16

Applying Advanced DNA Assembly Methods to Generate Pathway Libraries

Dawn T. Eriksen¹, Ran Chao¹, and Huimin Zhao¹,²

¹ University of Illinois at Urbana-Champaign, Department of Chemical and Biomolecular Engineering, 600 South Mathews Avenue, Urbana, IL 61801, USA
² University of Illinois at Urbana-Champaign, Departments of Chemistry, Biochemistry, and Bioengineering, 600 South Mathews Avenue, Urbana, IL 61801, USA

16.1 Introduction

Pathways, which are cascades of biochemical reactions catalyzed by enzymes, maintain the vitality of all living organisms. These biochemical routes have been exploited to produce numerous commodities since early civilization, such as beer, wine, and cheese. With the advance of biotechnology, various genetic tools have become available for construction and manipulation of pathways to efficiently convert renewable feedstock to value-added compounds such as specialty chemicals, pharmaceuticals, and biofuels [1]. Microbial production of these compounds is usually enabled by overexpressing endogenous or heterologous enzymes of the corresponding pathways. However, overexpression of pathway enzymes alone can be insufficient for optimal metabolite production due to an imbalanced flux through the pathway [1, 2]. A typical symptom of flux imbalance is the accumulation of unwanted and even toxic intermediates [3, 4], which can be detrimental to the productivity of desired compounds. There is seldom a straightforward strategy to resolve the non-product accumulation because enzymes within the pathway are not independent; instead the enzymes are intertwined and cross-regulated among the pathway enzymes and among the cell’s intricate metabolic networks. Due to this complexity, rationally engineering a pathway to improve its efficiency is a significant challenge. To this end, random approaches can be preferred over rational design in pathway engineering [5]. Random engineering approaches to optimize pathways generally screen through large and/or combinatorial pathway libraries. Pathway libraries have been constructed for diverse gene expression based on promoters of different strengths [6], varied intergenic regions affecting mRNA stability [4], or engineered ribosomal binding sites (RBSs) of diversified translational initiation rates [7].

In previous studies [4, 6–8], the pathway libraries were assembled by restriction digestion/ligation or overlap extension polymerase chain reaction (PCR).
These traditional assembly methods were limited in complexity of design, being forced to rely on the multiple-cloning site (MCS) for pathway assembly, and had low assembly efficiency. In recent years, a number of new DNA assembly methods have been developed, such as DNA assembler [9], sequence and ligation-independent cloning (SLIC) [10], Gibson assembly [11], circular polymerase extension cloning (CPEC) [12], Golden Gate cloning [13], and BioBrick standards [14]. These advanced DNA assembly methods have ameliorated the design constraints on heterologous pathway construction and simplified the assembly of multi-gene metabolic pathways. The improved efficiency of these methods allows for larger and unbiased library creation, while the modularity of the methods greatly facilitates the generation of complex combinatorial libraries (Figure 16.1). The following chapter will include a brief description of the advanced assembly methods that could be applied to combinatorial pathway libraries. Some of the most recent work in pathway library generation using these methods will then be discussed as well.

![Figure 16.1](image_url)  
**Figure 16.1** Overview of the combinatorial library approach for pathway improvement. When improving a multi-gene pathway, variations of the pathway components including promoters, RBSs, coding DNA sequences (CDSs), or transgenic regions are generated by either mutagenesis, homolog cloning, or *in silico* design (promoters and CDSs are used as examples in the figure). The diversified components are then assembled by various DNA assembly techniques to form a library of combinations. Cells hosting this pathway library will then be screened for the optima of the desired phenotype. Labels “p1-3” standard for promoters. Labels “t1-3” standard for terminators.
16.2 Advanced DNA Assembly Methods

The following methods have the potential to be used for pathway library creation (Table 16.1). The advanced assembly methods exploit diverse strategies for pathway construction such as homologous recombination, DNA polymerase extension, and advanced applications of restriction digestion/ligation.

The following assembly strategies are based on homologous recombination and DNA repair mechanisms: DNA assembler, Gibson assembly, and SLIC. In the DNA assembler strategy, the endogenous in vivo homologous recombination mechanism in yeast is used to create large pathways in a simple, one-step manner [9, 15, 16]. The DNA fragments to be assembled are PCR amplified by oligos designed with an 80-bp homologous region to the 5′ and 3′ neighboring DNA sequences within the pathway. The linear DNA fragments are cotransformed with the linear plasmid backbone into Saccharomyces cerevisiae, and the homologous regions are recognized by the endogenous homologous recombination machinery and “repaired” into a single DNA molecule.

Mimicking the in vivo homologous recombination mechanisms, in vitro assembly has been accomplished by SLIC and Gibson assembly. SLIC is a two-step DNA assembly method [10], which utilizes a 30-bp homology region. The linearized host vector and the insert DNA fragment are separately treated with T4 DNA polymerase in the absence of deoxynucleotide triphosphates (dNTPs), which chews back the 3′ terminal end. This generates a 5′ overhang that is homologous to the vector/insert. The second step involves addition of RecA and adenosine triphosphate (ATP), which can recombine the DNA fragments together into a single plasmid: any nicks generated are fixed after transformation. The Gibson assembly method [11] exploits a specific exonuclease to chew back the 5′ end to generate single-stranded complementary overhangs and ligases that are incorporated in the reaction mix to seal the DNA nicks. The DNA fragments are PCR amplified with 15–30 bp of homologous DNA regions to the 5′ and 3′ adjacent DNA sequences. In a single reaction, both vector and insert are subjected to T5 exonuclease that chews back the 5′ ends of the DNA fragments, and then the polymerase and ligase combine the homologous ends of fragments to a single circular DNA molecule.

Table 16.1 Summary of different advanced DNA methods that could be used for combinatorial library generation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Type of reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLIC</td>
<td>Exonuclease-based overhang generation and in vivo ligation</td>
</tr>
<tr>
<td>BioBrick standards</td>
<td>Step-wise modular restriction digestion and in vitro ligation</td>
</tr>
<tr>
<td>Golden Gate</td>
<td>Type II restriction enzyme digestion and in vitro ligation</td>
</tr>
<tr>
<td>DNA assembler</td>
<td>In vivo homologous recombination</td>
</tr>
<tr>
<td>Gibson assembly</td>
<td>Exonuclease-based overhang generation and in vitro ligation</td>
</tr>
<tr>
<td>CPEC</td>
<td>Overlap extension PCR</td>
</tr>
</tbody>
</table>
Homologous recombination is successful in DNA assembly, but basic polymerase extension mechanisms have also shown to be successful in the CPEC method to assemble DNA fragments into a plasmid [12]. The insert and vector are fused in an overlap extension PCR and circularize with extended overlapping stands, leaving only a nick in each strand. Then *Escherichia coli* repairs the nicks *in vivo* when transformed.

Another family of advanced DNA assembly techniques has been developed via the implementation of the type IIS endonucleases such as *Bsa*I, which cleave the DNA outside of their recognition sites, resulting in 5′ or 3′ DNA overhangs of nearly any user-defined nucleotide sequence [13, 17]. This strategy is more advanced than traditional restriction digestion/ligation method because it allows more flexibility in insertion location than cloning into the MCS on a plasmid. Use of type IIS endonucleases through the Golden Gate assembly method is a one-step reaction, which combines restriction digestion and ligation. This method has a high fragment assembly efficiency and proven to be effective in creating gene libraries [17]. A continuing area of research with this technique is investigating a more modular approach for pathway and pathway library construction [18, 19].

The need for modularity in gene and pathway cloning is becoming more significant with recent focuses on high-throughput DNA assembly and automation. One of the most established strategies for assembly standardization is the BioBrick system [14, 20–24]. The BioBrick and BglBrick standards (such as vectors, promoters, and RBS) rely on isocaudomer pairs of restriction enzymes to generate compatible cohesive ends and, upon ligation, result in a scar sequence that cannot be cleaved by either of the original restriction digests. DNA fragments flanked with these recognition sequences can be used for modular assembly of a pathway by iterative digestions and ligations.

Consideration of which assembly strategy to use for the generation of pathway libraries will greatly depend on the chassis, number of DNA fragments, and required assembly efficiency. *In vivo* homologous recombination is especially useful if the pathway is being expressed in *S. cerevisiae*. However, Gibson assembly and BioBrick standards are very useful if working in *E. coli*. Many DNA fragments to be assembled in the library can greatly decrease the assembly efficiency, which should be considered if a complex pathway is being investigated. If assembly efficiency is limiting, a strategy that allows for longer homology or linker region can be applied. Though no studies have linked library size to assembly strategies, some of the previous strategies might limit the library size, which could reduce the potential search space. Biases in assembly toward a certain gene or promoter can also reduce the potential search space. It is important to ensure that the library is diverse and random clones exhibit all potential genotypes of the library. One-pot assembly is also an important consideration, as iterative assemblies can be time consuming and can also reduce the potential library size.

### 16.3 Generation of Pathway Libraries

Combinatorial pathway library screening strategies, as compared with traditional pathway engineering strategies, can be more efficient in the identification of an optimized pathway. Traditional strategies optimize individual components
of the pathway one at a time to increase flux to the desired product [25–27], but pathway library screening strategies can tune multiple components of the pathway simultaneously. By varying multiple constituents concurrently, the likelihood of obtaining an optimized flux via balanced gene expression and protein activity within the pathway is increased. A more comprehensive exploration of the potential diversity of a target pathway can be achieved, which could identify unexpected synergistic effects [28, 29]. Many pathway optimization strategies are based on gene expression by varying promoter strength or RBS engineering. It is also possible to balance the flux through the pathway by exploring various combinations of enzymatic properties such as catalytic efficiency, cofactor specificity, stability, and substrate specificity. Currently, there are several examples of pathway libraries constructed through different advanced DNA assembly methods.

16.3.1 *In vitro Assembly Methods*

The Gibson assembly method was applied to generate a large combinatorial library of promoters and enzymes. The proof-of-concept pathway was the heterologous acetate utilization pathway in *E. coli*, comprised of an acetate kinase (*ackA*) and a phosphotransacetylase (pta) [30]. This combinatorial library was based on three promoter sequences with assorted strengths and four orthologous variants of both genes, generating 144 possible unique combinations of the promoters and genes. Each gene cassette was synthesized with an RBS, a terminator, and the promoter/gene variant. A unique 40-bp DNA linker sequence contains homologous DNA directly upstream and downstream of the gene at the terminal ends of the cassette (Figure 16.2). This linker region was used to ensure proper pathway sequence during assembly.

The total library size was approximately $10^4$, affording 70-fold coverage of the 144 possible combinations. Investigation of the assembly efficiency showed that over 80% (30/37) of the selected clones harbored a correctly assembled pathway. Further sequencing analyses showed that of the thirty correctly assembled pathways, 60% (18/30) had recognizable promoter sequences. Of the possible 144 promoter/gene combinations, 14 unique combinations were present in the 18 positively identified pathways. A bias was noted toward a specific combination of genes from certain organisms, even though each gene fragment was assembled in equal combinations. This bias could have been the result of an assembly bias, or it could be the result of a screening bias, as the library was screened on acetate and these genes could be the most efficient for acetate utilization in *E. coli*.

The Gibson assembly was also used by Coussement and coworkers in another example of creating a combinatorial library of transcription, translation, and protein sequence variability [31]. This strategy utilized a single-stranded assembly to introduce diversity in the double-stranded DNA of the promoter, RBS, and/or coding sequences. Optimization of the assembly found that two oligonucleotide fragments of similar lengths provided a nearly 100% efficiency of assembly. More DNA fragments or fragments of different lengths lowered the assembly efficiency. Promoter, RBS, and protein libraries using a single gene were all proven to have a large linear range and had diverse expression and activity. The assembly was tested for combinatorial pathway libraries using the
Figure 16.2 Preparation of DNA fragments for large library generation. Each unique design represents a unique promoter or gene. Varied strength promoters, orthologous genes, or mutated pathway components generate diversity. If the DNA is assembled with homology regions, upstream and downstream of the DNA fragment of interest, the pathway can assemble properly into many different combinations. Each strategy has incorporated different lengths of homology, which can contribute to the efficiency of correct assembly. These DNA fragments are then subjected to the desired DNA assembly reaction with the linearized vector and transformed into the host.
Generation of Pathway Libraries

reporter genes mKate2 and sfGfp. The promoter library was tested using this fluorescent pathway. One hundred and eighty-eight clones were randomly picked and profiled for complete representation of the potential expression landscape. Theoretical library size was $4^{32}$ and these 188 clones revealed a good representation of diversity. This pathway optimization is based on short fragment assembly of 50–150-bp assembly and has not been applied to larger DNA fragments. This strategy is efficient for shorter promoter regions of *E. coli* and point mutations of targeted protein engineering, which is one of the preferred strategies for protein engineering. However, pathways incorporating diversity in larger DNA fragments such as yeast promoters and other protein engineering strategies could not be accomplished through this current method.

The Gibson assembly was also utilized by Lee *et al.* to optimize a multienzyme pathway in the absence of a high-throughput assay [32]. This study took the pathway libraries assembly one step further and incorporated computational modeling to reduce the large search library that must be screened. For assembly, standardized vectors were constructed based on principles of the BglBrick-style cloning of protein fusions. The expression cassettes were flanked by pairs of homology sequences (20 bp) derived from yeast barcodes to allow for correct sequence assembly. Each promoter used was proven to work independently of DNA sequence directly downstream of it. Three-gene library assemblies resulted in 25–33% miss-assembly. Library assembly was tested in a three-gene fluorescent protein library, with a theoretical library size of 125. The triple library was shown to cover the complete three-dimensional expression space.

To apply this assembly to a pathway and construct a predictive model, the five-gene violacein biosynthetic pathway was utilized, resulting in a theoretical combinatorial library size of 3125. Ninety-one random transformants from the colony were characterized for geno- and phenotypic data. A linear regression model was then constructed from this data and used to predict optimal phenotypes. The authors suggest that a low sampling rate of 1–2% of the library could be sufficient for generating a predictive model. Four models were constructed for different intermediates and branched products of the violacein pathway. The model predictions and empirical data were high, with Pearson correlation coefficients being between 0.77 and 0.92 for the specific targets. The model was used to predict the top five expression-level combinations. These combinations were individually cloned and tested to determine if the desired product had increased production with the predicted expression levels. The model was able to predict and identify the expression level to yield the desired product with the highest production from the pathway.

A BioBrick-like assembly strategy was used in a combinatorial library of engineered RBSs [33]. This iterative assembly process utilizes the chloramphenicol resistance cassette paired with the library of RBS sequences. The resistance cassette is flanked by restriction digests and then can easily be removed to incorporate the next target gene and RBS library. To determine if the strategy could yield a library that spanned a multidimensional expression space, three reporter genes were used in a synthetic operon: CFP, YP, and mCherry. The RBS modulation was shown to span 100-fold in each dimension of the expression space. The seven-gene carotenoid biosynthesis pathway with the end product of astaxanthin
was used as a proof-of-concept study for this assembly in pathways. The theoretical library was 6^7 possible RBS combinations, and nearly 25,000 clones were visually screened, which is only 10% of the potential library. Through visual screening of the colonies' color of astaxanthin, 500 colonies were picked for further analysis. Fifty clones were identified to have the most intense color and screened for highest astaxanthin production through high-performance liquid chromatography (HPLC). This strategy yielded a clone with fourfold higher astaxanthin production than the wild-type pathway.

The aforementioned studies have all involved random, large pathway libraries. A new BioBrick standard platform, the ePathBrick system, allows for assembly of specific pathways, with the ability to vary specific components [34]. The ePathBrick system is a pathway fine-tuning toolkit that consists of a number of BioBrick-compatible plasmids with characterized regulatory signal elements. With this system, Xu and coworkers demonstrated a modular engineering approach for significant titer improvement of a multi-gene fatty acid metabolic pathway by fine-tuning gene expression through plasmid copy number and RBS engineering [35]. The *E. coli* fatty acid biosynthetic pathway was apportioned and overexpressed in three separate modules. These modules were successfully expressed on compatible ePathBrick vectors with varying plasmid copy numbers. The total fatty acid production was optimized by overexpressing each module on high, medium, or low copy number plasmids. Nine independent pathways were constructed through the ePathBrick standards and analyzed for fatty acid production. As has been noted before in product titer, the highest gene expression is not always optimal [6]. The greatest increase in fatty acid production occurred only when the final module was expressed highly, combined with a lower expression in the other modules. The balanced gene expression pathway produced a fourfold increase in fatty acid titer compared to the lowest-producing pathway. Similarly, three different strength RBSs were also tested in the modules, and a balance between strong and medium strength RBSs improved fatty acid production by twofold. This type of strategy can illuminate bottlenecks in the pathway. This study exemplified the importance of high concentrations of malonyl-CoA in fatty acid production.

A randomized BioBrick strategy has also been developed, which combines the power of Gibson assembly and the modularity of the BioBrick standards [36]. In this method, all promoters, RBSs, and transcriptional terminators were randomized within the pathway. These modular DNA fragments were derived from PCR-amplified BioBricks, and each component was cloned with 18–28-bp linkers of homologous DNA regions to the 5’ and 3’ DNA. Three promoters, three RBSs, and three terminators were simultaneously randomized for the three-gene pathway for the lycopene biosynthetic pathway, generating a library of nearly 20,000 unique clones. The library was assembled through Gibson assembly and was screened on plates for the orange-colored lycopene product. Of the red-orange colored colonies, 12 were selected, and DNA sequencing analysis demonstrated that 7/8 randomized pathways were distinct and four pathways had deletions. The study cautioned the metabolic burden placed on the cells during the library screening that could have caused the mutations.
16.3.2 In vivo Assembly Methods

E. coli does not have robust and efficient homologous recombination machinery; therefore in vitro assembly methods are highly needed. In contrast, plants and yeast have very vigorous and efficient homologous recombination machinery, allowing for facile pathway library creation in vivo. Two divergent strategies for in vivo homologous recombination have been developed: chromosomal integration and plasmid assembly.

16.3.2.1 In vivo Chromosomal Integration

Wingler and Cornish established a reiterative recombination method for the in vivo assembly of multi-gene pathway libraries directly into the chromosome [37]. The strategy utilized a pair of alternating orthogonal endonucleases and selectable markers. Homologous recombination and gap repair were used to construct a plasmid containing the gene of interest, marker, and endonuclease, which were recombined into an acceptor strain. This acceptor strain carries a predefined target locus for integration into the chromosome. Galactose-induced expression of the endonuclease cleaves the double-stranded DNA, triggering the homologous recombination and leading to integration of the gene of interest and the auxotrophic marker into the chromosome. The strains are then selected for the new auxotrophic marker and cured against excess donor plasmid. The proof of concept for pathway integration and mock library assembly was demonstrated using the lycopene biosynthetic pathway (crtE, crtB, and crtI). A large library of over $10^4$ was assembled: the mock library contained various ratios of crtB and crtI alleles that contained either nonsense or silent mutations, which would produce working or interrupted pathways. The diversity could be judged based on the actual and theoretical percentages of working pathways versus interrupted pathways, visualized based on the color of the colonies on the plate. Each library had the expected percentage of working pathways, indicating a non-biased library assembly into the chromosome.

Pathway library strategies have also been established in plant biotechnology to study secondary metabolites [38, 39]. Engineering secondary metabolism in plants can be a daunting task considering the complexity of the target pathways, which could have multiple branches, multifunctional and/or compartmentalized enzymes, and complex feedback inhibition. Zhu et al. established a novel method for the combinatorial nuclear transformation of multiple genes into a plant, generating a pathway library to simplify the study of multiple variables of secondary metabolites [38, 39]. Carotenoid production in cereal grains was used as a proof of concept. Embryos of the cereal-grain white maize were bombarded with metal particles coated with six unique constructs, consisting of a selection marker and five carotenogenic genes. The resultant library consisted of any combination of one or more expression phenotype from any of the five genes. This method of multiple gene transformation and pathway library screening allowed the identification of rate-limiting steps in the carotenogenic pathway. Total carotenoid production in cereal grains was improved 140-fold based on a unique combination identified from this multi-gene pathway library strategy.
16.3.2.2 In vivo Plasmid Assembly and One-Step Optimization Libraries

Chromosomal integration has been successful in pathway library creation, but assembling the pathways into a plasmid is also advantageous. A plasmid is a DNA molecule that can be easily transported across strains, which is an important characteristic to consider when excluding the possibility that the observed improvements are not a result of off-target genome modification.

An example of plasmid-based pathway libraries was constructed by the DNA assembly method and focused on a combinatorial library of different promoter strengths for all the genes within the library [40]. As a proof of concept in pathway library generation, the xylose and cellobiose utilization pathways for ethanol production were optimized. Efficient utilization of these biomass sugars is critical for economically feasible biofuel production. Promoters PDC1, ENO2, and TEF1 were mutagenized through nucleotide analog-based error-prone PCR to induce a very high mutation rate and produce promoters of various strengths. After mutagenesis, mutants for each promoter were assayed through fluorescence protein expression, and 10 promoters of defined strengths were selected for library construction. These 10 promoters in each position of the library resulted in a theoretical library size of $10^2$ and $10^3$ for the cellobiose and xylose utilization libraries, respectively. Each mutant promoter was cloned into a helper plasmid that contained 400-bp sequences homologous to the 5′ DNA region (Figure 16.2). The mutant promoter/gene expression cassettes were co-transformed into a yeast strain with a total library size of $10^5$. To confirm the diversity of the library, over 40 individual colonies from each library were screened from an antibiotic selection marker for plasmid-pathway assembly and not based on sugar utilization. Each colony from this plasmid marker selection exhibited a unique growth curve on its respective carbon source, which was indicative of a diverse library.

Improved sugar utilization was visualized in a high-throughput manner by inspection of colony size on agar plates, wherein larger colony sizes were suggestive of faster sugar utilization and improved growth. In the xylose utilization pathway, a very efficient mutant pathway was identified in a single step. This pathway conferred a xylose consumption rate of $0.73 \text{ g l}^{-1} \text{ h}^{-1}$, comparable with some of the fastest xylose consumption rates from strains that had been subjected to multiple generations of optimization strategies. The strain harboring the wild-type pathway did not produce any ethanol, while the mutant pathway conferred an ethanol productivity of $0.17 \text{ g l}^{-1} \text{ h}^{-1}$. In the cellobiose utilization strategy, the strain harboring the optimized pathway yielded a 5.4-fold improved cellobiose utilization rate and a 5.3-fold increase in ethanol productivity.

A similar pathway library strategy created a combinatorial library of homologous enzymes of the xylose utilization pathway, with fix-strength promoters [41]. The fungal xylose utilization pathway has been shown to be especially sensitive to cofactor imbalances and unbalanced enzyme expression [42–45]. A total theoretical library size of 8360 possible unique combinations of homologous enzymes for each of the five genes in the pathway was constructed through homologous recombination. Each enzyme was characterized to show varied activities and cofactor dependencies. The gene sequences were cloned into helper plasmid expression cassettes, containing promoters, terminators, and at
least a 400-bp region homologous to the 5′ and 3′ DNA regions of the pathway at the termini of the expression cassette (Figure 16.2). The expression cassettes were transformed into the three different yeast strains with an average library size of $1.3 \times 10^4$. To confirm library diversity and screening for optimal pathways, the same strategies established in the promoter-based library were applied [40]. Sequencing results of random colonies showed that all the genes were recognizable with no major mutations or hybrids, resulting in a 100% efficiency, and there was no significant bias toward a certain gene. The same library was screened in three different strains, and a unique combination of genes was discovered to be optimal in each individual strain. This unique combination for each strain is attributed to the different metabolic background of the strains and availability of precursors or cofactors.

16.3.2.3 In vivo Plasmid Assembly and Iterative Multi-step Optimization Libraries

Directed evolution, an iterative multistep optimization strategy, is an established strategy that is a very powerful technique in synthetic biology for optimizing protein activity [5, 46]. Application of the strategy has been expanded to include pathway-scale transcriptional engineering and protein engineering through the following pathway library studies. The directed evolution strategy on the pathway scale is particularly powerful because it allows for the optimal flux to be identified with no a priori information about pathway bottlenecks or specifics about the pathway enzymes. This directed evolution strategy on the pathway scale allows for all components to be screened/selected for a balanced activity, not just for high activity.

Yuan and coworkers applied directed evolution to mutant promoter pathway libraries of the cellobiose utilization [47]. An average mutation rate of 12–16-nucleotide substitutions per kilobase for each mutagenized promoter was obtained. The pathway genes were not mutagenized, and these non-mutated DNA fragments were co-transformed with the error-prone promoter library and a linearized vector for a total library size of $10^4$. The pathway phenotype improvement was assessed by fast sugar utilization, visualized by large colonies on agar plates. The first round of directed evolution identified a strain with a 5.7-fold increase in cellobiose consumption rate and a 5.5-fold increase in ethanol productivity. The further rounds of evolution yielded incremental subsequent increases (Figure 16.3). After characterizing the mutant promoters, it was found that the expression level ratios had significantly changed. While the parent BGL:CDT ($\beta$-glucosidase/cellodextrin transporter) relative expression ratio was $13.8:1$, the first round of mutagenesis altered the ratio to $2.5:1$. This significant increase in relative CDT expression suggested that this protein expression was a bottleneck.

Pathway-scale protein engineering strategies were also applied using homologous recombination [48]. In this study, both the BGL and CDT proteins were coevolved for balanced activity in a directed evolution manner. One amino acid substitution per protein was introduced through error-prone PCR, yielding a theoretical total library size of $9.9 \times 10^6$. No gene expression elements were mutagenized in this strategy and therefore were PCR amplified into the pathway
Figure 16.3 Fermentation profiles of the evolutionary rounds for the pathway libraries. (a,b) Cellobiose consumption and ethanol production of the cellobiose utilization pathway from the promoter-based directed evolution. The black square represents the parent pathway with no mutations in the PDC1 and ENO2 promoter. The circles are the first round of error-prone PCR of both promoters. The triangles represent the second and final rounds of directed evolution mutagenesis. (c,d) Cellobiose consumption and ethanol production of the cellobiose utilization pathway from the protein-based directed evolution. The black square represents the wild-type pathway with no mutations in the β-glucosidase and the cellodextrin transporter. The circle is the first round of error-prone PCR of both proteins. The triangle represents the second round of directed evolution.
16.4 Conclusions and Prospects

Advanced DNA assembly methods have allowed scientists and engineers extraordinary freedom in constructing pathways, greatly facilitating advances in pathway library generation. Pathway optimization through whole pathway libraries has expanded the potential diversity and possibilities for improving pathway phenotype. Furthermore, high efficiency and modularity of these advanced DNA assembly methods make in silico design [49] and automated assembly [50] of these libraries possible. Large combinations of library components can be individually constructed by robotic platforms and investigated by high-throughput screening for extensive investigations of improved pathway phenotypes. Despite the rapid progress of DNA assembly technologies, widespread application of pathway libraries is currently limited by high-throughput screening. Without the ability to easily and economically quantify the phenotype of interest, these large-scale pathway libraries will not be able to fulfill their maximum potential. Future high-throughput screening methods could be realized through microfluidic devices, with the ability to screen up to $10^8$ clones per day [51, 52]. Biosensors also have potential in high-throughput screening, as shown by a number of transcription factor-based biosensors that have been engineered to detect small molecules. These biosensors can link the small molecule concentration to an easily measurable signal such as fluorescence and cell growth via gene circuits [53–56]. Though there are challenges, the potential of using advanced DNA assembly methods to create pathway libraries to significantly improve microbial cell production of fuels and chemicals is significant, and future pathway engineering methods will benefit from these strategies.

Definitions

Pathway Coordinated heterologous and/or endogenous enzymatic reactions
Pathway engineering A research area that specializes in modifying or optimizing components of an enzymatic pathway for improved phenotype
Pathway optimization Strategies to improve the overall performance of an enzymatic pathway
Pathway libraries  A collection of mutant enzymatic pathways wherein multiple components (RBS, promoters, enzymes) within the pathway have simultaneously been mutated

Directed evolution  An evolutionary process for engineering biological systems that mimics Darwinian evolution in vitro and in vivo: rounds of random mutations are incorporated into the DNA sequence and selected for improved phenotype in an iterative fashion

DNA assembly  The process to conjoin several DNA fragments to create a large DNA molecule

References


