as most existing whole-cell biosensors have been constructed by assembling the reporter system and the sensor module on plasmids bearing antibiotic resistances (7–9).

The most common organization of an engineered biosensor circuit is the one composed of one transcriptional factor (TF) that responds directly to the primary environmental signal and a cognate promoter fused to a reporter gene with an enzymatic or an optical readout (fluorescence or luminescence). Various biosensors of this sort, aimed at detection of BTEX (a mixture of benzene, toluene, ethylbenzene and xylenes) and other aromatics have been engineered in soil bacteria (e.g., *Pseudomonas putida*) tailored for extensive application to sites suspect of containing this type of pollutants. One of the most elaborated examples of this sort involved the development of a dedicated genetic platform for stable implantation of different transcriptional regulators and reporter genes in a variety target bacterial genomes (10). Besides allowing chromosomal integration of thereby formatted gene fusions, the same platform facilitated deletion of all markers that were first employed for selection and screening of the constructs. This is an important requirement for bacteria destined for environmental release, since deliberate discharge of antibiotic resistance genes is considered hazardous and ultimately not acceptable. The downside of monocopy, chromosomally integrated sensor genetic circuits is that the intensity of the reporter readout is limited by the intrinsic specificity, sensitivity, and capacity (i.e., maximum output) of the primary regulator/promoter pair, which varies to a large extent. In this chapter, we describe a genetic platform which maintains the requirements applicable to whole-cell biosensors for environmental release (stable genomic integration, no antibiotic resistances) while multiplying the output by means of an artificial expression cascade based on the T7 polymerase. As a case study, we describe below the construction

and validation of a BTEX biosensor endowed with a simple genetic device based on the toluene-sensing XylR protein of *P. putida* and its cognate promoter *Pu* that amplifies the emission of luminescence when the engineered bacteria face aromatic compounds of this sort.

1.1. Rationale and Components of the T7 Polymerase-Based Amplifying Cascade

The biosensor-building strategy described in this Chapter attempts the formatting of a host strain (in principle, any Gram-negative bacterium) that acts as a standarized *landing pad* for multiple sensor modules. Moreover, since the genetic platform is devised for chromosomal integration of the signal-sensing device and for elimination of antibiotic resistance markers, the TF/promoter pairs can be introduced into such strain for eventual environmental release. Assembly of such whole-cell biosensor is based on vectors that incorporate functional parts of two different transposition systems (1) Tn*5*, that inserts randomly into the bacterial chromosome of a large variety of bacteria (11) and (2) Tn*7* that preferentially targets distinct *att*Tn*7* sites, an event that occurs at high frequency in optimal genomic sites and in one specific orientation (12, 13). The functional organization of these three vectors is sketched in Fig. 1. The first one is plasmid pTn*5*Tel [*T7*→ *lux*], which is the suicide delivery vector for a mini-Tn*5* transposon that carries the *luxCDABE* operon of *Photorhabdus luminescens* behind a strong P_{T7} promoter that can be activated by the corresponding T7 phage RNA polymerase (see Fig. 1a) in a fashion altogether independent of the housekeeping RNAP of the bacterial host. Insertion of this transposon into the target Gram-negative genome is selected by means of the telurite resistance genes, a nonantibiotic marker that is perfectly admissible for environmental release (14). The second component of the vector system is pTn*5*-*T7pol*, which delivers a second mobile element with the array of functional sequences indicated in Fig. 1b. The organization of this plasmid is basically identical to that of pTn*5*-LacZ (10), excepting that the *lacZ* insert has been replaced by a promoterless T7 RNA polymerase (*T7pol*). As a result, *T7pol* is placed downstream of the optimal site for Tn*7* insertion in *E. coli* (*att*Tn*7*; (15)), which lies adjacent to a kanamycin (Km) resistance gene. Finally, the third element of the genetic platform described herein is pTn*7*-*FRT* (10). This is a mini-Tn*7* delivery plasmid that operates as the cloning vector for any sensor module consisting of the gene for the desired signal-sensing TF and its matching promoter (see Fig. 1c). Insertions of the mini-Tn*7* can be selected with gentamicin (Gm) and the marker later deleted as explained below.

Besides the mobile element indicated in each case, the vectors of Fig. 1 bear the *bla* gene for resistance to beta-lactam antibiotics (*bla*) located outside the transposon ends. This allows discrimination, when the time comes, of true transposition events from mere co-integration of the delivery plasmid with the target genomes. Yet, the critical feature of the platform embodied in these plasmids is the presence in pTn*5*-*T7pol* and pTn*7*-*FRT* of purposefully located

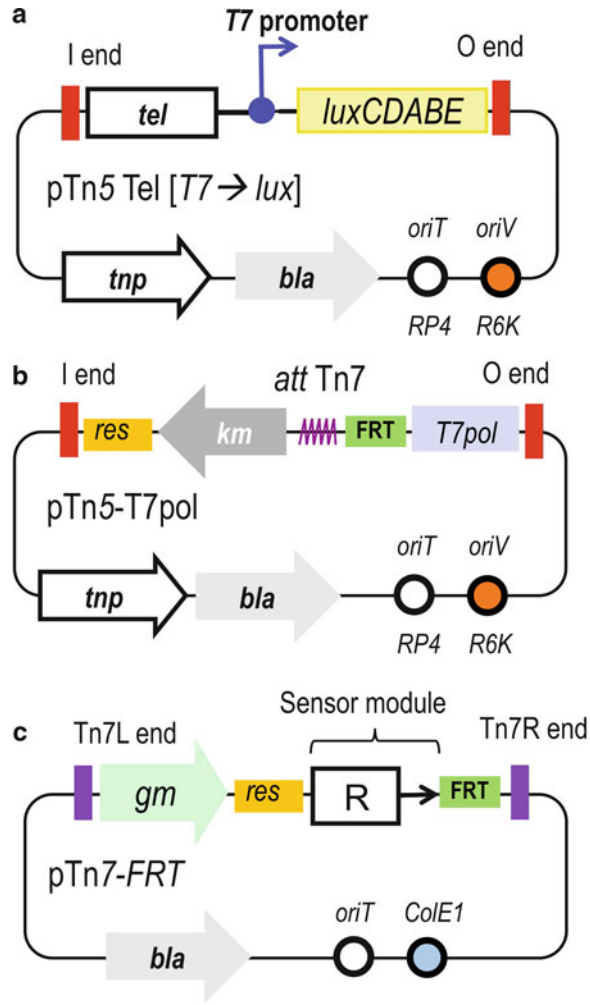


Fig. 1. Structure (not to scale) of the mini-transposon delivery vectors used in this work. (**a**) Structure of mini-Tn*5* transposon suicide delivery plasmid (pTn5 Tel [*T7→ lux*]) bearing nonantibiotic resistance marker (*tel*) and a fusion between the *luxCDABE* operon and the cognate promoter of the RNA polymerase of the bacteriophage T7. (**b**) Structure of mini-Tn*5* transposon suicide delivery plasmid (pTn*5-T7pol*) bearing an optimized *att*Tn*7* sequence, promoterless T7 RNA polymerase gene, *res, FRT* sequences, and Km resistance gene assembled in a conditional replication plasmid. The important elements for both mini-Tn*5* delivery vectors are colored: Tn*5* transposase coding gene *(tnp)*, the origins of replication (*R6K)*, and transference (*RP4*) and resistance gene to beta-lactamic antibiotics (*bla*). (**c**) Structure of mini-Tn*7* transposon delivery plasmid pTn7-*FRT* inserted with the transcriptional regulator and the cognate promoter that form the sensor module. *res, FRT* sequences, the genes that confer antibiotic resistance (*gm* and *bla*), and the origins of replication (*oriT*) and transference (*ColE1*) are colored.

FRT sequences from the yeast *FRT*/FLP site-specific recombination system (16), along with *res* sequences from the multimer resolution system of broad host-range plasmid RP4 (17). As disclosed later in the Chapter, the definite order in which these sequences are arrayed

in each of the vectors allow the eventual deletion of the antibiotic resistance markers and the tractable activation of the biosensor circuit (10).

1.2. Implantation of a Regulatory Cascade for Enhanced Biosensing of BTEX in P. putida Cells Destined for Environmental Release

As a study case, we describe below all steps for engineering a regulatory cascade that amplifies the response of *P. putida* to the aromatics present in BTEX with the judicious use of the three vectors described in Fig. 1. To this end, the overall strategy is briefly described in this section, followed below by complete protocols for each step of the process. The itinerary toward the enhanced and release-able BTEX biosensor starts with the insertion of the *T7→ lux* module of pTn*5*Tel [*T7→ lux*] into the genome of the target strain (*P. putida* KT2440) by selection for resistance to potassium tellurite. This operation enters the output-producing component of the circuit i.e., a P_{T7}→ *luxCDABE* fusion in the genome of the carrier strain. Note that in the absence of the T7 polymerase, transcription of the reporter cannot happen and emission of luminescence is virtually none. One exconjugant from such transposition procedure (see Fig. 2a) was named *P. putida* T7·LUX4 and employed as

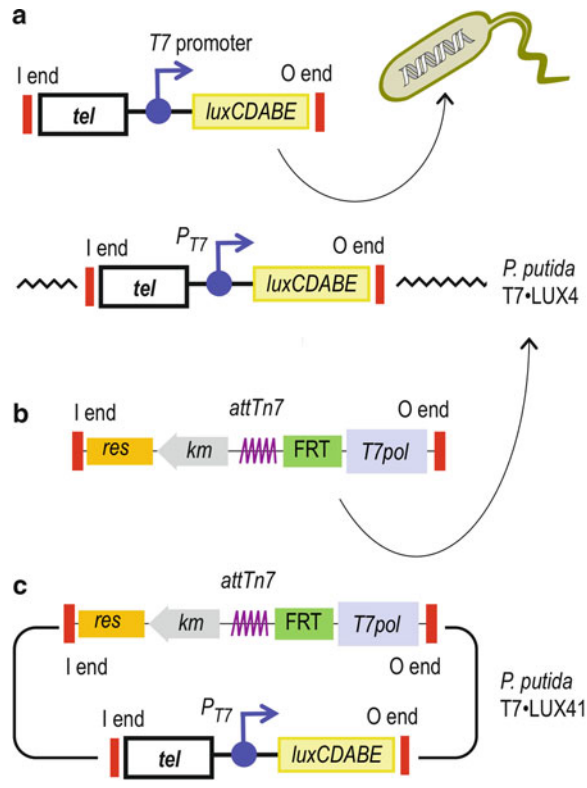


Fig. 2. Construction of the formatted strain that acts as genomic anchor for the sensor circuit in the *prerelease* strain. (**a**) insertion of Tn*5*Tel [*T7→lux*] in the chromosome of *P. putida* generates strain T7·LUX4. (**b**) Insertion of Tn*5-T7pol* in the chromosome of *P. putida* T7·LUX4 produces strain T7·LUX41. (**c**) Organization of relevant insertions in *P. putida* T7·LUX41 (not to scale).

recipient of a second transposition round, in which the donor plasmid was pTn*5-T7pol*. Selection of Km^R colonies allows the appearance of clones which besides the previously inserted Tn 5 Tel [*T7→ lux*], have the complex insert sketched in Fig. 2b in their chromosome. One of such clones (named *P. putida* T7·LUX41, see Fig. 2c) thus contains both segments with P_{T7}→ *luxCDABE* and T7pol, but the later lacks any expression signals and therefore luminescence emission remains as low as *P. putida* T7·LUX4. As indicated in Fig. 3a, strain *P. putida* T7·LUX41 becomes the standarized host for further insertion of the core sensor module, consisting of the gene for a transcriptional regulator along with a cognate promoter. In our example, the regulator is encoded by the *xylR* gene of the TOL plasmid, which acts on the *Pu* promoter (18) This module is first cloned in vector pTn*7-FRT* (see Fig. 1c and Note 1), yielding in our case plasmid pTn*7* [*FRT*-BXPu] (10). In this construct (see Fig. 3a), the TF/promoter segment becomes bracketed upstream and downstream by *res* and *FRT* sequences, respectively. This arrangement of functional segments is itself inside a Gm^R mobile element delimited by the L and R termini of Tn*7*. When this transposon is delivered to *P. putida* T7·LUX41, Gm^R/Km^R exconjugants which are sensitive to the beta-lactam antibiotic piperacillin (Pip) must contain the group of sequences of various origins represented in

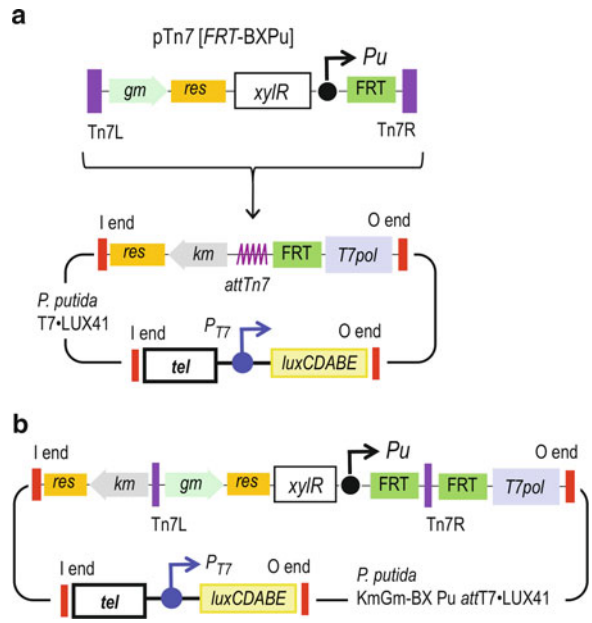


Fig. 3. Insertion of the sensor module into the *landing pad* of the formatted *P. putida* T7·LUX41. (**a**) insertion of mini-Tn*7* carrying *xylR* and its cognate promoter *Pu* into *P. putida* T7·LUX41. (**b**) Genetic structure of the *prerelease* strain carrying Km and Gm markers flanked by *res* sequences, the *xylR* gene and its corresponding promoter *Pu,* and the T7 RNA polymerase gene downstream Tn*7* right end, flanked by *FRT* sequences.

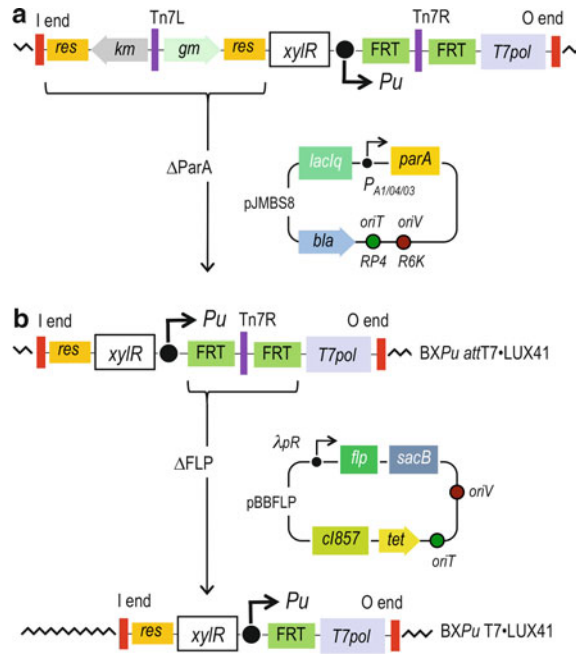


Fig. 4. Enabling the biosensor strain. (**a**) Deletion of antibiotic resistance markers of *P. putida* strain, Km Gm-BX Pu *att*T7·LUX41 by means of site-specific recombination between flanking *res* sequences bought about by transient expression of ParA from pJMBS8 plasmid. The organization of the segment of interest after the deletion is shown in the resulting strain *P. putida* BX Pu *att*T7·LUX41, not to scale. (**b**) Deletion of Tn7 right end placed between *Pu* and the T7 RNA polymerase-coding gene by expression of FLP mediated by pBBFLP plasmid and site-specific recombination between the flanking *FRT* sequences. Note the antibiotic-free, functional organization of the resulting module.

Fig. 3b. This is because the mini-Tn*7* borne by pTn*7*[*FRT*-BXPu] is targeted to the *att*Tn*7* site-engineered upstream of the *T7pol* gene of *P. putida* T7·LUX41.

The resulting strain (*P. putida* Km Gm-BX *Pu att* T7·LUX41; see Fig. 3b) already contains all regulatory and functional elements that are needed for the enhanced BTEX biosensor, but they are at this point in a silent state. The flow of transcription between the *Pu* promoter and the *T7pol* gene is blocked by the transcriptional termination activity of the Tn*7* right end located between them (10). Furthermore, the longer genomic implant of the strain encodes genes for Km and Gm resistance, which were inherited owing to the strategy for selection of each of the insertions. This precludes any possible application e.g., deliberate environmental release. Fortunately, the order and orientation of the business segments of *P. putida* Km Gm-BX *Pu att* T7·LUX41 (see Fig. 4, top) has been arranged to both delete antibiotic resistances, and to fully activate the designed cascade for BTEX detection. Enabling the prerelease strain *P. putida* Km Gm-BX *Pu att* T7·LUX41 (see Fig. 3b) into an active biosensor entails two steps (10). First, in vivo deletion of the

segment of the insert containing the antibiotic resistance markers that is limited by the two *res* sequences (see Fig. 4a, top). This is brought about by the conditional expression of the *parA* gene, encoding the cognate site-specific recombinase of RP4 that is present in the suicidal plasmid pJMBS8 (19) (see Fig 4a). As result of this site-specific recombination, strain *P. putida* Km Gm-BX *Pu att* T7·LUX41 loses the resistance cassettes without affecting the regulatory parts of the engineered signal-amplification cascade, thereby generating strain *P. putida* BX *Pu att* T7·LUX41 (see Fig. 4b). The final step of the enabling process is the removal of the terminator Tn*7*R placed between *Pu* and the promoterless *luxCDABE* reporter. In this case (see Fig. 4b), elimination of the unwanted sequence is affected by the site-specific recombination between flanking *FRT* sequences by expressing the yeast flippase (*flp*, FLP; (16)). The result of this concluding action is a completely functional and environmentally safe biosensor strain named *P. putida* BX *Pu* T7·LUX41 (see Fig. 4b). The plasmid encoding FLP (pBBFLP; see Fig. 4b) can be eliminated by plating the strain on a medium with sucrose, which counterselects the *sacB* gene present in the plasmid (10). The *scar* left after recombination between the promoter of the signal-sensing module (i.e., *Pu*) and the *T7pol* gene does not interfere with expression of the T7 RNA polymerase expression when XylR is activated by BTEX.

Strain *P. putida* BX *Pu* T7·LUX41 hosts all the necessary genetic parts for the amplification cascade sketched in Fig. 5a. In sum, XylR effectors bind this TF, which then activates *Pu* and thus causes expression of the *T7pol* gene. In turn, the produced T7 RNA polymerase triggers the P_{T7} promoter in front of the operon *luxCDABE* operon. As shown in Fig. 5b, this cascade endowed the host strain with a considerable response to the presence in the medium of 3-methylbenzylalcohol (3MBA, a proxy of the BTEX inducers). Depending on the growth phase, light readout was 5–10-fold higher than an equivalent strain bearing a direct fusion *Pu-luxCDABE* (10; see Fig. 5b, c). Needless to say that replacing the sensor module in vector pTn*7-FRT* by any other combination of signal-responsive TF/promoter pairs should suffice to generate similar cascades for any other target compounds or, in general, any environmental stimulus of choice.

2. Materials

2.1. Strains and Plasmids

All the strains and plasmids described in this study are listed in Table 1

2.2. Media and Reagents

1. LB medium: 10 g/L of tryptone, 5.0 g/L of yeast extract and 5.0 g/L of NaCl, dissolve in H_2O and autoclave. For LB-agar plates add 1.5% agar.

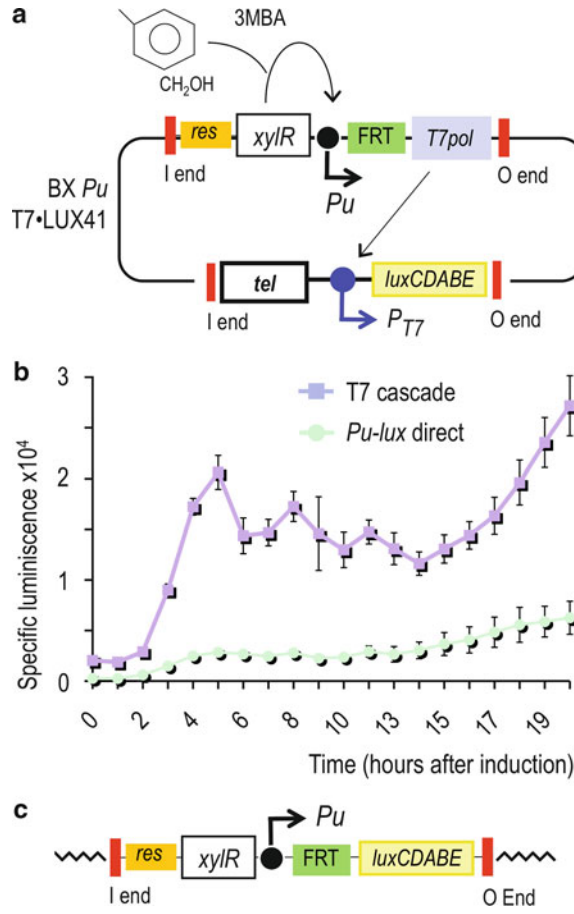


Fig. 5. Response of the BTEX biosensor strain engineered with a transcriptional amplification cascade. (**a**) Sketch of the regulatory circuit borne by *P. putida* strain BX *Pu* T7·LUX41. 3MBA activates XylR, which then triggers the *Pu* promoter and expression of the T7 polymerase. In turn, T7 RNA polymerase activates expression of *luxCDABE*, causing emission of light. (**b**) Luminescence produced by a culture of *P. putida* BX *Pu* T7·LUX41 added with 1 mM 3MBA. In the absence of 3MBA, luminescence is below the detection threshold (not shown). see comparison of the strain engineered with the T7-based cascade and one identical inserted instead with a direct *Pu→luxCDABE* fusion, the organization of which strain (*P. putida* BX *Pu* LUX14; *10*) is drawn in (**c**).

2. Minimal medium M9 preparation (see Note 2 below): First, set a 10× stock solution of M9 salts with 42.5 g Na_2HPO_4-$2H_2O$, 15 g KH_2PO_4, 2.5 g NaCl, and 5 g NH_4Cl, dissolved in 500 mL of H_2O and sterilize; prepare stocks of 1 M $MgSO_4$ and citrate at 20% (as selective carbon source for *Pseudomonas*); also, make a 1.6% agar solution dissolved in H_2O; and autoclave the four components separately. After sterilization, mix components to leave them at the following concentrations: 1× M9 salts, 2 mM $MgSO_4$, 0.2% citrate, and 1.4% agar.
3. Antibiotic concentrations:

Table 1
Strains and plasmids

Strain or plasmid(s)	Relevant characteristics	Reference
Strains		
E. coli CC118 λ*pir*	Δ(*ara-leu*), *araD*, Δ*lac*X174, *galE*, *galK*, *phoA*, *thi1*, *rpsE*, *rpoB*, *argE (Am)*, *recA1*, and lysogenic λ*pir*. Recipient of pTn5 Tel [*T7→lux*], pTn5-T7pol, and pTns1	(26)
E. coli DH5α	Routine host strain for cloning, recipient of pTn7 FRT-BXPu	(22)
E. coli HB101	Sm^r, *hsdR*$^-$*M*$^+$, *pro*, *leu*, *thi*, and *recA*. Host of RK600	(22)
P. putida KT2440	mt-2 cured of plasmid pWW0	(27)
P. putida T7·LUX4	Tel^R, KT2440 with *lux* operon under the control of *T7* promoter	This work
P. putida T7·LUX41	Km^R, *P. putida* T7LUX4 with *T7pol* gene downstream the *att*Tn7 site	This work
P. putida Km Gm-Bx *Pu att*T7·LUX41	Km^R, Gm^R *P. putida* T7·LUX41, and the *xylR/Pu* sensor module inserted upstream the *T7pol* gene	This work
P. putida BX *Pu att*T7·LUX41	*P. putida* Km Gm-Bx Pu *att* T7·LUX41 after the deletion of the antibiotic resistances cassettes	This work
P. putida BX *Pu* T7·LUX41	*P. putida* BxPu *att* T7·LUX41 after the deletion of the Tn7R end	This work
Plasmids		This work
pJMBS8	Ap^R, R6KoriV RP4 *oriT* PA1/04/03::*parA* source of ParA recombinase	This work
RK600	Cm^R; ColE1oriV RK2 *mob*$^+$ *tra*$^+$	(11)
pUC18Not	Ap^R, same as pUC18 but with polylinker flanked by *Not*I sites	(26)
pBBFLP	Tc^R, source of inducible FLP recombinase	(10)
pTn5 Tel[*T7→lux*]	Ap^R Tel^R, pUT mini-Tn5 Tel with a T7 promoter *luxCDABE* operon fusion	This work
pTn5-T7pol	Ap^R Km^R, pTn5 resL Km with *T7pol* gene downstream the *att*Tn7 site	This work
pTns1	Ap^R, RK6 replicon encodes the TnsABC + D specific transposition pathway	(12)
pTn7 [*FRT*-BX *Pu*]	Ap^R Gm^R, pTn7-*FRT* carrying the *xylR/Pu* sensor module	This work

- Ampicillin (Ap) stock: 150 mg/mL in H_2O, filter sterilized. Store at –20°C. For *E. coli* cells use it at a final concentration of 150 μg/mL.
- Kanamycin (Km) stock: 50 mg/mL in H_2O, filter sterilized. Store at –20°C. Use it at a final concentration of 50 μg/mL.
- Chloramphenicol (Cm) stock: 30 mg/mL in ethanol, filter sterilized. Store at –20°C and use it at a final concentration of 30 μg/mL.
- Piperacillin (Pip) stock: 40 mg/mL in H_2O, filter sterilized. Store at –20°C. Use it at a final concentration of 40 μg/mL.
- Tetracycline (Tc) stock: 10 mg/mL with H_2O, filter sterilized. Store at –20°C. Use it at a final concentration of 10 μg/mL.
- Gentamicin (Gm) stock: 10 mg/mL with H_2O, filter sterilized. Store at –20°C. Use it at a final concentration of 10 μg/mL.

4. Tellurite stock: K_2TeO_3 (potassium tellurite hydrate; Aldrich Chemical Co.) dissolved in water at a concentration of 40 mg/mL. The preparation was filtered and kept frozen indefinitely at –20°C. The working concentration of this salt in selective plates was 80 μg/mL.
5. Sucrose solution dissolved at 300 mM in H_2O. Sterilize and maintain at room temperature.
6. 3-methylbenzylalcohol (3MBA) stock solution was dissolved in dimethylsulphoxide (DMSO; Fluka #89890) at a concentration of 1.0 M. Use it at a final concentration of 1 mM. Store at room temperature.

2.3. PCR Primers Used to Verify Targeted Insertion of the pTn7 [*FRT*-BX*Pu*]

In this work we used the following primers (5′→3′) for confirming correct insertion of the *xylR/Pu* sensor module of pTn7 [*FRT*-BX*Pu*] into the att site of the mobile element borne by pTn5-*T7pol*:

- Pu1F: CCCGGGAAAGCGCGATGA, which anneals in the 5′region of the *Pu* promoter of *P. putida*
- T7pol2R: CGGCTTAGGAGGAACTACG, anneals in the T7 RNA polymerase gene present in plasmid pTn*5-T7pol*
- PpuglmS2R: GTGCGTGCCCGTGGTGG, anneals in glutamine synthetase *(glmS)* gene, present in the chromosome of *P. putida*

3. Methods

3.1. DNA Techniques

1. For plasmid preparations we routinely use the Wizard Plus SV Minipreps kit (Promega).
2. PCR-amplified DNA and agarose DNA extractions are commonly purified with the NucleoSpin Extract II kit (MN).

3. In DNA ligations the T4 ligase is employed.
4. PCR: for general PCR reactions prepare first a PCR reaction mix in an Eppendorf tube. The values given below are calculated for a final volume of 25 μL per tube:
   - 5 μL Buffer 5× (provided by the supplier)
   - 1.5 μL $MgCl_2$ 25 mM
   - 1 μL of upstream primer 5 μM
   - 1 μL of downstream primer 5 μM
   - 0.5 μL dNTPs 10 mM
   - 0.3 μL DMSO (specially recommend for organism with a high GC genome content)
   - 0.2 μL Taq Polymerase

   Add 15.5 μL of sterile H_2O into each of the PCR reaction tubes. Pick fresh single colonies directly from a plate and transfer it directly to the PCR reaction tube containing the 15.5 μL of H_2O. Vortex the PCR reaction mixture and distribute 9.5 μL into PCR tubes. Set up the PCR machine with the appropriate Tm (depending on primer composition) and extension time.

3.2. Engineering the Biosensor

P. putida T7·LUX41 strain (see Fig. 2c) was endowed with all the necessary genetic parts that allow construction whole-cell biosensors without antibiotic resistances and with a regulatory cascade that amplifies the output once the sensor module has been introduced. As explained above, this strain was engineered by successive insertions of the mobile elements carried by pTn*5*Tel [*T7*→ *lux*] and pTn*5*-*T7pol* into the genome of *P. putida* KT2440. The corresponding mini-transposons were delivered into recipient cells by three-parental mating procedure (see below). Both plasmids replicate through the *pir*-dependent R6K origin (20; see Fig. 1a, b), so they act as a suicide delivery vectors in any host lacking the π protein. This makes the entire genetic platform to be applicable to a wide variety of gram-negative bacteria so that given strains can be formatted identically to *P. putida* T7·LUX41 for hosting the same regulatory systems. For the sake of illustration, however, we focus below on a detailed description of the methods when applied to *P. putida*.

3.2.1. Construction of Plasmids and Transposons for Circuit Implantation

1. The *BamHI*/*Pst*I 5871 bp segment of pALux4 (10) carrying the *luxCDABE o*peron was cloned into pUC18NotT7 (21) and the corresponding 6071 bp *Not*I fragment (bearing a *PT7* →*luxCDABE* fusion) was passed into pJMT6 (14), producing pTn*5*Tel [*T7*→ *lux*].
2. For assembling pTn5-*T7pol*, the *Xma*I-T7pol-*Pst*I insert from pUC18Not/T7pol (21), carrying the T7 RNA polymerase

gene, was transferred into pBluescript SK- (Stratagene). Then a *BamH*I-T7pol-*Pst*I fragment was cloned into p*attFRT* (10), generating p*attFRT* T7pol. Finally, a *Not*I fragment of this plasmid carrying the T7 RNA polymerase gene was inserted into pUT/mini-Tn*5res*LKm (10), producing pTn5-*T7pol*.

3. pTn*7* [*FRT*-BX*Pu*] was engineered by excising the *Not* I fragment from pBXPu (10), which carries the *xylR/Pu* sensor module, and introducing it into pTn*7*-*FRT* (10).

3.2.2. Suicide Delivery of Mini-Tn5 Derivatives (pTn5Tel [T7→ lux] and pTn5-T7pol)

1. Aerobically grow overnight cultures of:
   - Donor cells: *E. coli* CC118λ*pir* (carrying either pTn*5*Tel [*T7→ lux*] or pTn*5*-*T7pol*) in LB with ampicillin 150 μg/mL and Km 50 μg/mL at 37°C.
   - Helper cells: *E. coli* HB101 (pRK600) in LB with chloramphenicol 30 μg/mL at 37°C.
   - Recipient cells: Grow *P. putida* KT2440 for pTn*5*Tel [*T7→lux*] or *P. putida* T7·LUX4 for pTn5-*T7pol* in LB at 30°C.
2. Mix the three strains in a 1:1:1 ratio (e.g: 100°μL of each strain) in an empty test-tube and complete with 4.7 mL of 10 mM $MgSO_4$.
3. Vortex the mixture and pass through a Millipore filter disc (0.45-μm pore-size, 13-mm diameter) using a 10 mL sterile syringe.
4. Place the filter onto an LB-agar plate and incubate for a maximum of 6 h at 30°C (in this case, this is the optimum temperature for recipient cells).
5. Take the filter from the LB plate with the help of flamed sterile forceps and introduce it into a test-tube with 5.0 mL of 10 mM $MgSO_4$. Vortex to resuspend cells.
6. Plate dilutions (typically 500, 100 and 50 μL from the 5 mL of the resuspended conjugation) on 140 mm diameter Petri dishes with agar-M9 citrate with Km 50 μg/mL or Tel 80 μg/mL (selective media for exconjugant *P. putida* cells inserted with pTn5-*T7pol* or pTn*5*Tel [*T7→lux*], respectively). As negative controls, plate separately *E. coli* CC118λ*pir* (pBAM1), *E. coli* HB101 (pRK600), and the recipient strain on the same medium.
7. Pick exconjugant colonies directly from selective plates with a sterile toothpick and patch on M9 citrate Km 50 μg/mL and Pip 40 μg/mL. This procedure is to make sure that we do not select recipient cells that co-integrated the plasmid instead of undergoing an authentic transposition.
8. Pick a few (5–10) kanamycin resistant and Pip-sensitive colonies. Re-streak single colonies in M9 citrate plus Km 50 μg/mL.

3.2.3. Cloning of Sensor Modules into the Multiple Cloning Site of the pTn7-FRT Vector

Assembly of the sensor module is facilitated by arraying the functional parts (transcriptional regulator and its promoter vector, along with an outward-going cognate promoter) in pUC18Not (11), although it is possible also to directly clone or amplify an existing module as a *Not*I segment (22). One way or the other, the segment is inserted in the *Not*I site of pTn*7-FRT*. In our example, we retrieved the *xylR/Pu* sensor module of pBXPu (10) as a *Not*I fragment, which was then inserted in pTn*7-FRT* (see Fig. 1c) to generate pTn*7* [*FRT*-BX *Pu*].

3.2.4. pTn7-FRT Derivative Delivery by Tetraparental Mating Conjugation

pTn*7-FRT* derivatives carrying any sensor module of interest can then be mobilized to *P. putida* T7·LUX41 strain (see Fig. 3b) by a tetraparental mating in which the Tn*7* transposase is provided in trans with using the helper plasmid pTn*s1* (12) (see Note 2). As before, the following protocol was optimized for *P. putida*:

1. Aerobically grow overnight cultures of:
   - Donor cells: *E. coli* CC118λ*pir* carrying the formatted mini-Tn7 delivery vector that carries the module *xylR/Pu* (pTn*7* [*FRT*-BX *Pu*]) in LB with ampicillin 150 μg/mL and Gm 10 μg/mL at 37°C.
   - Helper cells: HB101 (RK600) in LB with chloramphenicol 30 μg/mL at 37°C.
   - *E. coli* CC118λ*pir* bearing pTns1, the plasmid carrying the Tn*7* transposase genes *tnsABCD*$^+$ in LB with ampicillin 150 μg/mL.
   - Recipient cells: *P. putida* T7·LUX41 in LB at 30°C.
2. Mix the four strains in a 1:1:1:1 ratio (e.g: 100 μL of each strain) in an empty test-tube and complete with 4.7 mL of 10 mM $MgSO_4$.
3. Vortex the mixture and pass through a Millipore filter disk (0.45-μm pore-size, 13-mm diameter) using a 10 mL sterile syringe.
4. Place the filter onto an LB-agar plate and incubate for a maximum of 8 h at 30°C (the optimum temperature for recipient cells).
5. With the help of flamed sterile forceps take the filter from the LB plate and introduce it into a test-tube with 5.0 mL of 10 mM $MgSO_4$ and vortex to resuspend cells.
6. Plate dilutions (in general 500, 100 and 50 μL from 5 mL mix of conjugation) onto M9 citrate with Gm 10 μg/mL (selective media for exconjugant *P. putida* cells) onto 140 mm diameter Petri dishes.
7. Pick colonies directly from transformation plates, with a sterile toothpick and double patch onto M9 citrate Gm 10 μg/mL, Pip 40 μg/mL and on M9 citrate Gm 10 μg/mL. This procedure is to rule out co-integrates (see above).

8. Select gentamicin resistance and Pip-sensitive colonies.
9. Re-streak isolated single colonies in M9 citrate plus Km 50 μg/mL and Gm 10 μg/mL.

3.2.5. Checking the Insertion of the Sensor Module into the Formatted Host Strain by Colony PCR

The mobilization of the pTn*7-FRT* carrying the sensor module into the formatted host bacterium can lead to three different transposition events (1) insertion in the *attTn7* site present in the chromosome of many Gram negatives (12) and placed usually in the terminator of the gene *glmS* (15), (2) insertion into the optimized *att*Tn7 entered into the chromosome upstream the T7 RNA polymerase by means of pTn*5-T7pol* (see above; see Fig. 2a), and (3) insertion in both loci i.e., the naturally occurring *attTn7* site(s) and the engineered site (see Note 3 below). To discriminate among the three possibilities two PCR reactions are made for each colony analyzed. One uses a sequence corresponding to the promoter of the sensor module as forward primer and T7pol2R that anneals in the position +831 of T7 RNA polymerase gene as reverse primer. The second PCR involves the same forward primer and a sequence annealing in the coding region of the *glmS* as reverse primer (see Note 4 below). The right colonies will be those positives for the first PCR reaction and negatives for the second one. This is because those clones that respond to such a diagnostic PCR have the sensor module inserted upstream the *T7pol* gene, they are amenable for deletion of antibiotic resistance genes and suitable for eventual activation of the regulatory cascade. By using this method we introduced the *xylR/Pu* sensor module into *P. putida* T7·LUX41 strain, generating the strain *P. putida* Km Gm-BX Pu *att*T7 LUX41 (see Fig 3b). The protocol below is optimized for *P. putida* T7·LUX41.

1. Pick single, large colonies from sixteen different gentamicin resistant and Pip-sensitive exconjugants. Re-streak the picked colonies on agar plates with M9 citrate plus Km 50 μg/mL and Gm 10 μg/mL to maintain a backup for each screened colony.
2. Pick fresh single colonies directly from a plate and transfer each colony directly to two different PCR reaction tubes.
3. From each colony perform two different PCR reactions one using the primers Pu1F/T7pol2R and other using Pu1F/PpuglmS2R. The PCR conditions are 5 min at 95°C, 25 cycles of 30 s at 95°C, 30 s at 55°C, 2 min at 72°C, and with a final extra 5 min at 72°C.
4. Analyze the PCR reactions by gel electrophoresis on a 1.5% agarose gel in TAE buffer (run at a constant 60 V). When using the Pu1F/T7pol2R primer pair a single PCR fragment of 1,632 bp should be observed if Tn*7-FRT* is inserted in the *att*Tn7 site present in the *landing pad* provided by the mini-Tn*5*. A single PCR fragment of 837 bp should be observed when

using the Pu1F/PpuglmS2R primer pair if Tn*7*-FRT is inserted in the natural *att*Tn7 site.

5. Perform the following steps using the gentamicin resistant and Pip-sensitive colonies that produce a PCR fragment of 1,632 bp using Pu1F/T7pol2R primer pair, and no amplification in the PCR performed with the primers Pu1F/PpuglmS2R.

3.2.6. Deletion of the Antibiotic Resistance Cassettes

After the insertion of the sensor module into the formatted host strain, antibiotic resistance cassettes are deleted using the ParA/*res* components of the multimer resolution system (*mrs*) of broad host-range plasmid RP4 (19). The location of the sensor module into the engineered *landing pad* leaves the antibiotic resistances genes used during the selection steps (Km and Gm) flanked by *res* sequences of the *mrs* in the appropriate orientation. Expression of the ParA protein mediates recombination between both *res* sequences leading to the loss of the intervening chromosomal DNA fragment. In the case of *P. putida*, expression of ParA is achieved with plasmid pJMSB8 (19). This plasmid can be conjugatively transferred to the target cells, but has a R6K origin of replication (20). This means that pJMSB8 can be passed to the recipient but it cannot be established because the new host lacks the replication π protein (23). Yet, expression of the resolvase during the short residence of pJMSB8 in the *Pseudomonas* recipient is sufficient to effect the predicted excision of the DNA segment without any inheritance of the *parA* gene itself (see Fig. 4a). By using this procedure we generated the *P. putida* BX *Pu att*T7·LUX41 (see Fig. 4b) strain. As before, the protocol below was set up for *P. putida*.

1. Aerobically grow overnight cultures of:
   - Donor cells: *E. coli* CC118λ*pir* carrying pJMSB8, the source of the ParA protein. Grow in LB with ampicillin 150 μg/mL.
   - Helper cells: *E. coli* HB101 (RK600) in LB with chloramphenicol 30 μg/mL at 37°C.
   - Recipient cells: *P. putida* Km Gm-BX Pu *att*T*7*·LUX41 in LB at 30°C.
2. Mix the three strains in a 1:1:1 ratio (e.g: 100 μL of each strain) in an empty test-tube and complete with 4.7 mL of 10 mM $MgSO_4$.
3. Vortex the mixture and pass through a Millipore filter disc (0.45-μm pore-size, 13-mm diameter) using a 10 mL sterile syringe.
4. Place the filter onto an LB-agar plate and incubate for a maximum of 8 h at 30°C (optimum temperature for recipient cells).
5. With the help of flamed sterile forceps, take the filter from the LB plate and introduce it into a test-tube with 5.0 mL of 10 mM $MgSO_4$ and vortex to resuspend cells.
6. Plate dilutions (make serial dilutions from the mix of conjugation to get single colonies) onto M9 citrate plates without antibiotics.

7. Incubate overnight and pick colonies directly from conjugation plates, with a sterile toothpick. Patch the same clone on M9 citrate Gm 10 μg/mL, M9 citrate Km 50 μg/mL, and M9 citrate without any antibiotic.
8. Re-streak clones sensitive to both kanamycin and gentamicin in M9 citrate without antibiotics.

3.2.7. Activation of the Biosensor by Deleting the Tn7R End of the Tn7 Transposon

After deletion of the antibiotic resistance cassettes, the last step to generate the release-able, optimized whole-cell biosensor is the enabling of the reporter system. In order to do this, it is necessary to delete of the Tn*7*R end between the promoter of the sensor module and the *T7pol* gene. This is affected by a site-specific recombination event between the two *FRT* sequences that are placed at both sides of the segment that causes transcriptional termination activity (see Fig. 4b). The yeast *FRT*/FLP site-specific recombination system is known to be functional in a broad host-range of host cells (16). In addition, the *scar* left after recombination is known not to produce polar effects in downstream genes (24). On this basis any DNA segment flanked by properly oriented *FRT* sites can be deleted through the transient expression of the yeast flippase (FLP). The expression of the FLP protein is achieved in this case by introducing into the formatted biosensor strain the *flp*$^+$ plasmid pBBFLP (see Fig. 5a; (10)). This is a mobilizable construct based on the pBBR1MCS series (25), able to replicate in *Pseudomonas* and encoding expression of the yeast *flp* gene under the control of a thermosensitive λ repressor (16). This plasmid is endowed also with the *sacB* marker from pFLP2 (16), thereby its loss can be selected by sucrose selection. The expression of the FLP protein inside the *P. putida* BX Pu *att*T7·LUX41 led us to obtain the fully functional biosensor strain *P. putida* BX Pu T7·LUX41 (see Fig. 4b). The complete process of enabling the biosensor strain includes three steps (1) introduction of the vector pBBFLP, source of FLP, (2) screening of functional strains, and (3) curation of pBBFLP. The corresponding protocol for *P. putida* follows.

3.2.8. Delivery of pBBFLP

1. Aerobically grow overnight cultures of:
   - Donor cells: *E. coli* DH5α (pBBFLP). Grow in LB with Tc 10 mg/mL.
   - Helper cells: *E. coli* HB101 (pRK600). LB with chloramphenicol 30 μg/mL at 37°C.
   - Recipient cells: *P. putida* BX *Pu att*T7·LUX41 in LB at 30°C.
2. Mix the three strains in a 1:1:1 ratio (e.g., 100 μL of each strain) in an empty test-tube and complete with 4.7 mL of 10 mM $MgSO_4$.
3. Vortex the mixture and pass through a Millipore filter disc (0.45-μm pore-size, 13-mm diameter) using a 10 mL sterile syringe.

4. Place the filter onto an LB-agar plate and incubate for a maximum of 6 h at 30°C.
5. Collect the filter from the LB plate, introduce it into a test-tube with 5.0 mL of 10 mM $MgSO_4$, and vortex to resuspend cells.
6. Plate dilutions (100 and 50 μL from the mix of conjugation) on agar plates with M9 citrate, Tc 10 μg/mL and incubate at 37°C (see Note 5 below).

3.2.9. Identification of Functional Strains

1. Pick colonies directly from conjugation plates with a sterile toothpick and streak out on plates with M9 citrate, Tc 10 mg/mL.
2. Pick fresh, well-isolated colonies directly from the plate and transfer them directly to different PCR reaction tubes.
3. From each colony, perform a PCR reaction using the pair of primers described previously (Pu1F and T7pol2R). The PCR conditions are: 5 min at 95°C, 25 cycles of 30s at 95°C, 30s at 55°C, 2 min at 72°C, and with a final extra 5 min at 72°C.
4. Analyze the PCR products by gel electrophoresis on a 1.5% agarose gel in TAE buffer (run at a constant 60 V). A single PCR fragment of 1195 bp should be observed when using the Pu1F/T7pol2R primer pair in those colonies that have deleted the Tn*7*R sequence because of the recombination between the *FRT* sequences. In the cases where the Tn*7*R sequence has not been deleted, the PCR fragment should be of 1,632 bp. In the successful instances, elimination of the Tn*7*R sequence allows read-through transcription from the *Pu* promoter into the *T7pol* gene (see Fig. 4b). One of the colonies is then taken to the next step of the procedure.

3.2.10. Curation of pBBFLP

Since pBBFLP contains the *sacB* counterselection marker that confers sucrose sensitivity to Gram-negative bacteria, the spontaneous loss of the plasmid can be selected by plating in a medium with sucrose. Almost all sucrose-resistant colonies may have been cured of pBBFLP, although in some few cases sucrose resistance may be attributable to *sacB* point mutations. These two possibilities can be distinguished by verifying if the sucrose-tolerant colonies are still resistant to Tc (the selection marker of pBBFLP).

1. Streak out the colony from the previous procedure (3.2.9) on agar-M9 citrate.
2. Grow the colony on the same liquid medium supplemented with 5% sucrose, followed by an overnight incubation at 30°C.
3. Plate dilutions of the overnight culture on a M9 citrate 5% sucrose plate.
4. Pick colonies with a sterile toothpick directly from the M9 citrate 5% sucrose plate and patch on plates either with M9 citrate 5% sucrose, Tc 10 mg/mL or M9 citrate 5% sucrose plate.

5. Select a sucrose-resistant and tetracyclin-sensitive colony.
6. Re-streak isolated single colony on agar-M9 citrate. The resulting clone is the concluding biosensor strain that were after.

3.3. Measuring Bioluminescence Production

Once the biosensor strain is in place, the next step is to verify its response toward aromatic compounds. Although the ultimate target in our case is BTEX, we can use 3MBA as proxy of the aromatic mixture, because this chemical is also a good XylR inducer. Production of bioluminescence can be recorded with different devices for detection of light emission. The procedure to capture the output of *P. putida* BX *Pu* T7·LUX41 when exposed to 3 MB is the following:

1. Grow the biosensor *P. putida* strain overnight in 2 mL of LB medium at 30°C, set in 10 mL tubes.
2. Dilute to an OD_{600} of 0.05 in 100 mL flasks and grow to $OD_{600} = 1.0$, at which point 3MBA is added at a concentration of 1.0 mM.
3. Place three replicates of 200 μL aliquots of the cultures (induced and not induced) in 96-well plates (NUNC).
4. Record the growth (OD) and light emission along with the time in a Victor II 1420 Multilabel Counter (Perkin Elmer). The bioluminescence output is calculated as the ratio between total light emission (in arbitrary units) divided by the optical density of the culture (OD_{600}).

4. Notes

1. pTn7-*FRT* replication is based on a ColE1 origin of replication; this fact implies that it cannot be used in bacteria where it can replicate. An alternative is to enter the sensor modules in the previously described vector pCAC5 (10), which replicates through a *pir*-dependent R6K origin (20). This makes the plasmid to be a suicide delivery vector in any host lacking the π protein.
2. Instead of the M9 citrate medium used here for *Pseudomonas*, other bacterial species may require different selective media that allows only the growth of recipient exconjugants.
3. If a high frequency of double insertions of the pTn7-*FRT* carrying the sensor module is found, it is possible to decrease the concentration of the Gm in the media to enrich for exconjugants with only one insertion.
4. For verification of insertions of the mini-Tn7 transposons derived from pTn7-*FRT*, it is recommended to employ a primer matching

a *glmS* coding region somewhat distant from the minimal sequence required for Tn*7* insertion *att*Tn*7* (15).

5. FLP production from pBBFLP is under the control of the temperature-labile λ repressor (16). It is thus recommended to plate the exconjugants after the introduction of pBBFLP at 37°C (yet, complete excision is often observed at 30°C as well).

Acknowledgments

Esther Fernandez is kindly acknowledged for her technical help. This work was defrayed by generous grants of the CONSOLIDER program of the Spanish Ministry of Science and Innovation, by the BACSINE and MICROME Contracts of the EU and by funds of the Autonomous Community of Madrid.

References

1. Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in *Escherichia coli, Nature 403*, 339–342.
2. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation, *Nature 434*, 1130–1134.
3. Ninfa, A. J., Selinsky, S., Perry, N., Atkins, S., Xiu Song, Q., Mayo, A., Arps, D., Woolf, P., and Atkinson, M. R. (2007) Using two-component systems and other bacterial regulatory factors for the fabrication of synthetic genetic devices, *Meth Enzymol 422*, 488–512.
4. Wu, C. H., Le, D., Mulchandani, A., and Chen, W. (2009) Optimization of a whole-cell cadmium sensor with a toggle gene circuit, *Biotechnol Prog 25*, 898–903.
5. Garmendia, J., de las Heras, A., Galvao, T. C., and de Lorenzo, V. (2008) Tracing explosives in soil with transcriptional regulators of *Pseudomonas putida* evolved for responding to nitrotoluenes, *Microbiol Biotech 1*, 236–246.
6. Tecon, R., Binggeli, O., and van der Meer, J. R. (2009) Double-tagged fluorescent bacterial bioreporter for the study of polycyclic aromatic hydrocarbon diffusion and bioavailability, *Environ Microbiol 11*, 2271–2283.
7. Fiorentino, G., Ronca, R., and Bartolucci, S. (2009) A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein, *Appl Microbiol Biotech 82*, 67–77.
8. Olaniran, A., Motebejane, R., and Pillay, B. (2008) Bacterial biosensors for rapid and effective monitoring of biodegradation of organic pollutants in wastewater effluents, *J Environ Monit 10*, 889.
9. Niazi, J., Kim, B., and Gu, M. (2007) Characterization of superoxide-stress sensing recombinant Escherichia coli constructed using promoters for genes zwf and fpr fused to lux operon, *Appl Microbiol Biotech 74*, 1276–1283.
10. de Las Heras, A., Carreño, C. A., and de Lorenzo, V. (2008) Stable implantation of orthogonal sensor circuits in Gram-negative bacteria for environmental release, *Environ Microbiol 10*, 3305–3316.
11. de Lorenzo, V., and Timmis, K. N. (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons, *Methods Enzymol 235*, 386–405.
12. Choi, K. H., Gaynor, J. B., White, K. G., Lopez, C., Bosio, C. M., Karkhoff-Schweizer, R. R., and Schweizer, H. P. (2005) A Tn7-based broad-range bacterial cloning and expression system, *Nature Meth 2*, 443–448.
13. Peters, J. E., and Craig, N. L. (2001) Tn7: smarter than we thought, *Nat Rev Mol Cell Biol 2*, 806–814.
14. Sanchez-Romero, J. M., Diaz-Orejas, R., and de Lorenzo, V. (1998) Resistance to tellurite as

a selection marker for genetic manipulations of *Pseudomonas* strains, *Appl Environ Microbiol 64*, 4040–4046.

15. McKown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988) Sequence requirements of *Escherichia coli att*Tn*7*, a specific site of transposon Tn*7* insertion, *J Bacteriol 170*, 352–358.
16. Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J., and Schweizer, H. P. (1998) A broad-host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants, *Gene 212*, 77–86.
17. Eberl, L., Kristensen, C. S., Givskov, M., Grohmann, E., Gerlitz, M., and Schwab, H. (1994) Analysis of the multimer resolution system encoded by the *parCBA* operon of broad-host-range plasmid RP4, *Mol Microbiol 12*, 131–141.
18. Ramos, J. L., and Marques, S. (1997) Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators., *Annu Rev Microbiol* 51:, 341–373.
19. Kristensen, C. S., Eberl, L., Sanchez-Romero, J. M., Givskov, M., Molin, S., and de Lorenzo, V. (1995) Site-specific deletions of chromosomally located DNA segments with the multimer resolution system of broad-host-range plasmid RP4, *J Bacteriol 177*, 52–58.
20. Wu, F., Levchenko, I., and Filutowicz, M. (1994) Binding of DnaA protein to a replication enhancer counteracts the inhibition of plasmid R6K gamma origin replication mediated by elevated levels of R6K pi protein, *J Bacteriol 176*, 6795–6801.
21. Herrero, M., de Lorenzo, V., Ensley, B., and Timmis, K. N. (1993) A T7 RNA polymerase-based system for the construction of *Pseudomonas* strains with phenotypes dependent on TOL-*meta* pathway effectors, *Gene 134*, 103–106.
22. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*, Cold Spring Harbor, New York.
23. Kolter, R., Inuzuka, M., and Helinski, D. R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K, *Cell 15*, 1199–1208.
24. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc Natl Acad Sci USA 97*, 6640–6645.
25. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd, and Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes, *Gene 166*, 175–176.
26. Herrero, M., de Lorenzo, V., and Timmis, K. N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria, *J Bacteriol 172*, 6557–6567.
27. Franklin, F. C., Bagdasarian, M., Bagdasarian, M. M., and Timmis, K. N. (1981) Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring *meta* cleavage pathway, *Proc Natl Acad Sci USA 78*, 7458–7462.

Chapter 18

Metabolic Engineering for Acetate Control in Large Scale Fermentation

Yong Tao, Qiong Cheng, and Alexander D. Kopatsis

Abstract

Escherichia coli is the most commonly used microorganism for production of recombinant proteins for different applications. Acetate accumulation during aerobic growth on glucose has significant negative impact on recombinant protein production in *Escherichia coli*. Various strategies, such as process and genetic approaches have been developed to limit acetate formation to increase the productivity of recombinant proteins. We developed a strategy to combine inactivation of pyruvate oxidase (*poxB*) and over-expression of acety-CoA synthetase (*acs*) in *E. coli* K strain for controlling acetate accumulation. A recombinant peptide was expressed and produced in the engineered strains with a very low acetate formation in a 10-L fermentation process.

Key words: Acetate accumulation, Recombinant protein, Pyruvate oxidase, Acetyl-CoA synthetase, Metabolic engineering, P1 transduction

1. Introduction

Excess carbon influx into *Escherichia coli* cells result in accumulation of acetate in the aerobic culture medium (1, 2). The rate of acetate formation is directly related to the rate of cell growth or the rate of glucose consumption. In a fed-batch process, *E. coli* generates acetate when the cells grow above a threshold growth rate regardless of availability of oxygen (3). Over-expression of recombinant proteins in *E. coli* depletes more intermediates from the central carbon metabolism, which further disrupts the balance between formation and consumption of pyruvate and acetyl-CoA. This results in more acetate accumulation in the recombinant strains (4).

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_18,

One approach to reduce acetate formation is by process modification (3, 5). Glucose-limited fed-batch process could be used to force cells to grow below the threshold specific growth rate. A consequence of limiting the growth rate of cells is that biomass is generated at a slower rate than the cells are capable of achieving. Alternatively, the produced acetate can be removed from the culture during the fermentation process by a dialysis method (6). But this method could generate significant issues when the fermentation process is scaled-up. The process modification for acetate reduction is extensively used in the industry. It is not the best solution because it undermines maximum growth and production capacity, not mention the complicated fed-batch algorism for exponential cell growth. Therefore, a genetic approach to minimize acetate formation would be a good alternative.

E. coli B strains generally generate less acetate than *E. coli* K12 derivatives because B strains have an activated glyoxylate shunt (4). We notice that the level of pyruvate oxidase B (*poxB*) expression is higher in K strain comparing to B strain, which leads to more acetate accumulation in K strain (4). We also notice that the acety-CoA synthetase (*acs)* gene expresses at very low level in K strain comparing to in B strain. Therefore, we developed a genetic strategy to combine inactivation of pyruvate oxidase (*poxB*) and overexpression of acety-CoA synthetase (*acs*) in *E. coli* K strain for controlling acetate accumulation.

2. Materials

2.1. Strain Construction

1. *PfuUltra*® II Fusion HS DNA Polymerase.
2. GeneAmp® PCR System 9700.
3. QIAquick PCR Purification kit (Qiagen, Valencia, CA).
4. Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).
5. Strain JW0855: BW25113 (*lacI*q *rrnB*T14 Δ*lacZ*WJ16 *hsdR514* Δ*araBAD*AH33 Δ*rhaBAD*LD78) plus Δ*poxB*, (Keio University, Yamagata, Japan).
6. P1*vir*.
7. pCP20 (7).
8. pKD4 (7).
9. pRedET1 (GeneBridges, Germany).
10. pCre-tet (GeneBridges, Germany).
11. Gene Pulser® (Bio-Rad, Hercules, CA).
12. Gene Pulser® Cuvette.

13. 20% L-arabinose.
14. P1 salts solution: 10 mM $CaCl_2$, 5 mM $MgSO_4$.
15. 1 M sodium citrate.
16. LB (Luria Broth) medium: 10 g/L tryptone, 5.0 g/L yeast extract and 5.0 g/L NaCl.
17. Appropriate antibiotics: ampicillin, tetracycline and kanamycin.
18. Fermetation seed medium: 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl.

2.2. Fermetation Tank Medium (Complex)

1. Presterilization part: 9 g/L KH_2PO_4, 4 g/L $(NH_4)_2HPO_4$, 1.7 g/L citric acid, 5 g/L yeast extract, 0.1 mL/L Biospumex 153 K. Adjust pH to 6.7 using 20% (v/v) phosphoric acid or 20% (v/v) ammonium hydroxide. Sterilize the contents of the aforementioned complex medium in-place, and add the following ingredients (2–4) in a sterile manner.
2. Post-sterilization part:0.6 g/L $MgSO_4.7H_2O$, 4.5 mg/L thiamine.HCl, 20 g/L Glucose (see Note 1).
3. Batch trace elements (stock solution; 100×): 840 mg/L EDTA, 250 mg/L $CoCl_2.6H_2O$, 1,500 mg/L $MnCl_2.4H_2O$, 150 mg/L $CuCl_2.2H_2O$, 300 mg/L H_3BO_3, 250 mg/L $Na_2MoO_4.2H_2O$, 1,300 mg/L $Zn(CH_3COO)_2.2H_2O$, 10,000 mg/L Fe(III) citrate. Filter sterilize (do not autoclave) the batch trace elements stock solution. Add 10 mL stock solution/L tank medium.
4. Uracil: 50 mg/L.

2.3. Feed Medium Composition

1. Glucose feed medium: 600 g/L Glucose (autoclave separately), 2.0 g/L $MgSO_4.7H_2O$, 10 mL/L feed trace elements (stock solution; 100×).

   $MgSO_4$ and feed trace elements filtered sterilized and then added, using sterile technique, into the (sterile) glucose stock solution. The composition of the Feed trace elements stock solution (100×) is as follows:
2. Feed trace elements stock solution (100×): 1,300 mg/L EDTA, 400 mg/L $CoCl_2.6H_2O$, 2,350 mg/L $MnCl_2.4H_2O$, $CuCl_2.2H_2O$]: 250 mg/L, 500 mg/L H_3BO_3, 400 mg/L $Na_2MoO_4.2H_2O$, 1,600 mg/L $Zn(CH_3COO)_2.2H_2O$, 4,000 mg/L Fe(III)citrate.

   Filter sterilize (do not autoclave) the feed trace elements stock solution.
3. Induction medium composition: 1 g/L L-arabinose, 1 g/L KH_2PO_4, 1 g/L $(NH_4)_2HPO_4$, 4.5 mg/L thiamine HCl, 10 mg/L uracil, 20 mg/L $CaCl_2.7H_2O$.

2.4. Dry Cell Weight Determination

1. 0.1 M, pH 7.0 phosphate buffer (1 L): 61.5 mL 1 M K_2HPO_4 solution, 38.5 mL, 1 M KH_2PO_4 solution.
2. Aluminum weighing pan.
3. Oven.

2.5. High Performance Liquid Chromatograph System

1. HPLC Model 1100 with ChemStation® software (Agilent, Santa Clara, CA).
2. Shodex SH 1011, 300×8 mm, with a guard column SH-G (Waters, Milford, MA).
3. HPLC reagents and standards.
   (a) Sulfuric Acid.
   (b) Pivalic Acid (used as an internal calibration standard).
   (c) Succinic Acid.
   (d) Fumaric Acid.
   (e) Citric Acid Monohydrate.
   (f) Potassium Phosphate Monobasic.
   (g) Glycerol.
   (h) Glucose.
   (i) Acetic Acid.

2.6. SDS-PAGE Analysis

1. 4× sample buffer: 424 mM Tris HCl, 564 mM Tris Base, 8% SDS, 40% Glycerol, 2.04 mM EDTA, 0.88 mM SERVA Blue G250, 0.7 mM Phenol Red, pH 8.5 (Invitrogen, Carlsbad, CA) Prepare 10 mL 1× sample buffer: 2.5 mL 4× sample buffer, 7.3 mL 8 M Urea, 84 µL 1 M DTT, 84 µL H_2O.
2. MES SDS Running buffer: 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3.
3. Nu-Page 4–12% Bis-Tris gel (Invitrogen, Carlsbad, CA).
4. Xcell SureLock Mini-Cell.
5. Simply Blue staining solution.

3. Methods

3.1. Construction of poxB Knockout in Chromosome

There are two pathways for acetate production in *Escherichia coli*: conversion of acetyl-CoA through phosphotransacetylase (*pta*) and acetate kinase (*ackA*) and conversion of pyruvate via pyruvate oxidase B (*poxB*). The level of *poxB* expression is higher in K strains comparing to B strains, which leads to more acetate accumulation in K strains (4). Thus the major approach to limit acetate formation

during fermentation of *E. coli* K strains is to directly knockout *poxB* gene in chromosome of K strains.

P1 transduction was used to transfer the *poxB* knockout genotype from the Keio collection into *E. coli* KK2000 (MG1655 containing *araBAD* knockout). Strain JW0855 in the Keio collection has *pox*B mutation constructed using Wanner method (7, 8). P1*vir* stock was prepared according to reference ((9), see Note 2).

3.1.1. Amplify Wild-Type P1 Lysate

1. Subculture 50 µL overnight culture of wildtype *E. coli* strain (such as MG1655) into 5 mL fresh LB medium containing 0.2% glucose and 5 mM $CaCl_2$, grow at 37°C with shaking.
2. After about 30–45 min incubation, Add 100 µL of a P1*vir* stock (titer less than 10^9 pfu/mL).
3. Incubate at 37°C with shaking for another 3 h, until the culture lyses.
4. Add a few drops of $CHCl_3$ and continue shaking a few minutes more.
5. Centrifuge the culture for 10 min at 9,200 × *g*, 4°C, collect the supernatant, store at 4°C as wild-type P1*vir* stock. The titer should be around 10^9–10^{10} pfu/mL.

3.1.2. Prepare poxB⁻ Kan^R P1 Lysate

1. Subculture 50 µL overnight culture of the strain JW0855 (see Note 3) into 5 mL fresh LB medium containing 25 µg/mL kanamycin, 0.2% glucose and 5 mM $CaCl_2$, grow at 37°C with shaking.
2. After about 30–45 min incubation, Add 100 µL of a recently prepared 10^9 to 10^{10} pfu/mL P1*vir* stock.
3. Incubate at 37°C with shaking for another 3 h, until the culture lyses.
4. Add a few drops of $CHCl_3$ and continue shaking a few minutes more.
5. Centrifuge the culture for 10 min at 9,200 × *g*, 4°C, collect the supernatant, store at 4°C.

3.1.3. Perform P1 Transduction

1. Inoculate 5 mL LB medium with the *E. coli* KK2000. Incubate overnight with shaking.
2. The following day, centrifuge 1.5 mL of the culture to pellet the cells.
3. Discard the medium and resuspend the cells at one-half the original culture volume in sterile P1 salts solution.
4. Mix 100 µL of the cells/P1 salts mixture with varying amounts of P1 lysate obtained in Subheading 3.1.2 (1, 10, and 100 µL). As a control, include a tube containing 100 µL of cells without P1 lysate.

5. Allow phage to adsorb to the cells for 30 min at room temperature.
6. Add 1 mL LB medium plus 200 μL of 1 M sodium citrate.
7. Incubate for 1 h at 37°C with aeration.
8. Centrifuge each culture to pellet the cells. Remove and discard the supernatant.
9. Suspend the cells in 50–100 μL of LB medium and spread the entire suspension onto LB plates containing 25 μg/mL kanamycin. Also spread 100 μL of the P1 stock on another plate. Incubate the plates overnight at 37°C.
10. The next day check the plates. The control plates with cells only or P1 phage only should be free of bacterial colonies. From the plate that shows growth from the least amount of P1 phage, pick single colonies, streak to isolate single clones, and incubate plates overnight at 37°C. Repeat this process a second time.

3.1.4. PCR Amplification to Verify poxB⁻ Mutation

1. Setup PCR mixture for each PCR reaction as following:

Distilled water (ddH_2O)	41.0 μL
10× PfuUltra® II Fusion HS DNA polymerase buffer	5.0 μL
dNTPs (25 mM each dNTP)	1.0μL
Forward primer (100 ng/μL)	1.0 μL
Reverse primer (100 ng/μL)	1.0 μL
PfuUltra® II Fusion HS DNA polymerase (2.5 U/μL)	1.0 μL
Total reaction volume	50.0 μL

2. Add the components in order while mixing gently into a thin-wall PCR tube.
3. Resuspend small amount of cells in the PCR tube. The forward primer was poxB-up: 5′-CCGTTAGTGCCTCCTTTCTC-3′. The reverse primer was kan-down: 5′-CATAGCCGAAT AGCCTCTCCACCCA-3′. Cover the PCR tube with a cap or an aluminum foil.
4. Put the tube into a PCR cycler with a preheated lid (GeneAmp® PCR System 9700), close the lid.
5. Perform PCR using the following cycling conditions: First, 95°C for 5 min Then, 35 cycles of 95°C for 30 s, 59°C for 30 s, 72°C for 2 min Last, 72°C for 10 min Hold at 4°C.

   The PCR product from the Kan^R *poxB*⁻ strain should be around 570 bp; whereas the PCR product from the wild type strain should be invisible.

3.1.5. Removal of the Kan^R Marker

The Kan^R marker flanked by the FRT sites was removed by FLP recombinase by transforming pCP20 plasmid into the *poxB*$^-$ strain. Plasmid pCP20 is temperature sensitive and contains thermal induced FLP recombinase that catalyzes homologous recombination between the two FRT sites. The ampicillin resistant pCP20 was transformed into the Kan^R *poxB*$^-$ strain constructed in Subheading 3.1.3.

1. Streak the Kan^R *poxB*- strain onto a LB plate containing 50 μg/mL kanamycin, incubate overnight at 37°C.
2. Inoculate a single colony into a culture tube with 3 mL of LB; incubate overnight at 37°C with shaking of 250 rpm.
3. Inoculate 1 mL overnight culture into a 500-mL flask containing 100 mL fresh LB medium, incubate at 37°C with shaking of 250 rpm for 3–5 h until the OD_{600} reach around 0.6.
4. Transfer the cultures into two 50 mL tube, chill on ice for 10 min, then centrifuge at 4°C with 6,000 × *g* for 10 min and disregard the supernatant.
5. Resuspend the cell pellets with 50 mL of prechilled water, leave on ice for 10 min, then centrifuge at 4°C with 6,000 × *g* for 10 min and discard the supernatant.
6. Repeat step 5 with 25 mL of prechilled water.
7. Repeat step 5 with 5 mL of prechilled 10% glycerol and combine together.
8. Resuspend the cell pellets with 1 mL of prechilled 10% glycerol; make 100 μL aliquots into prechilled 1.7 mL microfuge tubes.
9. Flash freeze the competent cells in liquid nitrogen and store at −80°C.
10. Thaw a tube of the competent cells on ice and prechill two 1.7 mL microfuge tubes.
11. Transfer 30 μL competent cells into each of the prechilled 1.7 mL tubes.
12. Add each of 1 μL pCP20 plasmid DNA into the competent cells, tap the tubes to mix.
13. Transfer the mixture of DNA and cells into the prechilled 1 mm eletroporation cuvettes, respectively.
14. Set the electroporator at voltage = 1.8 kV, Resistance: 200 (Ω) and Capacitance = 25 microFaradays (μF).
15. Electroporate the cells.
16. Immediately after electroporation, add 1 mL SOC medium into the electroporation cuvette and mix by pipetting.
17. Transfer cells into a culture tube; incubate at 30°C with 250 rpm shaking for 45 min.

18. Plate the transformation mixture (50 μL per plate) onto LB plates containing 100 μg/mL ampicillin.
19. Incubate the plates at 37°C for overnight.

3.1.6. Isolation and Confirmation of Kan^S Intergration Strain

1. Inoculate single colonies on the ampicillin plates obtained from Subheading 3.1.5, each into 1 mL fresh LB.
2. Incubate at 30°C with shaking for 1 h.
3. Switch to 37°C with shaking for another 3–5 h.
4. Use an inoculating loop to streak the culture on to LB only plates, incubate at 37°C for overnight.
5. Pick individual colonies from each plate onto three sets of LB plates with different antibiotics (LB/Amp 100 μg/mL, LB/Kan 25 μg/mL, and LB only).
6. Restreak colonies which grew on LB only plates but not on the other plates with antibiotics.
7. Check the colonies by PCR with primer poxB-up 5′-CCGTTAGTGCCTCCTTTCTC-3′ and primer poxB-down500: 5′-AAATTCCCATGCTTCTTTCA-3′ using a PCR protocol in the Subheading 3.1.4.
8. The correct clone should have a PCR product expected to be around 620 bp after removing the Kan gene.
9. Confirmation of the markerless knockout of the *poxB* gene is performed by sequencing using the poxB-up primer or poxB-down500 primer.

3.2. Overexpression of acs in the Chromosome of E. coli K Strain by Promoter Replacement

The *acs* gene encodes acetyl CoA synthetase which catalyzes an acetate assimilation reaction: coenzyme A + acetate + ATP →acetyl-CoA + diphosphate + AMP. Alignment of the *acs* gene from the *E. coli* K strain (MG1655) and the *E. coli* B strain (BL21) showed that the amino acid sequences are identical, although there are 17 bp silent differences at the DNA level. Microarray studies indicated that *acs* gene expresses at very low level in the K strain comparing to the B strain (4), which usually does not accumulate much acetate in fermentations. One approach to reduce acetate accumulation in fermentation of *E. coli* K strains is to overexpress the *acs* gene by replacing its native promoter with stronger constitutive promoters. This was accomplished in the *poxB* knockout of the K strain as described in Subheading 3.1.

3.2.1. PCR Amplify the Promoter and Kan^R Cassettes

Plasmid pDCQ702 contained the T5 promoter (10) downstream of the kanamycin resistance cassette flanked by two loxP sites. Plasmid pDCQ703 contained the *cat* gene promoter from pBHR1 (MobiTec GmbH, Goettingen, Germany) downstream of the kanamycin resistance cassette flanked by two loxP sites. Plasmids pDCQ702 or pDCQ703 were used as the template to amplify the

1,255 bp loxP-Kan^R-loxP-PT5 cassette or the 1,351 bp loxP-Kan^R-loxP-Pcat cassette by PCR.

1. PCR reaction is as same as in Subheading 3.1.4 by using primers listed below (see Note 4).

acsPdown-catR	5′-cgatgttggcaggaatggtgtgtttgt-gaatttggctcat TTTAGCTTCCTTAGCTCCTG-3′
acsPdown-T5R	5′-cgatgttggcaggaatggtgtgtttgt-gaatttggctcat ATGAGTTAATTTCTCCTCTT-3′
acsPup-loxKan5′	5′-gttgtctttaatcaattgtaagtgcatg-taaaataccact GTGCGTAGTCGTTGGCAAGC-3′

   For T5 promoter replacement, acsPup-loxKan5′ primer was used as the forward primer and acsPdown-T5R primer was used as the reverse primer with pDCQ702 as the template. For Pcat promoter replacement, acsPup-loxKan5′ primer was used as the forward primer and acsPdown-catR primer was used as the reverse primer with pDCQ703 as the template.
2. After PCR reaction, add 5 volumes of DNA binding buffer PB of the QIAquick PCR Purification kit to 1 volume of the PCR sample. Mix by vortex briefly.
3. Load the sample to the QIAquick spin column and centrifuge at maximal speed in the Eppendorf microcentrifuge for 30 s. Discard the flow-through.
4. Add 0.75 mL of DNA Wash Buffer PE to the column and centrifuge for 30 s. Discard the flow-through. Centrifuge the column for an additional 1 min.
5. Add 30 μL of distilled water to the column matrix. Place column into a new 1.5 mL microcentrifuge tube and centrifuge for 60 s to elute DNA.
6. Measure DNA concentration using Nanodrop 1000 spectrophotometer.
7. Choose “DNA-50” for the sample type selection column on the computer screen.
8. Pipet 1 μL distilled water onto the sample pedestal.
9. Press “blank” button on the screen. Wait until the calibration completes.
10. Open the arm and dry the surface of the pedestal with a lint-free lab tissue wiper.
11. Pipet 1 μL DNA sample onto the sample pedestal. Close the arm.
12. Press “measure” button on the screen. Wait until the measurement completes.
13. Record the concentration reading of the DNA sample.

3.2.2. Integration of the Promoter and Kan^R Cassettes into Chromosome by Lambda Red Recombination

The loxP-Kan^R-loxP-PT5 cassette or the loxP-Kan^R-loxP-Pcat cassette was integrated into the chromosome by Lamda Red technique ((7), see Notes 5 and 6). The *poxB* knockout strain containing the pRedET1 plasmid was used as the host. Plasmid pRedET1 (GeneBridges, Germany) encodes lambda Red recombinases that catalyze homologous recombination. The tetracycline resistant pRedET1 was transformed into the *poxB* knockout strain constructed in Subheading 3.1. The resulting strain designated as Red1012 was later used as the host for lambda Red recombination.

1. Transform pRedET1 plasmid into the *poxB* knockout strain using a standard protocol as described in Subheading 3.1.5. Resulting strain was designated as Red1012.
2. Transfer 130 μL overnight Red1012 culture into 13 mL fresh LB/tet5 medium, grow at 30°C with shaking. After about 2 h incubation, OD will reach around 0.3.
3. Add 200 μL 20% L-arabinose to final concentration of 0.3%.
4. Incubate at 37°C with shaking for another 1 h, stop growth and cool on ice.
5. Collect bacteria pellet with 15 min centrifuge at 6,000 × *g*.
6. Wash pellet twice with 12 and 8 mL cold ddH_2O.
7. Wash pellet with 1 mL cold 10% glycerol.
8. Resuspend the pellet with cold 10% glycerol to final ~100 μL competent cells.
9. Follow the standard electroporation protocol as in Subheading 3.1.5 by mixing 2 μL clean PCR products (about 0.2–0.3 μg/μL) obtained in Subheading 3.2.1 with 50 μL Red1012 competent cells, recover in 1 mL SOC at 37°C shaker for 3 h.
10. Plate on the LB plates containing 25 μg/mL kanamycin; incubate at 37°C overnight.
11. Colonies obtained from above were streaked on fresh LB plates containing 25 μg/mL kanamycine. PCR as described in Subheading 3.1.4 was used to confirm integration of the cassette upstream of the *acs* gene (see Note 7). Loss of the temperature sensitive pRedET1 plasmid was confirmed as sensitive to tetracycline.

3.2.3. Removal of the Kan^R Marker

The Kan^R marker flanked by the loxP sites was removed by Cre recombinase by transforming pCre-tet into the integrated strain. Plasmid pCre-tet (GeneBridges, Germany) is temperature sensitive and encodes Cre recombinase that catalyzes homologous recombination between the two lox recognition sites.

1. The tetracycline resistant pCre-tet was transformed into the Kan^R integration strain using a standard protocol as described in Subheading 3.1.5.

2. Inoculate single colonies obtained on the tetracycline plates each into 1 mL fresh LB. Incubate at 30°C with shaking for 1 h.
3. Switch to 37°C with shaking for another 3–5 h.
4. Use an inoculating loop to streak the culture on to LB only plates, incubate at 37°C for overnight.
5. Pick individual colonies from each plate onto three sets of LB plates with different antibiotics (LB/Tc10 μg/mL, LB/Kan25 μg/mL, and LB only).
6. Restreak colonies which grew on LB only plates but not on the other plates with antibiotics.
7. Check the colonies by PCR with primers acs-up/acs-280rev (see Note 7). The correct clone should have a PCR product expected to be around 600 bp after removing the Kan^R gene.
8. Confirmation of the markerless promoter replacement for *acs* gene is confirmed by sequencing using the acs-up primer or acs-280rev primer.

3.3. Construction of Peptide Production Strains

Standard transformation as described in Subheading 3.1.5 was used to transform the peptide production plasmid pLR538 into the chromosomally engineered *poxB* knockout combined with *acs* overexpression host as describe above. Plasmid pLR538 is ColE1-type of plasmid encoding a peptide gene HC415 fused to an inclusion body tag IBT139 (5C) expressed by the L-arabinose inducible promoter. The peptide production strain containing the *acs* gene overexpressed by the Pcat promoter was designated as QC1557. The peptide production strain containing the *acs* gene overexpressed by the T5 promoter was designated as QC1559.

3.4. Evaluation of the Strains in Fermentors

The performance of two engineered strains with inactivated pyruvate oxidase (*poxB*) and over-expressed acety-CoA synthetase (*acs*) gene under the Pcat promoter (QC1557), or under the PT5 promoter (QC1559), expressing the peptide construct IBT139(5C)-HC415, were tested in 10 L fermentation runs, against the "mother" strain (LR0940) with wild type *poxB* gene and native *acs* gene expression.

3.4.1. Seed Vial Preparation

1. Add a thawed seed vial from the "mother" lot, containing the genetically engineered strain of interest, into a 250-mL flask containing 120 mL of Luria Broth (LB) with appropriate antibiotics.
2. Grow the cells in a shaker incubator to an OD_{550} 1.0–1.1 for a targeted amount of time and at the optimum temperature (usually 37°C).
3. Prepare and label the sterile cryogenic vials.

4. Transfer 100 mL of the broth and 50 mL of 60% glycerol (sterile) solution into a second sterile 250 mL flask. The final glycerol concentration is 20%.
5. Mix the cells and glycerol well and aseptically transfer 1.5 mL aliquots to the (previously) labeled cryogenic vials.
6. Quickly freeze the new cryogenic vials into liquid nitrogen, or into a bath of dry ice/methanol, or dry ice/ethanol.
7. Store the frozen vials into a −80°C freezer.

3.4.2. Fermentation Medium Preparation

1. Sterilize the seed medium (see Subheading 2) and add, following sterile procedures, 125 mL into a 250-mL conical culture flask. Add the contents of a thawed seed vial of the engineered strain of interest into the culture flask, place it in a shaker incubator (37°C, 260 rpm) and grew the culture for 5–7 h to an OD_{550} of 2–3 U.
2. Prepare 6 L of fermentation medium, feeding medium and stock solutions (see Note 1) based on the composition listed in Subheading 2.
3. Several hours (5–7 h) after the inoculation of the seed flask with the seed vial of interest, the seed flask is sampled to measure the pH and OD_{550}.
4. When the optical density of the seed broth reaches OD_{550} 1–3, 124 mL of the seed broth (see Note 8) are added, using sterile techniques, into the prepared, sterile fermentation vessel, thus accomplishing the inoculation and the commencement the fermentation run. Inoculation of the fermentation vessel marks the beginning of the fermentation run and the commencement of the data acquisition program, which records all the key fermentation operating variables at a preset time frequency.

3.4.3. Peptide Production Scheme

1. Biomass build-up (batch & fed-batch phases; initial 18–22 h): after inoculation, the *E. coli* strain grows through the lag and early exponential phases, utilizing the residual glucose (~20–40 g/L), which is added before the onset of the batch phase (more see Note 9).
2. Induction and production phase (later 10–20 h): Having achieved a reasonably high level of cell density (28–30 g dry cell weight/L), production of the targeted recombinant protein commences by the addition of the induction medium. At the same time, the exponential glucose feeding is changed to a less aggressive schedule. This, in combination with the selected induction system, further increases the amount of biomass in the vessel (more see Note 10).

The results of a typical protein production scheme are outlined in Fig. 1.

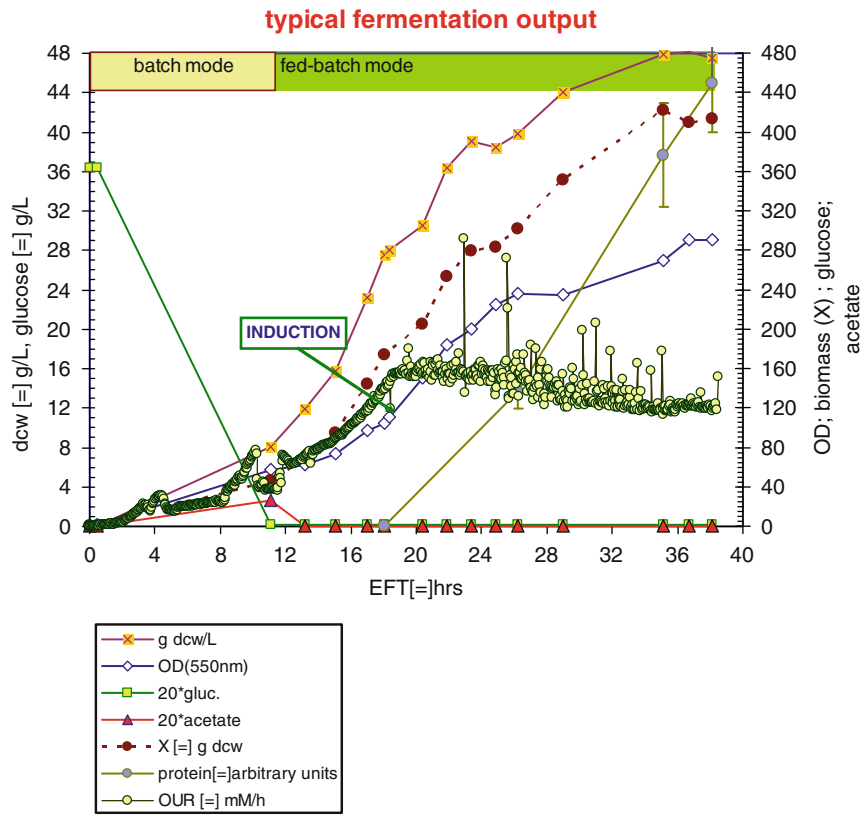


Fig. 1. An experimental outcome of a typical protein production run (10 L vessel). Biomass build-up phase (18–22 h): utilizing the residual glucose (~20–40 g/L), which is added before the onset of the batch phase. Induction and production phase (10–20 h): when cells reach to 28–30 g dry cell weight/L, production of the target recombinant protein commences by the addition of the induction medium.

3.5. Fermentation Sample Analysis

3.5.1. Dry Cell Weight Analysis

1. Dispense desired amount of fermentation broth into a 15-mL centrifuge tube.
2. Add equal volume of 0.1 M phosphate buffer to tube.
3. Mix well.
4. Centrifuge the mixture for 30 min at 7,500 × *g*, 10°C.
5. Discard the supernatant (clear liquid).
6. Add the same total volume of phosphate buffer to the 15 mL centrifuge tube containing the cell pellet (volume of broth + volume of buffer from first wash procedure).
7. Resuspend the cells and mix well either by vortexing or by mechanical means if the pellet is tightly spun (using a 1-µL inoculation loop works well).
8. Repeat steps 4 and 5.
9. Add half of the total volume used in step 2 above to the 15 mL centrifuge tube containing the cell pellet, this time using deionized water in place of phosphate buffer.

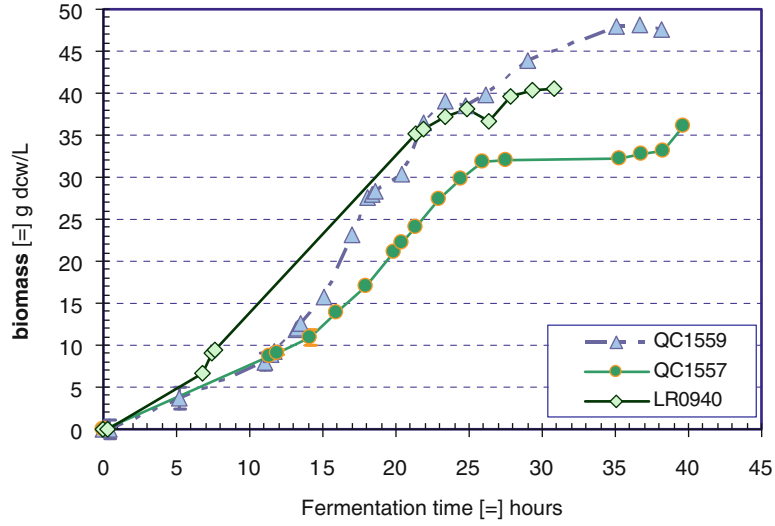


Fig. 2. Biomass growth of the two engineered strains QC1557, QC1559 and control LR0940 during the 10 L fermentation. The dry cell weight analysis was carried out every 2 h during the fermentation as described in Subheading 3.5.1.

10. Resuspend the cells and mix well either by vortexing or by mechanical means if the pellet is tightly spun.
11. Label and preweigh an aluminum pan for each sample and record tare weight.
12. Pour the pellet/water mixture in each tube into corresponding sample pan.
13. To ensure that all of the cell debris is rinsed from the tube, again add half the total volume of water to the tube, mix and dispense into pan.
14. Completely dry the pan & cells under vacuum @ 95°C with an N_2 purge (usually overnight).
15. Weigh pans again after drying.
16. Determine dry cell weight by subtracting pan weight after drying (step 14) from pan tare weight (step 11).

Biomass growth of the two engineered strains QC1557, QC1559 and control LR0940 during the 10 L fermentation is shown in Fig. 2.

3.5.2. Acetate and Organic Acids Analysis

1. Centrifuge the fermentation broth and use the supernatant for analysis.
2. Dilute, using an automatic pipettor, 1:1 using pivalic acid solution (e.g., 200 µL supernatant and 200 µL internal standard solution) (see Note 11).
3. Tightly cap and vortex.

4. Filter the mixture with 0.2 μm nylon filter and transfer to an HPLC vial.
5. Inject, 10 μL of each sample solution.
6. Run HPLC using following conditions:

Column	Shodex® SH1011 sugar column and SH-G guard column
Column temperature	50°C
Mobile phase	100% 0.010 N sulfuric acid in water
Flow rate	0.8 mL/min
Injection volume	10 μL
Detector wavelength (UV)	210 nm bandwidth 4 nm
Reference wavelength	350 nm bandwidth 80 nm
RI detector setting	40°C
Run time	40 min

Approximate retention time (min):

Compound	RT
KH_2PO_4	7.97
Citric acid	7.88
Glucose	8.60
Pyruvate	8.60
Succinic acid	10.55
Lactate	11.20
Glycerol	11.57
Formic acid	11.92
Acetic acid	12.86
Fructose	9.12
Pivalic acid	18.83

7. Data processing software is used that employs peak areas of each compound to determine the concentration of the samples. Standard solutions containing known amounts of each compound are first analyzed to create a calibration. The calibration points are constructed by calculating Amount vs. Response for each level in the calibration table. An equation for the curve through the calibration points is then calculated using linear least squares and the slope and intercept is determined. This equation is used to calculate the amount in the sample solutions.

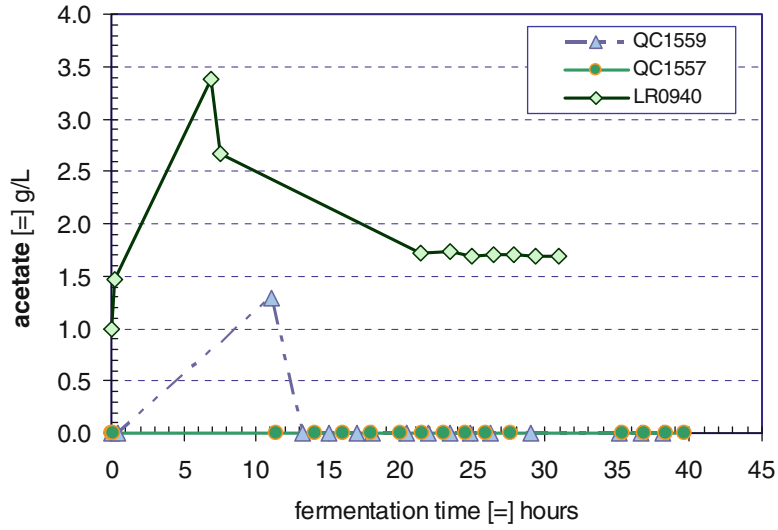


Fig. 3. Acetate accumulation from the two engineered strains QC1557, QC1559 and control strain LR0940 in 10 L fermentation. Acetate concentration of fermentation broths of three strains was analyzed for each time points based on the method described in Subheading 3.5.2.

Acetate accumulation from the two engineered strains QC1557, QC1559 and control strain LR0940 in 10 L fermentation is shown in Fig. 3.

3.5.3. Protein Analysis

1. Spin down 200 µL of the broth sample in a microcentrifuge at 14,000 rpm, 5°C for 5 min.
2. Resuspend the entire pellet in a volume of water to make 20 OD/mL. vortex sample to completely resuspend.
3. Add 5 µL of resuspended sample to 60 µL of 1× sample buffer.
4. Incubate 10 min. @ 95–100°C, gently shaking after 5 min. Do not vortex, BUT centrifuge lightly for a minute or so to remove any insoluble debris.
5. Load Nu-Page 4–12% Bis-Tris gel(s) into Xcell SureLock Mini-Cell.
6. Fill gel chamber with ~200 mL 1× MES running buffer, enough to completely cover the gel.
7. Load 1 cell well with 10 µL of protein ladder
8. Load remaining cell wells with 20 µL of prepared sample (about 0.033 OD cells, 3.3×10^7 cells each lane).
9. Fill outer chamber with ~600 mL 1× MES running buffer (enough to fill chamber ½ way)
10. Start power supply at 50 V.

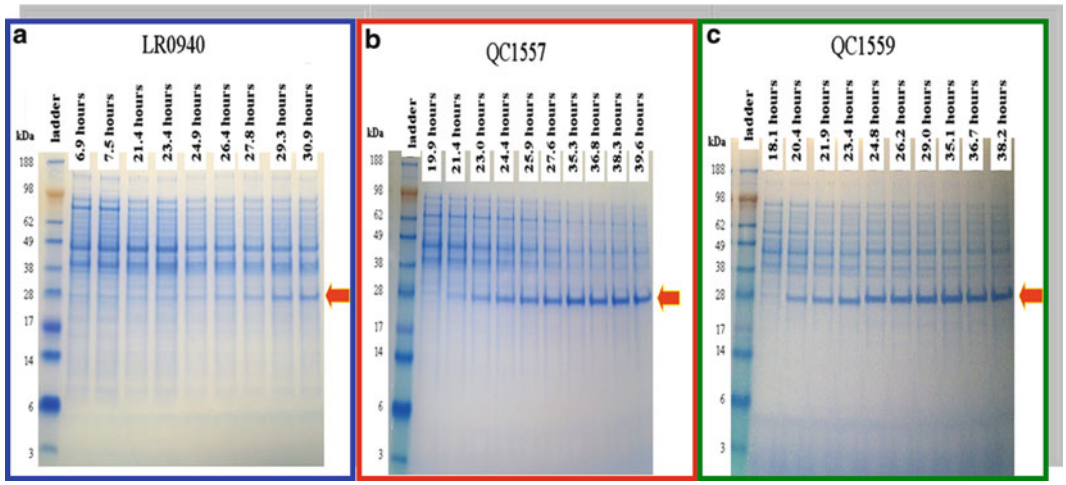


Fig. 4. SDS-PAGE showing the target protein (marked by the *arrow*) produced in control strain LR0940, and the two engineered strains QC1557 and QC1559. The SDS-PAGE analysis of the fermentation samples was carried out as described in Subheading 3.5.3. The time points were labeled in the top of each gel. The whole cell proteins from about 33 millions of cells were loaded at each time points. (**a**) Is for LR0940; (**b**) for QC1557 and (**c**) for QC1559.

11. Run for ~20 min or until an even band forms just below wells.
12. Increase power to 150 V and run ~30 min or until band is just above slit at bottom of gel.
13. Rinse gels, open gel cartridge and rinse gel itself with deionized water.
14. Remove gel from cartridge (must cut along bottom).
15. Soak gel in Simply Blue stain in pan with gentle agitation until bands are sufficiently developed (may take overnight for good development).
16. Remove gels from stain and rinse with deionized water.
17. Soak gels in water in pan with gentle agitation until background is clear (at least 3 h).

SDS-PAGE showing the target protein (marked by the arrow) produced in control strain LR0940, and the two engineered strains QC1557 and QC1559 is in Fig. 4. Relative protein production (see Note 12) from the two engineered strains QC1557, QC1559 and control strain LR0940 is shown in Fig. 5.

3.5.4. Summary of Performance of QC1557 and QC1559

Fermentation data of three strains indicated that two engineered strains with inactivated pyruvate oxidase (*poxB*) and over-expressed acety-CoA synthetase (*acs*) gene under the Pcat promoter (QC1557), or under the PT5 promoter (QC1559) performed significantly better in terms of protein production and acetate control. Although the biomass growth of three strains are between

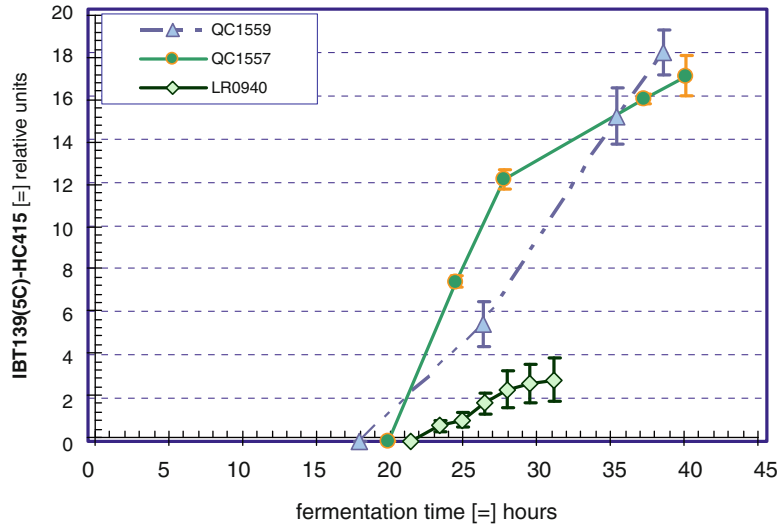


Fig. 5. Relative protein production from the two engineered strains QC1557, QC1559 and control strain LR0940. The relative and absolute quantities of the targeted protein were quantified using Agilent's 2100 Bioanalyzer™ unit and method as described in Note 12.

35 g dcw/L (QC1557) and 47 g dcw/L (QC1559) in 31–40 h fermentation processes (Fig. 2), two strains with *poxB* knockout and *acs* over-expressed (QC1557 and QC1559) produced much more target protein compared to control strain LR0940 (Fig. 4). QC1557 and QC1559 had more than sixfolds increase of target protein production compared to control strain (Fig. 5, 18 vs. 3 relative units/L). Although the same exponential glucose feeding schedule was applied for all three strains, the acetate level for control strain LR0940 is between 1.7 and 3.4 g/L during the fermentation. In contrast, strains QC1557 and QC1559 generated no detectable amount of acetate after 15 h of fermentation (Fig. 3).

In Summary, two engineered strains QC1557 and QC1559 have better acetate control and thus generate more target protein in 40 h in a 10-L fermentation.

4. Notes

1. While preparing the fermentation media, it's recommended to filter-sterilize the individual components and then sterilely add them, when dealing with mixtures of salts containing magnesium, phosphate and ammonium ions.
2. The P1 transduction was described very well in reference (9). As nonexpert, we used a virulent phage mutant, P1*vir*, to produce the lysate because the *vir* mutation prevents the phage from

generating P1 lysogens among transductants. The host genetic background, such as recombinant functions will play a significant role for a successful P1 transduction (9).

3. As the Keio collection of *E. coli* mutants is readily available for research communities and commercial parties through licensing (8), one could use *pox*B mutant from this collection for P1 lysate preparation. We suggest that one should use PCR to recheck all gene disruption sites to avoid chromosomal rearrangements, particularly in the case that *E. coli* recipient strain already has multiple mutations before P1 transduction. For others who have no access to the Keio collection, they can use a Lambda Red recombination protocol in Subheading 3.2 to incorporate poxB knockout in their strain background. The primer pairs for recombination can be:
   The 20 base nucleotides in lowercase matched to the flank sequences of kan^R cassettes and FRT sites in plasmid pKD4. The 40 base nucleotides in uppercase matched to the upstream or downstream flank sequences of *poxB* gene in chromosome.

poxB-kan-up	5′-CCGTTAGTGCCTCCTTTCTCTCCCA TCCCTTCCCCCTCCG tgtaggctggagctgcttcg-3′
poxB-kan-down	5′-TTTACCTTAGCCAGTTTGTTTTCGCCA GTTCGATCACTTC catatgaatatcctccttag-3′

4. In these acs primers, the 20 base nucleotides in upper case matched to the sequences of the cassettes. The 40 base nucleotides in lower case matched to the sequences of the *acs* promoter region.
5. It is important to use an *E. coli* strain with Δ*araBAD* mutation for higher Lambda Red recombination efficiency. Datsenko and Wanner reported that MG1655 would generate few transformants (7) because the wildtype AraBAD proteins would consume significant amount of L-arabinose needed for induction of the lambda recombinases.
6. It is critical to pay attention to the temperatures for Lambda Red recombination. The strain containing the pRedET1 should be initially grown at 30°C since the pRedET1 plasmid is temperature sensitive. After induction by L-arabinose, the strain should be switched to 37°C to activate expression of the Red recombinases. Following Lambda Red recombination, plates should be incubated at 37°C to isolate strains that have lost the pRedET1 plasmid.
7. To confirm the promoter replacement of *acs* gene by PCR, the forward primer was acs-up: 5′-tgctcgcccctatgtgtaa-3′; the reverse primer was acs-280rev: 5′-gtcgccttcccagatgatg-3′. The PCR

product from the wild type strain should be around 600 bp; whereas the PCR product from the integrated strain should be around 1.8 kb.

8. The 2-to-100 inoculum-to-fermentation volume ratio is typical of laboratory runs. This ratio should be kept similar while working on laboratory scale fermentation runs, typically, 1–100 L. As the fermentation volumes increase to pilot-scale values (several thousands to tenths of thousands liters), and to commercial scale runs, the seed train protocol needs to be adjusted accordingly.
9. Following the exhaustion of the residual glucose, the fed-batch phase starts via a scheduled exponential glucose feeding to further increase the biomass, while also restricting the total amount of glucose available to the cells. The aim is to restrict the growth of *E.coli* to low specific growth rates, approximately one-half to one-third of the maximum specific growth rate. Thus, the residual glucose concentration, throughout the fed-batch phase, is minimal (<0.5 g/L).
10. The biomass increases from 28 to 30 g dry cell weight/L before induction to 40–45 g dry cell weight/L at the end of the run, while increasing the protein cell-specific productivity (i.e., protein to biomass yield coefficient). In turn, the product concentration increases with time.
11. A known amount of Pivalic acid is used as an internal standard for calibration purposes. Pivalic acid is used as the internal standard due to its stability, its nonbiologic nature and the fact that it elutes away from the common, biologically produced (organic acid) compounds.
12. Protein quantification of the target protein, in Fig. 5, was accomplished by using Agilent's 2100 Bionalyzer™ unit, an automated protein gel-electrophoresis system. The BioAnalyzer™ unit produces electrophoresis "chromatograms" (electropherograms). Using the system's software, these electropherograms can be used to quantify the target protein, relative to the upper molecular weight marker. In addition, a fixed amount of a known protein can be used as an internal standard to quantify the targeted protein in absolute values, thereby calculating the target protein's concentration in the broth (titer).

Acknowledgment

This work was done at DuPont. Yong Tao was responsible for engineering the *poxB* knockout; Qiong Cheng was responsible for engineering the *acs* overexpression; Alexander Kopatsis was

responsible for fermentation evaluation. The authors thank Professor Barry Wanner for providing Lambda Red strains and plasmids used in this work. We also want to thank Timothy Schwartz, Qi Chen and Neal Perkins for their technical assistance.

References

1. Diaz-Ricci J.C., Regan L.and Bailey J.E.(1991) Effect of alteration of the acetic acid synthesis pathway on the fermentation pattern of *Escherichia coli*. *Biotechnol Bioeng.* **38**:1318–24.
2. Shiloach J. and Fass R. (2005) Growing *E. coli* to high cell density--a historical perspective on method development. *Biotechnol Adv.* **23**(5):345–57.
3. Eiteman, M.A. and Altman E. (2006) Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends in Biotechnol.* **24**, 530–536.
4. Phue, J.-N., and Shiloach, J., (2004) Transcriptional levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). *J. Biotechnol.* **109**, 21–30.
5. Lee, SY (1996) High cell-density culture of *Escherichia coli Trends in Biotechnology* 14, 98–105.
6. Nakano K., Rischke M., Sato S., and Märkl H. (1997) Influence of acetic acid on the growth of Escherichia coli K12 during high-cell-density cultivation in a dialysis reactor. *Appl Microbiol Biotechnol.* **48**(5):597–601.
7. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* using PCR products. *Proc. Natl. Acad. Sci. USA.* **97**, 6640–6645.
8. Baba, T *et al* (2006). Construction of *Escherichia coli* K12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* doi:10.1038/msb4100050.
9. Thomason LC et al. (2007). *E. coli* genome manipulation by P1 transduction. *Curr Protoc Mol Biol.*
10. Yuan, L. Z., Rouvière, P. E., LaRossa, R. A., and Suh, W. (2006) Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* **8**, 79–90.

Chapter 19

Minimization and Prevention of Phage Infections in Bioprocesses

Marcin Los

Abstract

Phage infections in bacterial bioprocesses constitute one of the most devastating threats to the productivity of the biotechnology facilities. There are several factors, which can decide if an infection would occur, and if it would turn into an outbreak and heavy contamination of the production facility. This issue is discussed on the basis of literature survey and experience of Phage Consultants.

Key words: Phage, Contamination, Fermentation, Bioreactor, Contamination prevention, Phage outbreak, Decontamination

1. Introduction

Bacteriophages are viruses, which are natural components of every community which contains bacteria. The estimated number of phage particles on Earth is about 10^{31} (1). They are in fact the most numerous entities on Earth. However, usually they have relatively narrow host range, which means that only small fraction of all bacteriophages can be dangerous for particular bacterial species. In practice, due to bacteriophage abundance, bioprocesses based on bacterial activities are in constant threat of bacteriophage infections. The infection by bacteriophage in fermentation facility may lead to destruction of bacterial cultures and subsequent contamination of the facility. This may in turn lead to paralysis of facility productivity, and thus to heavy financial loses. Thus the ability to prevent phage infection and to recover fast, if any infection already occurred, is a key element in all types of fermentation facilities which utilize bacteria for production. However, the optimal

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4_19,

way of contamination prevention and eradication strongly depends on the process and facility itself. In this chapter, literature survey and some of general remarks derived from work performed by Phage Consultants will be given.

2. Phage Sources

2.1. Facility Location

The fact that bacteriophages usually show relatively narrow host range also has an impact on their distribution in environment. Usually bacteriophages active against particular bacterial species are to be found in these environments which are habitats for the target bacteria. From the bioprocess point of view, it is the main concern when choosing new location for the facility. For example, fermentations based on one of the commonly used host—*Escherichia coli*—will be subject to the most danger when located near water treatment plants, rivers, lakes, and shores contaminated with municipal sewage, and in areas where human or animal manure is used as a natural fertilizer for soil (2, 3). It was also advised to keep them away from bird migration paths (3).

Some bacterial species used in fermentation are so widespread in environment, that it is almost impossible to find a relatively phage-free location, especially in places where infrastructure will be developed to serve as a fermentation facility. Among these bacterial species are mostly bacteria belonging to such genera like, e.g., *Bacillus*, *Clostridium*, or *Pseudomonas*. In such situations whole environment outside the facility is to be considered as a bacteriophage reservoir. In some cases, it is still possible to choose location supporting lower burden of specific bacteria and thus load of bacteriophage which can attack, e.g., by keeping away from farmlands when *Clostridium acetylobutylicum* will be used in fermentations, but this will not assure safety. Additionally, placing fermentation facility in remote location may affect the logistics of supplies and products, and thus it can impact the profit margin of the fermentation.

Use of the relatively rare bacterial species, preferably those from very rare bacterial genera, provides process safety advantage. This reduces the possibility that bacteriophage reservoir will be located near the facility, and the chances that phages will be introduced into the fermentation. Obvious disadvantage of such hosts is that they may not be very suitable for genetic engineering, and thus construction of production strains may not be feasible or economical.

2.2. Raw Materials, Equipment, and Bioprocess

Bacteriophages can infect the process using different entry ways. Sometimes it is not easy or even impossible to find out which measure of protection failed. The most common sources are those that are relatively hard to control. One of the major sources is air used to aerate fermentation or to keep fermentation overpressurized.

The main reason for that may be the presence of filtration membrane failures or the inability of membrane filters to fully stop particles of bacteriophage sizes. The performance of air filters may be dependent on the air moisture. Also, the ability of air filters to stop bacteriophage particles may be seriously affected during improper sterilization procedures or excessive usage of filter and resulting deterioration of filter condition (3).

Another possible source of bacteriophage in fermentation is raw materials. All measures of precaution would be worthless if media was not properly sterilized. As this task seems to be very easy in case of small volume fermentors, it is sometimes not simple to properly treat the media for hundreds- or thousands-cubic meters fermentors. Thus, the proper choice of sterilization method and equipment is crucial for successful phage contamination prevention. In case of large bioreactors sterilization of reactor itself may prove to be relatively difficult, especially when fermentors cannot be overpressurized and thus sterilization conditions used in autoclave cannot be achieved. In such situations combination of effective cleaning method, proper construction of fermentor, and adequate sterilization may considerably reduce the risk.

Water jacket was previously mentioned as a possible source of contamination of the process. The fermentation vessel integrity should be tested regularly and water used in water jacket should be assured to be phage-free (3).

Basically, all parts of bioreactors which come into contact with raw materials or bacterial culture itself are not easy to clean for mechanical reasons, and can be potential sources of phage outbreak. Tubes and valves are frequently culprits as they are hard to clean and are prone to temperature loss during sterilization. Special attention should also be given to sampling ports, as they may provide direct contact between content of bioreactor and outside world. Due to frequent manipulation, sampling ports may be easily contaminated during fermentor operation.

Inoculation is a very critical moment, as it is relatively easy to contaminate the fermentor if the procedure of inoculation or equipment used were not properly designed. Moreover this part of fermentation is most prone to operator errors.

2.3. Personnel

Despite the fact that the location of the facility in some cases can be a key factor in mitigating the risk of phage contamination, properly trained personnel is critical to the success of the bioprocess. Each person is a carrier of bacteriophages active against *E. coli*, and thus each operator is a potential source of phage infection and subsequent phage outbreak in the fermentation facility utilizing this host. As a result even perfectly located and planned facility cannot provide full protection without proper cooperation of employees. Considerable risk can be introduced by personnel by lack of proper precautions when designing the facility, by improper

choice of equipment, by poor design of the process itself (from the point of view of contamination security) including any of preparation, harvesting, and subsequent steps. One thing to keep in mind is that even the most skilled process microbiologists usually had little or no contact with bacteriophages during their courses and practices. They had their first opportunity to learn by trial-and-error method when contamination already occurred, and they have little chance to prevent phages from causing long-lasting and very expensive phage outbreak. From our experience, it is clear that improperly treated phage outbreak may paralyze facility productivity for up to 10 months. Although it is an extreme case recorded by us so far, much shorter periods of phage problems also cause large financial loses.

2.4. Bacterial Strains

To block phage infection, some authors recommend utilization of phage-resistant mutants (2, 3). This strategy is widely used in dairy industrial fermentation; however, in vast majority of cases it is very ineffective. There is a need for constant change of starter cultures used in dairy fermentations. The reason for that is there is almost unlimited variation of receptors utilized by phages to bind bacteria, and there are many counter-defense resistance mechanisms of bacteriophages. Introduction of any resistance mechanism in the host will rapidly result in selection of these phages, which can easily overcome the host resistance. Thus it is most important for long-term process protection, to work out proper work techniques and to use appropriate equipment in properly constructed facility. Production of resistant strains can be only an additional activity, and it cannot be a main approach to prevent a process contamination.

One of the most common myths is the ability of *tonA* mutants to prevent any T1-like phage from infecting the strain. From our practice, it became obvious that the vast majority of T1-like phages isolated from failed fermentation utilize different receptors, thus *tonA* (*fhuA*) mutants are not safe against T1-like phage infection. Moreover, majority of these phages which were sequenced or partially sequenced show much higher homology with Rtp phage than with T1. Rtp phage was also isolated from failed fermentation, and it most probably utilizes LPS as a receptor (4). It seems that the utilization of FhuA-TonA related infection mechanism by T1 and T5 among selected T-series phages was a coincidence, and thus no general conclusions can be made on protection against T1-like phages by *tonA* mutation.

Some bacterial strains show relatively high resistance against phage infections. The resistance mechanisms were reviewed by other authors (e.g., ref. (2) and references therein). They can be attributed to a few changes in bacterial cells like, e.g., cell wall alteration, receptor mutations, restriction modification systems, exclusion systems encoded by prophages, and other mobile genetic elements. Some mechanisms will work quite well on solid surfaces while in

liquid cultures they will be completely ineffective, e.g., spatial separation of phage from bacterial cell by secretion of large amounts of exopolysaccharides. Production of phage-resistant strains may help in prevention of phage infection by the same phage, which was used for selection of strain, or at best case a group of closely related phages. However, this cannot replace all the preventive measures of precaution, as subsequent phage infection by different phages can easily occur and cause outbreak, since the resistance never covers all possible phage types.

Another, quite frequently overlooked issue is the presence of prophages in production strains. Prophages can have very adverse effect on the strain performance. The effects range from occasional lysis of culture due to prophage induction or mutation causing virulence (5), through phage contamination and infection of other bacterial strains (6), to the contamination of fermentation by lysogenic conversion gene products (7). Prophages usually are present in media due to spontaneous induction and release from the host. They can be present in final product from copurification. Properties of some temperate bacteriophages make them quite hard to detect, and relatively easy to spread horizontally (6).

3. Phage Outbreak

3.1. Phage Primary Contamination

Phage outbreak in the facility starts with primary contamination. There are some basic methods, which can prevent it, or make it less likely (e.g., ref. (8)), but they will never prevent all contaminations (9). Phage primary contamination differs from subsequent infections as the source of phage is not located within the facility itself, and the phage may not be adapted yet to the host to give maximal potential phage yield. When it happens, it would be recognized only by the well-prepared personnel, and an outbreak would be prevented only by following well-prepared emergency procedures. However, surprisingly often no emergency procedures are in place, and the personnel have not undergone proper training. Thus primary infection in many cases has the potential to turn into a regular outbreak.

Symptoms of phage primary infection strongly depend on several factors such as virulence of the phage, what stage of fermentation phage contamination occurred, type of fermentation, properties of bacterial strain itself, and even factors like the mechanism of lysis utilized by the phage. Due to the fact that airborne phage can be introduced to the process relatively late, infected culture may not show any obvious signs of phage propagation, and thus fermentation may be considered successful. Also the phage may not be adapted yet to the host, and thus the phage growth may not be fast enough to lyse bacterial culture. However, in most cases the occur-

rence of phage contamination is relatively obvious, especially if phage was introduced early with inoculum or raw materials were not properly sterilized. In such situation phage development starts from the very beginning of fermentation, and thus even relatively unadapted phage could destroy bacterial culture. Sometimes action of bacteriophage can be partially masked by growth of resistant mutants or lysogenized bacterial cells. In such situation bacterial growth may restore after decrease of culture optical density (OD). In general, occurrence of fermentation failure with foaming and OD value drop should be always considered as a possibility of phage lysis, and in vast majority of cases phages are to blame for such failure. Subsequent actions after fermentation failure strongly influence the probability of occurrence of next infection. The most critical thing in phage primary infection is to prevent any phage release from the infected fermentor. As not all primary infections are very obvious to operators, all batches should be handled with care in order to prevent potential phage release. If any material from contaminated culture was released inside the facility, the probability of occurrence of next, secondary contamination would increase dramatically. In general, it is a good practice to keep aliquots of all fermentation samples stored at –80°C; it is very important to store samples from all suspicious fermentations to check later if any fermentation irregularities were caused by phages. In case of regular problems, on the basis of stored samples we could determine if the phage is a resident one, and thus we are dealing with secondary contamination, or the problems are caused by serial primary contaminations. This knowledge is very important, as the method of problem-solving is strictly dependent on phage origin.

The probability of occurrence of primary contamination depends on season. The various dependences were previously suggested by different authors. Primrose (2) suggested that dry weather facilitates phage spreading due to ability of dust particles with adsorbed bacteriophages to be transmitted by wind. Contrary to these observations, Bogosian (3) noticed that wet air makes air filters less effective. Moreover, the infections analyzed by this author occurred most frequently in autumn (October–November) and winter–spring (January–March). This pattern was suggested to be probably dependent on a few possible factors like seasonal increase in phage load in soil and sewage (10), planting and harvesting activities, or bird migrations. We observed additionally that wet monsoon season facilitates phage infection. Due to fact that there are no major bird migrations during wet monsoon, the other previously mentioned hypothesis seems to be more probable.

The most common bacteriophage type observed by us as a causative agent of *E. coli* fermentation failures is T1-like phages. They were previously reported by several authors (e.g., ref. (11)). It is hard to judge, whether they are the most common, or they simply cause serious problems that require external help to solve

the problems. Their legendary ability to resist desiccation, ability to infect nongrowing hosts, and short development cycles with very high burst sizes make them very dangerous in fermentation environment. The infections which they cause are relatively easy to recognize as they usually cause evident culture OD drop with extensive foaming, unless they occurred in very late stages of fermentation. These symptoms are accompanied with sharp increase in dissolved oxygen due to drop in respiration and increase of gas exchange caused by extensive foaming.

Second most common contaminants found in *E. coli* fermentations are phages belonging to Myoviridae. Their presence in fermentation was also reported previously (e.g., ref. (12)). They seem to be easier to eradicate and to prevent, as majority of them are not very resistant to desiccation. Other types of bacteriophages occur relatively seldom in *E. coli* fermentations. It is possible that some types of contaminations may underwent unnoticed, as not all of them are able to cause evident lysis. Similar pattern seems to be observed in other bacterial species—Siphoviridae and Myoviridae are the most common contaminants detected (13–17)

3.2. Phage Secondary Contamination and Phage Evolution

If phages from primary contamination were not prevented from spreading in facility, or equipment was not efficiently decontaminated, the occurrence of secondary infection would be highly probable. As we observed quite frequently, as well as others reported in the literature (18), phages quite often evolve in fermentation facility to improve their ability to produce higher burst sizes in shorter time, and to improve their ability to survive between infection cycles. The reason for that adaptation is obvious, when one take into consideration, that even a few minutes shorter development period and/or a few percent higher burst size after several infection cycles during one infection would make it possible for more efficient phage strain to produce much bigger amount of progeny phage when compared with the original strain. Probability of any mutation occurs among phage progeny from single contamination is very high, as the final titers of phage may reach as high as 10^{13}/mL (19). The mutations that confer adaptation advantages to the phage may be selected.

The most serious problems in case of secondary contamination is not the gradual adaptation of the phage to the host, it is the fact that the facility is contaminated. Even the smallest spillage may introduce billions of progeny phages to the facility environment, they immediately become dominant specie. If phages are resistant to desiccation, cleaning procedures are ineffective, and continued production despite occasional or frequent failures, phage load inside facility may increase even more considerably, decreasing the chances of successful fermentations.

As a result of the facility contamination, all subsequent fermentation failures are caused by phages which were released during

primary contamination. The fermentation irregularities caused by the same phage should be easier to recognize by operators. It is possible to implement and optimize detection methods, including process monitoring for phage presence.

3.3. Phage Development in Bioprocess

Growth of bacteriophages during fermentation is dependent on several factors. In feeding phase of fed-batch fermentation, phage development may be slower than in batch phase or in batch fermentations due to lower bacterial host resources, which can slow down the development and limit phage burst size. High bacterial density allows for much faster phage development by increasing the adsorption rate, as it minimizes the unproductive stage of phage life cycle when genome is encapsulated. Development of some bacteriophages may be slower if fermentation is conducted at temperature lower than usual for particular host, or when bacterial growth rate is slower (20). Also the type of phage itself may influence development. Phages that are able to grow on starved hosts (e.g., T1, T7) are less prone to slower growth of host cells, while development of other phages can be easily stopped by shutdown of feeding (21). Ability of phages to achieve very high growth rates in bacterial cultures is also dependent on the size of virion, which determines the amount of resources to be consumed to produce single progeny particle, and on host enzymatic machinery and many other factors. As a result, we observe relatively wide range of phage growth rates in fermentation, which span from 0.2 to 4 orders of magnitude per hour. Even the less effective bacteriophages are able to destroy fermentation when introduced early enough. For the most virulent phages, the complete lysis of the culture is achievable in just a few hours, even if introduced to the fermentation at late stages.

3.4. Phage Detection

Phage and prophage detection at first glance looks like relatively uncomplicated task. The most commonly used approach is double layer agar plates (22). However, the performance of this approach is strictly dependent on the skills of the personnel and the phage itself. When properly performed, the test can give outstanding results with excellent sensitivity and relatively rapid results. The time of development of plaques, which are indication of phage presence in this assay, strongly depends on phage itself and it may take from a few hours to 2 days. However, this method can give false negative results for some bacteriophages. Some ways of increase of the efficiency of phage detection using this method are already known, and they are highly recommended to use (23, 24).

Majority of alternative rapid detection methods are possible only for well-characterized phages, or at least they require proper preparations, since they utilize antibodies against phage (25) or amplification of phage nucleic acid (26). These methods are useful, when a facility suffers a recurrent infection by the same phage.

As it is the most common scenario, implementation of these methods should be always considered when facility suffers from phage-based production irregularities.

Prophages, often present in production strains, represent a different challenge than virulent phages. First of all, they are much harder to detect. Due to their life cycle, the lysogenic strain may behave perfectly normal, while producing vast amounts of phage particles. They are dangerous to other strains, which do not carry the same prophage. This is an important issue in facilities such as tollers that handle different strains in fermentation. One challenge in detection of prophages is the fact that there are many of them which cannot be easily induced using agents activating SOS response like, e.g., P2. Thus it is not a good practice in prophage detection to rely solely on mitomycin C or UV as induction agents in prophage detection. Moreover, when induced, sometimes temperate phages do not form plaques on double agar layer plates (23). So far, it is hard to predict, even on the basis of complete sequence, if given phage is active or will become active, and with some exceptions, what will be the triggering factor, which would cause induction of a prophage. This sometimes limits the possibility to cure a strain from prophages. To make things more complicated, some prophages offer benefits for the host, which can highly influence strain performance in fermentation (27–29). Thus, the decision on the deletion of prophage sometimes may cause unforeseen effects.

3.5. Phage Eradication

In general, the ability to clean up contamination and successfully eradicate phages is strongly dependent on facility construction and equipment installed. This issue should be considered at the stage of facility blueprints, but in practice it is usually neglected and makes future problems much more expensive and time-consuming to deal with.

First of all, possible phage contamination issues originated from future facility location, the facility plan, and instrumentation should be consulted with experts in phage eradication and prevention in order to prevent some obvious mistakes (9). The most common mistakes are connected with ventilation systems, location of various parts of facility, and lack of control elements to prevent phage spreading in facility. The location and construction of fermentation waste treatment stations, burden water reusage and treatment stations, and sewage systems are the most common sources of reinfection, even when bacteriophages seemed to be eradicated from the facility.

One most common approach in phage decontamination is the use of various disinfectants to clean up the facility and to kill bacteriophages. One must note that not all disinfectants are effective against phages, and various phages show varied susceptibility to a given disinfectant. Effectiveness of different disinfectants may be modified by types of contamination and various properties of

matrix to be disinfected such as type of surface, composition of media, presence of aggregates, solid and/or hydrophobic particles, etc. In general, disinfectants that were tested and effective on bacteriophage or adenoviruses should be chosen.

Another common approach in phage decontamination is heat treatment of equipment. It is important to remember that temperature is not the only important factor in heat treatment. One key element in sterilization is the energy transfer, which is most effective in water solutions, less effective in steam, and least effective in arid air. When dry heat is used for any equipment sterilization, extensive temperature and prolonged time should be used in order to achieve satisfying results. In case of liquids, temperatures as high as 80–90°C can be effective in phage eradication; however, the time of necessary exposure varies from phage to phage (30).

All attempts of phage eradication should be concluded with proper tests to detect possible phage presence despite sterilization efforts. Sampling should be focused on, but not limited to, areas where raw materials and bacterial cultures are prepared, and where biomass transfer to and from reactor/seed reactor(s) normally takes place. Regular testing programs should be implemented in order to detect problems before they become disasters. Another very important aspect, as also recommended by others (e.g., ref. (2)), is the development of emergency procedure that should be implemented immediately after contamination detection. The procedure should be triggered by phage detection or by observation of certain fermentation irregularities like drop in OD, extensive foaming, etc. At the moment, there is possibility to construct rapid, semiautomatic detection methods based on nucleic acid amplification; however, their usefulness may be limited to recurrent infections. Despite this fact, they can provide very effective help in eradication of phage from the facility.

Acknowledgments

This work was partially supported by the European Union within European Regional Development Fund, through grant Innovative Economy (POIG.01.01.02-00-008/08).

References

1. Whitman, W.B., Coleman, D.C., Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA*. **95**, 6578–6583.
2. Primrose, S. B. (1990). Controlling bacteriophage infections in industrial bioprocesses, p. 1–10. *In* J. Reiser (ed.), Applied molecular genetics. Springer-Verlag, Berlin.
3. Bogosian, G. (2006) Control of bacteriophage in commercial microbiology and fermentation facilities. In Calendar R, Abedon ST (Ed.), "The Bacteriophages. 2nd ed"., Oxford University Press, New York.
4. Wietzorrek, A., Schwarz, H., Herrmann, C., Braun, V. (2006) The genome of the novel

phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1 . *J. Bacteriol.* **188**, 1419–1436.

5. Bruttin, A., Brüssow, H. (1996) Site-specific spontaneous deletions in three genome regions of a temperate Streptococcus thermophilus phage. *Virology* **219**, 96–104.
6. Rotman, E., Amado, L., Kuzminov, A. (2010) Unauthorized horizontal spread in the laboratory environment: the tactics of Lula, a temperate lambdoid bacteriophage of Escherichia coli. *PLoS One* **5**:e11106.
7. Los, M., Kuzio, J., McConnell, M.R., Kropinski, A.M., Wegrzyn, G, Christie, G.E., (2010) Lysogenic Conversion in Bacteria of Importance to the food Industry in "Bacteriophages In the Control of Food- and Waterborne Pathogens". ASM press, Washington, DC, USA. 157–198.
8. Los, M., Czyz, A., Sell, E., Wegrzyn, A., Neubauer, P., Wegrzyn, G. (2004) Bacteriophage contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? *J. Appl. Genet.* **45**, 111–120.
9. Los, M. (2010). Contamination concerns. *European Biopharmaceutical Review*, **51**, 78–80.
10. Ogata, S. 1980. Bacteriophage contamination in industrial processes. *Biotechnol. Bioeng.* **22**(Suppl. 1), 177–193.
11. Wu,W.-W., Yoshinaga, K., Kanda, K., Kato, F., Murata, A., (1991). Phage S2, another new phage for serine-producing *Eschericha coli*. *Bull. Fac. Agr. Saga Univ.* **71**, 123–132.
12. Wu, W.-W., Tanaka, K., Kato, F., Murata, A., (1991) Phage S1, new phage for *Eschericha coli*. *Bull. Fac. Agr. Saga Univ.* **71**, 91–100.
13. Teuber, M., Andresen, A., Sievers, M. (1987) Bacteriophage problems in vinegar fermentations. *Biotechnol. Lett.* **9**, 37–38.
14. Koptides, M., Barak, I., Sisova, M., Baloghova, E., Ugorackova, J., Timko, J. (1992) Characterization of bacteriophage BFK20 from Brevibacterium flavum. *J. Gen. Microbiol.* **138**, 1387–1391.
15. Jones, D.T., Shirley, M., Wu, X., Keis, S. (2000) Acetone Butanol (AB) Fermentation Process. *J. Mol. Microbiol. Biotechnol.* **2**, 21–26.
16. Maeda, A., Ishii, K., Tanaka, M., Mikami, Y., Arai, T., (1986) KMl, a Bacteriophage of *Clostvidium butyvicum J. Gen. Microbiol.* **132**, 2271–2275.
17. Bartholomew, W. H., Engstrom, D. E.,Goodman, *S.* S., O'Toole, A. L., Shelton, J. L.,Tannen L. P. (1974) Reduction of contamination in an industrial fermentation plant. Biotechnol *Bioeng.* **16**, 1005–1013.
18. Josephsen, J., Petersen, A., Neve, H., Waagner, E. (1999) Development of lytic Lactococcus lactis bacteriophages in a Cheddar cheese plant. Int. J. Food Microbiol. 50, 163–171.
19. Seregant, K., Yeo, R.G. (1966) The production of bacteriophage m2. Biotechnol. Bioeng. 8, 195–215.
20. Los, M., Wegrzyn, G., Neubauer, P. (2003) A role for bacteriophage T4 *rI* gene function in the control of phage development during pseudolysogeny and in slowly growing host cells. *Res. Microbiol.* **154**, 547–552.
21. Los, M., Golec, P., Los, J.M., Weglewska-Jurkiewicz, A., Czyz, A., Wegrzyn, A., Wegrzyn, G., Neubauer, P. (2007) Effective inhibition of lytic development of bacteriophages lambda, P1 and T4 by starvation of their host, Escherichia coli. BMC Biotechnol. 7:13.
22. Adams, M.H. (1959). Bacteriophages. Interscience Publishers, New York, pp. 450–456.
23. Los, J.M., Golec, P., Wegrzyn. G., Wegrzyn, A., Los. M. (2008). Simple method for plating Escherichia coli bacteriophages forming very small plaques or no plaques under standard conditions. *Appl. Environ. Microbiol.* **74**, 5113–5120.
24. Lilehaug, D. (1997). An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. *J. Appl. Microbiol.* **83,** 85–90.
25. Los, M., Los, J.M., Blohm, L., Spillner, E., Grunwald, T., Albers, J., Hintsche R., Wegrzyn, G. (2005). Rapid detection of viruses using electrical biochips and anti-virion sera. *Lett. Appl. Microbiol.* **40**, 479–85.
26. Los,M., Los, J.M., Wegrzyn, G. (2008). Rapid identification of Shiga toxin-producing Escherichia coli (STEC) using electric biochips. Diagn. Mol. Pathol. **17**, 179–184.
27. Chen, Y., Golding, I., Sawai, S., Guo, L., Cox, E.C. (2005) Population Fitness and the Regulation of Escherichia coli Genes by Bacterial Viruses. *PLoS Biol.* **3**, 1276–1282.
28. Edlin, G., Lin, L., Bitner, R., (1977) Reproductive fitness of P1, P2, and Mu lysogens of Escherichia coli. *J Virol.* **21**, 560–564.
29. Lin, L., Bitner, R. Edlin, G. (1977), Increased Reproductive Fitness of Escherichia coli Lambda Lysogens J Virol. **21**, 554–559.
30. Pollard, E., Reaume, M. (1951) Thermal inactivation of bacterial viruses. Arch. Biochem. **32**, 278–287.

INDEX

Qiong Cheng (ed.), *Microbial Metabolic Engineering: Methods and Protocols*, Methods in Molecular Biology, vol. 834,
DOI 10.1007/978-1-61779-483-4, © Springer Science+Business Media, LLC 2012

H

I

L

M

N

O

P

R

S

T

Y