Synbiological systems for complex natural products biosynthesis

Jianhua Li a, Hailin Meng b, Yong Wang a,∗

a Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
b Bioengineering Research Center, Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou 511458, China

A R T I C L E   I N F O

Article history:
Received 17 August 2016
Received in revised form 24 August 2016
Accepted 24 August 2016

Keywords:
Natural products
Synbiological system
Synthetic biology
Biological parts
Systematic optimization

A B S T R A C T

Natural products (NPs) continue to play a pivotal role in drug discovery programs. The rapid development of synthetic biology has conferred the strategies of NPs production. Synthetic biology is a new engineering discipline that aims to produce desirable products by rationally programming the biological parts and manipulating the pathways. However, there is still a challenge for integrating a heterologous pathway in chassis cells for overproduction purpose due to the limited characterized parts, modules incompatibility, and cell tolerance towards product. Enormous endeavors have been taken for mentioned issues. Herein, in this review, the progresses in naturally discovering novel biological parts and rational design of synthetic biological parts are reviewed, combining with the advanced assembly technologies, pathway engineering, and pathway optimization in global network guidance. The future perspectives are also presented.

© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction ........................................................................................................... 221
2. Synthetic biological parts: screening, characterization and rational design .................................................. 222
  2.1. Discovery of novel BioParts .............................................................................. 222
  2.2. Rational design of BioParts ............................................................................. 224
3. Parts assembly and pathway engineering in chassis cells ............................................................................. 224
  3.1. BioParts assembly .......................................................................................... 224
  3.2. Pathway engineering in chassis cells ................................................................. 225
    3.2.1. Isoprene ................................................................................................. 225
    3.2.2. Amorphadiene ....................................................................................... 226
    3.2.3. Rebaudioside A ..................................................................................... 226
    3.2.4. Compound K ......................................................................................... 226
    3.2.5. Rare ginsenosides ............................................................................... 226
    3.2.6. Transporter engineering ....................................................................... 226
4. Computer-aided guidance for systematic optimization .............................................................................. 227
5. Perspective ............................................................................................................ 227
   Acknowledgement ............................................................................................... 228
   References ............................................................................................................. 228

1. Introduction

Natural products, produced by bacteria, fungi, and plants, play a highly significant role in the drug discovery and development

* Corresponding author.
E-mail address: yongwang@sibs.ac.cn (Y. Wang).
process [1]. Many of them, such as paclitaxel [2], digitalis [3], codeine [4], and erythromycin [5], are highly concerned with astonishing biological activities and are used for diverse purposes such as anticancer, congestive heart failure treatment, alleviating pains, and antibiotic, respectively. Unfortunately, these desired products are commonly isolated from native organism in low yield or synthesized with inefficient or unfeasible because of their complex structures [2,5].

Synthetic biology provides an alternative approach to produce these valuable products for industry application [6,7]. It aims to reduce the complex biosynthetic systems in originated organism to a simplified, reliable, quality-controlled heterologously artificial biological network to achieve our special goals. The biological parts, devices, or modules related to the targeting biosynthetic pathway are assembled and transplanted from the natural host into a genetically tractable host system such as Escherichia coli, Saccharomyces cerevisiae.

The classical engineering strategies have been widely provided for our purposes in this approach. For example, overexpress enzymes responsible for putative bottleneck steps in biosynthetic pathway [8], delete competing steps [9], transfer biosynthetic machinery to an amenable heterologous host [10], re-regulate regulatory circuits to awaken unknown natural compounds [11], or even create enzyme variations by domain shuffling [12]. Nevertheless, there is still a challenge when integrating a heterologous pathway for complex natural products in chassis cells for overproduction purpose due to the limited characterized parts, modules incompatibility, and cell tolerance towards product. Thus in this review, we will discuss the strategies to deal with the problems mentioned above for products production.

2. Synthetic biological parts: screening, characterization and rational design

Biological parts (BioParts), including promoters, terminators, ribosome binding sites (RBS), and protein coding sequences (CDSs), etc. are the basic building blocks widely used in synthetic biology (Fig. 1). [13] The characterized biological parts can be designed and grouped into a biological device at large-scale to obtain specific biological functions in an engineered cell. With the development of genome/transcriptome sequencing and bioinformatics, huge numbers of novel biological parts can be precisely predicted [14–29], and many of them were functionally characterized and standardized to build up open access libraries by academy application, e.g., the Registry of Standard Biological Parts established by Massachusetts Institute of Technology (http://partsregistry.org) and the Joint BioEnergy Institute Inventory of Composable Elements (JBEI-ICEs, http://www.jbei.org) set up by Jay D. Keasling’s group (Table 1). [30–35] However, only very limited biological parts have been explored to date. Take microorganism for example, only 1% of total resource can be cultured by current technology. Traditional culture methods result in an extremely low efficiency in exploration of natural biological element. As to plant, the problem becomes more prominent: much larger genome but trace has been sequenced, enormous novel parts are still cryptic and eager to be discovered and characterized. In terms of announced biological parts, it is quite necessary to perform manual construction based on the existing knowledge and make full use of the library resources of natural elements.

2.1. Discovery of novel BioParts

In most case, BioParts can be obtained directly from native organism. And among all of the BioParts, genes encoding the biosynthetic enzymes play a prominent role. Serious efforts are under way for the discovery of novel catalytic parts continuously, though it is not a trivial work. Recent advances in genome sequencing and bioinformatics have shown that in bacteria and fungi, the genes responsible for the assembly of a metabolite are generally clustered together within the chromosome [3,36–38].

But the metabolites derived from plant are formed in much more complex condition, the genes exist in a cassette with
separated promoter and terminator, and these genes are not always cultured in genome. Generally, transcriptome data provided useful information for the desired gene mining. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. Thus, the simplest idea behind the metabolic engineering has been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. The plants UGTs contain a highly conserved consensus signature sequence called the PSPG motif. With the aims to elaborate the biosynthetic pathway of ginsenosides, which were further used in the metabolic engineering, have been to simply redirect the metabolites by altering genetic programs. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools. These genes are interchangeable units in separated organisms and the equivalent genes may be discovered by bioinformatic tools.
Colletotrichum gloeosporioides TA67 were capable of producing taxol which were validated by HPLC-MS. Thanks to the relatively small genomes and metabolically genetic cluster, these microorganisms provided a fascinating reservoir for the diversity of taxol biosynthetic genes, thus could significantly expand the number of known taxol biosynthetic genes to elucidate the whole pathway and provide the basis for heterologous production.

2.2. Rational design of BioParts

Generally, there are two available main approaches for manual construction of library according to the design principle: the first one is based on random mutagenesis and library screening, and the other one is artificial design based on quantitative prediction models [42]. Random mutagenesis can be easily introduced via error prone PCR by slightly adjusting the reaction conditions (e.g., different concentration of magnesium and manganese ions, unbalanced the concentration of dNTPs, or DNA polymerase with low fidelity). However, it is still a laborious work with low efficiency to acquire a controllable and quantitative biological part from different kinds of libraries for complex biological systems design. So, it is a strong demand for regulatory element from random different kinds of libraries for complex biological systems design.

So, we can get a controllable and quantitative part from random different kinds of libraries for complex biological systems design. Therefore, it is a strong demand for regulatory element from random different kinds of libraries for complex biological systems design.

More recently, great advances in synthetic biology have brought back the sequence-activity modeling to the forefront. For instance, Rhodius et al. [43] scored various motifs of E. coli c^ binding promoters and correlated promoters scores based on position weight matrix (PWM) models. Salis et al. [44] targeted translation initiation process and developed an equilibrium statistical thermodynamics model for designing synthetic RBSs. De Mey et al. [45] established a correlation between the entire sequence and strength by applying partial least squares (PLS) regression method. However, building precise models that can predict the activity of regulatory elements and quantitatively design elements with desired strength is still a challenge.

As a powerful machine learning algorithm, artificial neural network (ANN) simulates the structure and functional aspects of human brain neural network. The weight of neuron connection can be changed to a suitable value after learning knowledge from training data set. To predict the correlation of expression strength and regulatory sequences [46], a precise ANN model was built in our lab to predict the strength of Trc promoter & RBS elements with a high regression correlation coefficient of 0.98 for both model training and test. Furthermore, the quantitative design sequences with desired strength were also successfully applied to improve the expression of a small peptide Bmk1 and fine-tune a key enzyme gene des for pathway engineering of terpenoids biosynthesis in E. coli. This research demonstrates that the methodology based on machine learning models can de novo and quantitatively design regulatory elements with desired strength, which are of great importance for synthetic biology applications.

As to natural protein coding sequences, it is crucial for metabolic engineering a target in biosynthetic biology. For several decades, biosynthetic scientist assembled the native coding sequences to create the genetic pathway in a heterologous host. But the titer is always remarkable low. Codon optimization, modifying the coding sequence of an individual gene without altering the amino acid sequence, are widely used in heterologous cells in order to increasing the yield of protein expression under the control of a given promoter. It has been known that the choice of synonymous codons in many species is strongly biased and that a correlation exists between high expression and the use of selective codons in a given organism. Additionally, mutagenesis was operated by computational chemistry to model binding sites and active sites with key residues mutated to confer altered enzymatic activity or high-level expression. However, the 3D structure is not always easy to obtain leading to the modification of proteins are not always highly effective. And the mutations were operated basing on semi-empiricism such as terminal truncation, domains replacement. Synthetic biologists are still working on designing the most efficient enzyme for a specificity engineering pathway.

3. Parts assembly and pathway engineering in chassis cells

For the purpose of synthesizing a desirable natural product in chassis cells, the biosynthetic pathway need to be totally designed, assembled and engineered. According to the synthesis pathway and the character of module parts, many different microbes, Escherichia coli, Saccharomyces cerevisiae, and Myxococcus xanthus has been used as the chassis cells for production of many natural products successfully. It is a great advance that the interchangeable biological parts are used and assembled quickly in synthetic biology, which had to be standardized previously. Biological parts, or devices are not independent objects, they typically function within a cellular environment. When these artificial parts were introduced into the cells, drastic influence on cells’ physiology would be appeared. It this section, we will describe the parts assembly and the pathway engineering in chassis cells.

3.1. BioParts assembly

Synthetic biologist seeks to engineer new function in biological system for variety of applications. A key limitation in synthetic biology is the time taken to assemble genes or other DNA parts into new devices, pathways and systems, or to alter these assemblies for recycling purpose. Considerable efforts have been invested into developing new tools for DNA assembly over past four decades. At present, popular DNA assembly methods are mainly based on four disciplines: restriction endonuclease based assembly, homology-dependent assembly, site-specific recombination strategies, and homing endonucleases based assembly technology. (summarized in Table 2, Fig. 1).

One of existing assembly methods rely on restriction enzymes digestion followed by ligation to join DNA segments together. Although this approach works well for the insertion of a single DNA sequence into a vector, it is often hard to find enough distinct restriction sites for the cloning of multiple DNA fragments. BioBrick and BglBrick, are two standardized techniques in the process of DNA assembly by using standard restriction sites [47–49]. Biological parts were flanked by four restriction sites and the assembly products of the first round can subsequently be used in the next round, thus permitting the assembly of multiple parts. This approach has proven to be very effective while consuming as well as forbidden sites limitation especially for multiply sequences assembly. Golden gate assembly is an advanced approach, which uses the activity of type II restriction enzymes (eg., Bsa I and Bsm BI) to cut outside of their recognition sequences to expose designer overhangs. Multiply DNA sequences assembly can be enabled in one-pot reaction with designed direction seamlessly [50,51].

Another class is homology-dependent assembly in vitro or in vivo that work by joining DNA fragments with homologous sequences, usually between 20 bp and 40 bp at the end. The long sequence homology ensures high efficiency and specificity of DNA assembly, meaning that most long-overlap-based methods can easily assemble five or more DNA parts together in one step. These methods are particularly popular for manipulating larger DNA
perform protein engineering of the present Cpf1s to enhance the assembly. Therefore, it is very important to some target sequences, cleavage nearby the two sites could also be downstream of the protospacer adjacent motif (PAM) site. But, for a class 2 type V CRISPR-Cas systems protein FnCpf1[60]. It mainly stand, was developed recently with a crRNA-guided endonuclease, gously expressed. Another DNA assembly stand, namely C-Brick sequences (sequences, there is usually no need for modi
quences, as there is usually no requirement for the removal of restriction sites from within parts. These include SLIC (sequence and ligase independent cloning) [52],In-Fusion technologies [53],the Gibson isothermal assembly method [54],CPEC (circular polymerase extension cloning) [55],LCR (ligase cycling reaction) [56]and so on. The mechanism of these methods varies greatly. For example, CPEC is based on overlap extension PCR and is essentially a high-fidelity PCR amplification, infusion technology is worked in the present of recombinase in vitro. Site-specific recombination omits any need for restriction endonucleases and instead uses phage integrases, which are site-specific recombinases that recognize versions of attachment (att) sequence motifs and catalyse DNA rearrangement between them. Integrases are utilized in the popular commercial Gateway cloning method, which uses λ integrase in vitro to catalyse directional cloning of DNA parts that are flanked by orthogonal versions of the attB and attP sites recognized by the integrase. Similar non-commercial systems have also been developed that use alternative phage integrases. Zhang et al. [57]efficiently, accurately, and tandemly assembles epothilone biosynthetic gene cluster by Streptomyces phage φBT1 integrase. Colloms et al. [58]assembled functional carotenoid biosynthetic pathways containing three, four or five genes pathways in E. coli by bacteriophage φC31 integrase.

Homing endonucleases, which recognize long DNA sequences is newly developed DNA assembly technology. A new standard iBrick, which uses two homing endonucleases of I-SceI and PI-PspI was introduced by Liu et al. [59] Both enzymes recognize long DNA sequences (>18 bps), which in extremely rare in natural DNA sequences, there is usually no need for modification of the DNA sequences regardless of their length. Using this standard, the carotenoid biosynthetic cluster was successfully assembled and the actinorhodin biosynthetic cluster was easily cloned and heterologously expressed. Another DNA assembly stand, namely C-Brick stand, was developed recently with a crRNA-guided endonuclease, a class 2 type V CRISPR-Cas systems protein FnCpf1 [60]. It mainly cleaves target DNA sequences with the “18–23” cleavage pattern downstream of the protospacer adjacent motif (PAM) site. But, for some target sequences, cleavage nearby the two sites could also be detected, which cause the lower assembly efficiency and incorrect assembly. Therefore, it is very important to find new Cpf1s or perform protein engineering of the present Cpf1s to enhance the accuracy of Cpf1 digestion in future.

3.2. Pathway engineering in chassis cells

A module is a compartmentalized set of parts/devices with interconnected functions that performs complex tasks. Generally speaking, in a cell, modules are specific pathway, such as a metabolic pathway or a signal transduction pathway. However, the optimization and balancing of multigene pathways are a challenge when introduced into a heterologous cell. Multiple rounds of construction, debugging, and fine-tuning are needed for chemical improvement, which is inefficient and time-consuming. Therefore, module pathway engineering has emerged as a promising strategy to solve this problem, which artifically divides metabolic pathway in to various modules, constructs artificially controlled modules with various expression level, and assemblies of multiple modules simultaneously for generating strain library.

Biologist always divides a metabolic pathway into at least two modules-upstream module and downstream module, and the different function module were assembled into separated vectors. The former produce the primary metabolites, which is the mainly the precursor of target chemical and act as the reactant for the downstream action. And the upstream module might present in the chosen chassis. Typically the flux toward the product is naturally low but through the use of classical strain improvement or the use of directed genetic modifications, it is possible to increase the flux toward the product and it is the routine strategy for high yield. Many other strategies have also been employed for modules engineering including: (1) Over-expression of the limited step enzymes; (2) Deletion or down regulation of bypass genes; (3) Promoters engineering; (4) Interchange of natural proteins from range of livings; (5) Modification of rate-limiting enzymes; (6) transporter engineering. In this section, I will introduce several cases for natural product pathway engineering with the strategies above mentioned (Table 3).

### Table 2

| Classes                      | Standards | PCR required | Forbidden restriction sites | Multipart assembly | Scar | References |
|------------------------------|-----------|--------------|-----------------------------|--------------------|------|------------|
| Restriction and ligation     | BioBrick  | No           | 4                           | No                 | Yes  | Knight [47]; Shetty et al. [48] |
|                              | BglBrick  | No           | 4                           | No                 | Yes  | Anderson et al. [49] |
|                              | Golden gate assembly | Yes | 1                           | Yes                | No   | Engler et al. [50] |
| Homology-dependent assembly  | In-fusion | No           | No                          | Yes                | No   | Berrow et al. [53] |
|                              | SLIC      | No           | No                          | Yes                | No   | Li et al. [52] |
|                              | Gibson isothermal assembly | No | No                          | Yes                | No   | Gibson et al. [54] |
|                              | CPEC      | Yes          | No                          | Yes                | No   | Horton et al. [55] |
| Site-specific recombination  | SSRTA (φBT1 integrase) | No | No                          | Yes                | No   | Zhang et al. [57] |
| Homing endonucleases         | iBrick    | No           | No                          | Yes                | No   | Colloms et al. [58] |
|                              | C-Brick   | No           | No                          | Yes                | No   | Liu et al. [59] |

CPEC, circular polymerase extension cloning; LCR, ligase cycling reaction; SIAR, serine integrase recombinational assembly; SLIC, sequence and ligase independent cloning; SSRTA, site-specific recombination-based tandem assembly.

### 3.2.1. Isoprene

The IspS enzymes have very high Kₘ values. Because of the low affinity of IspS for DAMPP, the protein fusions of IPP isomerase (IDI) and IspS were constructed based on the hypothesis that these fusions might improve isoprene production by providing a high local concentration of DAMPP to IspS. Overexpression of the IDI and P. alba IspS enzyme fusions, especially the IDI-IspS fusion, has increased isoprene production in a cyanobacterium [40]. In vitro assays using equimolar amounts of purified enzymes confirmed that the IDI-P. alba IspS fusion effectively increased the rate of IPP conversion to isoprene. Presumably, the DMAPP produced by IDI was channeled toward the active site of IspS in the fusion without being released into the surrounding environment, thereby increasing the rate of isoprene synthesis. This work demonstrates that the use of synthetic fusion proteins can successfully improve product yields in cyanobacteria.
226

227

was further assembled and incorporated into the highest steviol
228
ducing high-level of DMAPP/IPP in
229
same pathway is also present in many living organisms. For pro-
230
The precursor DMAPP/IPP is originated from MEP pathway. And the
231
possess bioactivities of anti-in
232

3.2.4. Compound K

233

Codon optimization; and protein fusion (IDT-ispS)

234

Exploiting exogenous MEP pathway genes dxs, ispD, ispF, and id from four
235
bacteria Erwinia taxa, Streptomyces, avermitilis, Saccharopolyspora erythraea, and
236
Bacillus subtilis to improve isoprenoid production.

237

Transporter engineering (Overexpression of native efflux pumps AcrAB-ToIC,
238
MdTEF-ToIC; Overexpression of exogenous pumps mexAB-OprM; and
239
Modulating the copy number of pump).

240

404 mg L⁻¹[68]

241

ent-Kaurene

242

E. coli

243

Transporter engineering.

244

32 mg L⁻¹[68]

245

Rebaudioside A

246

E. coli

Modules standardization; Codon optimization; N terminus truncation and
247
substitution; and KAH replacement.

248

10 mg L⁻¹[68]

249

Compound K

250

S. cerevisiae

EST database mining: Enzymatic UGT characterization; and Prompter
251
replacement.

252

1.4 mg L⁻¹[39]

253

Ginsenoside F1

254

S. cerevisiae

UGTPs characterization; Structural modeling; Protein chimeric.

255

42 mg L⁻¹[65]

256

Ginsenoside Rh1

257

S. cerevisiae

Same as ginsenosides F1.

258

92 mg L⁻¹[65]

259

Ginsenoside Rh2

260

S. cerevisiae

Yeast chromosomes integration; Cell factories establishment.

261

0.17 g L⁻¹[66]

262

Ginsenoside Rg3

263

S. cerevisiae

Same as ginsenosides Rh2.

264

0.5 g L⁻¹[66]

265

6-DeB

266

E. coli

Transporter engineering; and regulatory factor modulating.

267

59 mg L⁻¹[69]

268

Table 3

269

Heterologous biosynthesis of important natural products.

270

Target products

271

Hosts

272

Engineering strategies

273

Yield

274

References

275

Isoprene

276

Synechococcus elongatus

277

Codon optimization; and protein fusion (IDT-ispS)

278

1.2 g L⁻¹[40]

279

Gao et al.[40]

Amorphadiene

280

E. coli

Exploiting exogenous MEP pathway genes dxs, ispD, ispF, and id from four
281
bacteria Erwinia taxa, Streptomyces, avermitilis, Saccharopolyspora erythraea, and
282
Bacillus subtilis to improve isoprenoid production.

283

232 mg L⁻¹[61]

284

Wang et al.[61]

ent-Kaurene

285

E. coli

Transporter engineering (Overexpression of native efflux pumps AcrAB-ToIC,
286
MdTEF-ToIC; Overexpression of exogenous pumps mexAB-OprM; and
287
Modulating the copy number of pump).

288

404 mg L⁻¹[68]

289

Wang et al.[68]

ent-Kaurene

290

E. coli

Transporter engineering.

291

32 mg L⁻¹[68]

292

Wang et al.[68]

Rebaudioside A

293

E. coli

Modules standardization; Codon optimization; N terminus truncation and
294
substitution; and KAH replacement.

295

10 mg L⁻¹[68]

296

Wang et al.[68]

Compound K

297

S. cerevisiae

EST database mining: Enzymatic UGT characterization; and Prompter
298
replacement.

299

1.4 mg L⁻¹[39]

3.2.2. Amorphadiene

300

Amorphadiene is the important intermediate of artemisinin. The precursor DMAPP/IPP is originated from MEP pathway. And the same pathway is also present in many living organisms. For producing high-level of DMAPP/IPP in E. coli, Wang et al.[61] screened the elementary library to get the rate-limiting genes dxs, ispD, ispF, id from a wide range of microorganisms and assembled into the native MEP pathway. The results show that dxs2 derived from Streptomyces avermitilis, ispF from E. coli, ispD from S. erythraea, and id from Bacillus subtilis highly improve the metabolic flux of MEP pathway. Finally, this interchange strategy enhanced the yield of amorphadiene to 15.5 -fold in E. coli.

301

3.2.3. Rebaudioside A

302

Rebaudioside A (RA) is an intense natural diterpenoid sweeteners isolated from Stevia rebaudiana. The biosynthetic pathway of RA in Stevia plants has been largely characterized and it involved nine enzyme-catalytic reactions from isopentenyl diphosphate/dimethylallyl diphosphate [62,63]. The pathway then was divided into three metabolic modules: terpene synthetic module, cytochrome P450 module, and glycosylation module for de novo biosynthesis of RA in E. coli.[64] Each module has been tested and optimized before transferred into the host. The first module has been established in the precisely work and the maximum yield of ent-kaurene, the precursor of RA, was 194 mg/L (in shake flask) and 1.8 g/L (in a 5-L bioreactor). Of the second module building, two P450 enzymes ent-kaurene oxidase (KO) and KAHI are responsible for the selective C19 oxidation and C13 hydroxylation of ent-kaurene, respectively. The KO was incorporated along with an electron transfer partner CPR, and the KAHI was replaced by CYP714A2 from A. thaliana together with N terminus engineering resulted in 17str29CYP714A2 to produce 15.47 mg/L of steviol, which was enhanced to 25.6-fold compared with KAHI2. Afterwards, the UGT module UGT85C2/UGT91D2/UGT74G1/UGT76G1 was further assembled and incorporated into the highest steviol producer with 17str29CYP714A2 to achieve the complete biosynthesis of GA in E. coli and the yield reached 10.0 mg/L.

3.2.4. Compound K

303

Compound K (CK) is the major ginsenoside-type metabolite detected in vivo after oral administration of ginseng in mammals, which has not been detected from Panax plants. It is testified to possess bioactivities of anti-inflammation, hepatoprotection, anti-diabetes and anti-cancer. The identification of the novel UDP-glycosyltransferase, UGT7g1, which could selectively glycosylate the C-20(S)-OH of PPD is the key point to build a yeast cell factory to produce CK from simple sugars. After co-expression of UGT7g1 with the PPD biosynthetic pathway in the yeast chassis, CK could be produced from cheap monosaccharide via microbial fermentation [39]. A novel compound, 20S-O-β-(β-glucosyl)-dammarenediol II was discovered and identified in the CK producing yeast, which is readily converted to CK by a P. ginseng cytochrome P450 in vitro. Thus, two parallel pathways for CK production in the engineering yeast cell as well as in Panax plant are supposed. The ‘one-pot’ biosynthesis of CK in yeast provides not only a potential low-cost CK manufacturing method for its clinical applications, but also the scientific bases for understanding the biosynthetic pathways of ginsenosides in Panax plants.

3.2.5. Rare ginsenosides

304

Rh2, Rg3, Rh1 and F1 are all rare ginsenosides, and their contents in ginseng are extremely low. For example, Rh2 is almost undetectable in the total ginsenosides of Panax plants (P. ginseng, P. quinguefolium, and P. notoginseng), and the Rg3 is only detected in the total ginsenosides of P. ginseng; its content in the dry ginseng is about 0.0003%. To build yeast cell factories to produce rare ginsenosides, a yeast chassis to produce PPD or PPT with high yield was built by enhancing the MVA pathway of S. cerevisiae as well as integrating the genes related to PPD or PPT biosynthetic pathway (gene encoding PgDDS, CYP716A47, PgpCPR1and CYP716A53v2) into its chromosome [65,66]. And then the characterized UGT7g1s were introduced into the PPD or PPT producing chassis to obtain the yeast cell factory producing different rare ginsenosides.

3.2.6. Transporter engineering

306

The tolerance of chassis cells to heterogenous compounds are key point for high level production. Thus, expelling heterologous compounds out of hosts by transporters is a potential strategy to enhance product titers in microbial cell factories. Microorganisms have evolved membrane transporters that recognize and export toxic compounds from the cell and sustain their survival by exporting a wide range of substrates, antibiotics, chemotherapeutic agents, and solvents [67]. The general engineering strategies for the product transportation modules are: (1) over-expression of efflux pump genes; (2) combinatorial expression of efflux pump components; (3) introduction of exogenous transportation modules and design of artificial transportation system.

The pleiotropic resistant pumps, such as AcrAB-ToIC from E. coli and MexAB-OprM pump from Pseudomonas aeruginosa has been verified the function that expel a wide range of relevant antibiotics
from cells. For enhancing isoprenoid-amorphadiene (sesquiterpene) and kaurene (diterpene) production, Wang et al. [68] overexpressed, systematically assembled and modulated these efflux pumps in E. coli. The overexpression of AcrB and ToIC components can effectively enhance the specific yield of amorphadiene and kaurene, e.g., 31 and 37% improvement, respectively. The heterologous MexB component can enhance kaurene production with 70% improvement which is more effective than ToIC and AcrB. The results suggest that the three components of tripartite efflux pumps play varied effect to enhance isoprenoid production. Considering the highly organized structure of efflux pumps and importance of components interaction, various component combinations were constructed and the copy number of key components AcrB and ToIC was finely modulated as well. The results exhibit that the combination ToIC and ToIC and AcrB improved the specific yield of amorphadiene with 118%, and AcrA and ToIC and AcrB improved that of kaurene with 104%.

Another example is that Yang et al. [69] modulated tripartite multidrug efflux pumps MacAB-ToIC, AcrAB-ToIC, MdtEF-ToIC, and MexAB-OprM for increasing heterologous polyketide 6-deoxyerythronolide B production. Compared with the control, overexpression of a single component of efflux pumps (except oprM) repressed 6dEB production, but modulation of two components MacA and MacB, or the complete pumps MacAB-ToIC and MdtEF-ToIC significantly improved 6dEB titer by 100, 118, and 98%, respectively. In addition, to avoid the challenging fine-tuning components of pumps, the transcriptional regulators of efflux pumps were modulated to improve the 6dEB production. Overexpression of RpoH (activator of MdtEF-ToIC) and EnvA (activator of EmrKY-ToIC and AcrAD-ToIC) strongly increased 6dEB titer by 152 and 142%, respectively. These two cases strongly suggested that transporter engineering is a potentially effective strategy to enhance the yield of natural products in a heterologous biological system.

4. Computer-aided guidance for systematic optimization

The functional behavior of a module in a cell depends not only on its component devices and their connectivity, but also on the cellular context in which the module operates. Because synthetic modules and endogenous cellular processes condition each other’s behavior, any fluctuations in the host cell processes are relayed to the module and affect its output and vice versa. This presents a problem for engineering predictable, reliable biological systems. One approach to solving this problem is predict the process utilized computer-aided mathematical model and take into account a module’s connection to the host’s cellular context. Although simplification, specification, and standardization make engineering easier, it may not be advantageous to hide all the information about the host cell.

Using the knowledge-based empirical approaches, some targets of the MEP pathway (e.g., dxa, idi, and ispdF gene) were straightforwardly engineering for enhancing production of isoprenoids in E. coli. [70] A novel approach: flux distribution comparison analysis (FDCA) was developed for discovery of genomic scale metabolic targets in E. coli. 51 knockout, down-, and up-regulated targets were predicted and experimentally tested to enhance lycopene production. Five significant targets gdh A, eut D, tpi A, omp E, and ompN were combined the lycopene titer improved by 174% in shake-flask. For further better understand the imbalance of modules in heterologous expression system, we will give the following example for 6dEB production by antisense RNAs. 6-deoxyerythronolide B (6dEB) is a key intermediate of erythromycin. Heterologous biosynthesis of 6dEB has been successful in E. coli, but the conversion is still a very low molar yield range from 0.7 to 2.1%. On the contrary is that 6dEB derived from propionate could reach up to 11.2% in E. coli, which was first evaluated utilizing the maximum theoretical molar yield (MTMY) by Meng et al. [71] This extremely gap is often caused to a large extent by the imbalance of heterologous biosynthetic pathway and endogenous metabolic network.

To demonstrate the metabolic module interactions for understanding their imbalance, a genome-scale metabolic model (GSMM) was constructed based on iAF1260, and another two flux distribution analysis methods, FDCA and LMOMA were used to calculate the MTMYof 6dEB in E. coli [72]. Metabolic pathways of the cell that interact with 6dEB biosynthesis module were totally divided into 8 major functional modules, including (i) propionate metabolism and 6dEB biosynthesis module; (ii) TCA cycle module; (iii) glycolysis module; (iv) PP pathway module; (v) nucleotide metabolism module; (vi) pyruvate metabolism module; (vii) cell membrane constituent biosynthesismodule; and (viii) amino acid metabolism module. The mathematic model, which simulate the optimal growth of E. coli and 6dEB production characterize an “ideal situation” (maximize product formation rate).

In comparison with the “actual situation”-transcriptome analysis of 6dEB production strain by microarray assay, the significant discrepant genes of these two situation mainly involved in the pentose phosphate pathway module and nucleotide metabolism module. All 25 predicted targets at these two modules were tested for improving the 6dEB production in E. coli via synthetic antisense RNAs. Down-regulation of 18 target genes leads to more than 20% increase in 6dEB yield. Combinatorial repression of targets with greater than 60% increase in 6dEB titer, e.g., anti-guaB/anti-zwf led to a 296.2% increase in 6dEB production (210.4 mg/l in flask). This study strongly demonstrated that the synthetic 6dEB module not only interfaced with connecting pentose phosphate pathway module, but also with nucleotide metabolism module in cellular.

5. Perspective

Synthetic biology is an emerging filed with the development with DNA sequencing technology, DNA recombinant technology, computational biology, and biological engineering. It distinguishes itself from other disciplines in both its approach and its choice of object. Synthetic biology should be considered a hybrid discipline, combining both engineering and science to achieve its goal of engineering synthetic organisms. The aim of this discipline for producing a genetic product in a host cell may be not an issue in term of biological parts characterization and DNA assembly, but the modules interfacing in the future.

Biological systems are dynamic, fine regulated, nonlinear complex systems [73]. It possessed a far greater degree of integration of their parts than that of non-living systems. However, the behavior of most synthetic modules has been studied in isolated cells. The function of a synthetic module may fundamentally affect host cell processes, thus altering the cellular context. This situation is further complicated and reprogramming the cell is required, which involves the creation of synthetic biological components by adding, removing, or changing genes or proteins [74]. In the multiply rounds of design, construction, and testing of new cellular components, researchers need to develop feasible and precisely predicted tools and effective methods for module construction. As a result, design of synthetic biological systems has become an iterative process of screening, modeling, construction, and testing that continues until a system achieves the desired behavior. Rational redesign based on mathematical models improves system behavior in such situations. These retooled systems are once again tested experimentally and the process is repeated as needed. Many synthetic biological systems have been engineered successfully in this fashion because the methodology is highly tolerant to uncertainty.
Synthetic biology will benefit from further such development and the creation of new methods that manage uncertainty and complexity.

Acknowledgement

This work was supported by the National Basic Research Program of China (“973” Program, grant No. 2012CB721041), the National Natural Science Foundation of China (grant No. 31170104).

References

[1] Li JW, Vederas JC. Drug discovery and natural products: end of an era or an endless frontier? Science 2009;325:161–5.
[2] Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 1971;93:2325–7.
[3] Akera T, Brody TM. The role of Na+, K+–ATP in the inotropic action of digitoxin. Pharmacol Rev 1977;29:187–220.
[4] Sindrup SH, Brosen K, Bjerring P, Arendt-Nielsen L, Larsen U, Angelo HR, et al. CoT scores and antimutagenic activity of sparteine. Clin Pharmacol Ther 1990;48:686–93.
[5] Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, Hallinan J, et al. Standard bioconjugation and standardization of synthetic biological parts knowledgebase. PLoS One 2011;6:e17005.
[6] Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new tools for new ideas. Nature 2009;325:161–9.
[7] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for detection of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol 2007;8:R22.
[8] Pickens LB, Tang Y, Chooi YH. Metabolic engineering for the production of CoQ(10). Microb Biotechnol 2009;2:1753–60.
[9] Li R, Townsend CA. Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in Streptomyces coelicolor. Metab Eng 2011;13:733–44.
[10] Carlson JM, Chakravarty A, DeZiel CE, Gross RH. BEAM: a beam search algorithm for the identification of orthologous genomic sequences. Nucleic Acids Res 2003;31:3540–7.
[11] Winter JM, Behnken S, Hertweck C. Genomics-inspired discovery of natural products. Curr Opin Chem Biol 2011;15:107–15.
[12] Anderson JC, Duber JE, Leguia M, Wu GC, Goler JA, Arkin AP, et al. Bglbricks: a virtual biological parts knowledgebase. PLoS One 2011;6:e17005.
[13] Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 2007;4:251–5.
[14] Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. Nucleic Acids Res 2015;43:D129–34.
[15] Galdzicki M, Rodriguez C, Chandran D, Sauro HM, Gennari JH. Standard biological parts knowledgebase. PLoS One 2011;6:e17005.
[16] Ingmar JS, Deynoot Z, Plab H, Chen J, Hilson NJ, Keating JD. Design, implementation and practice of iBEI-ICE: an open source biological parts registry platform and tools. Nucleic Acids Res 2012;40:e141.
[17] Li SY, Zhao GP, Wang Y. Dynamic genome of China (W686). J. Li et al. / Synthetic and Systems Biotechnology 1 (2016) 221–229
[62] Brandle JE, Telmer PG. Steviol glycoside biosynthesis. Phytochemistry 2007;68:1855–63.
[63] Ceunen S, Geuns JM. Steviol glycosides: chemical diversity, metabolism, and function. J Nat Prod 2013;76:1201–28.
[64] Wang JF, Li SY, Xiong ZQ, Wang Y. Pathway mining-based integration of critical enzyme parts for de novo biosynthesis of steviolglycosides sweetener in Escherichia coli. Cell Res 2016;26:258–61.
[65] Wei W, Wang P, Wei Y, Liu Q, Yang C, Zhao G, et al. Characterization of Panax ginseng UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. Mol Plant 2015;8:1412–24.
[66] Wang P, Wei Y, Fan Y, Liu Q, Wei W, Yang C, et al. Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts. Metab Eng 2015;29:97–105.
[67] Lv HJ, Li JH, Wu YY, Garyail S, Wang Y. Transporter and its engineering for secondary metabolites. Appl Microbiol Biotechnol 2016;100:6119–30.
[68] Wang JF, Xiong ZQ, Li SY, Wang Y. Enhancing isoprenoid production through systematically assembling and modulating efflux pumps in Escherichia coli. Appl Microbiol Biotechnol 2013;97:8057–67.
[69] Yang JY, Xiong ZQ, Song SJ, Wang JF, Wang Y. Improving heterologous polyketide production in Escherichia coli by transporter engineering. Appl Microbiol Biotechnol 2015;99:8691–700.
[70] Wang JF, Meng HL, Xiong ZQ, Zhang SL, Wang Y. Identification of novel knockout and up-regulated targets for improving isoprenoid production in E. coli. Biotechnol Lett 2014;36:1021–7.
[71] Meng HL, Lu ZG, Wang Y, Wang XN, Zhang SL. In silico improvement of heterologous biosynthesis of erythromycin precursor 6-deoxyerythronolide B in Escherichia coli. Biotechnol Bioproc E 2011;16:445–56.
[72] Meng HL, Xiong ZQ, Song SJ, Wang JF, Wang Y. Construction of polyketide overproducing Escherichia coli strains via synthetic anti-sense RNAs based on in silico fluxome analysis and comparative transcriptome analysis. Biotechnol J 2016;11:530–41.
[73] Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 2009;10:410–22.
[74] Liu Y, Shin H, Li J, Liu L. Toward metabolic engineering in the context of system biology and synthetic biology: advances and prospects. Appl Microbiol Biotechnol 2015;99:1109–18.