Combinatorial pathway optimization for streamlined metabolic engineering

Funding acknowledgement:
289326 - Standarization and orthogonalization of the gene expression flow for robust engineering of NTN (new-to-nature) biological properties. (EC)

Combinatorial Pathway Optimization for Streamlined Metabolic Engineering

Authors: Markus Jeschek[1]*, Daniel Gerngross[1] and Sven Panke[1]*

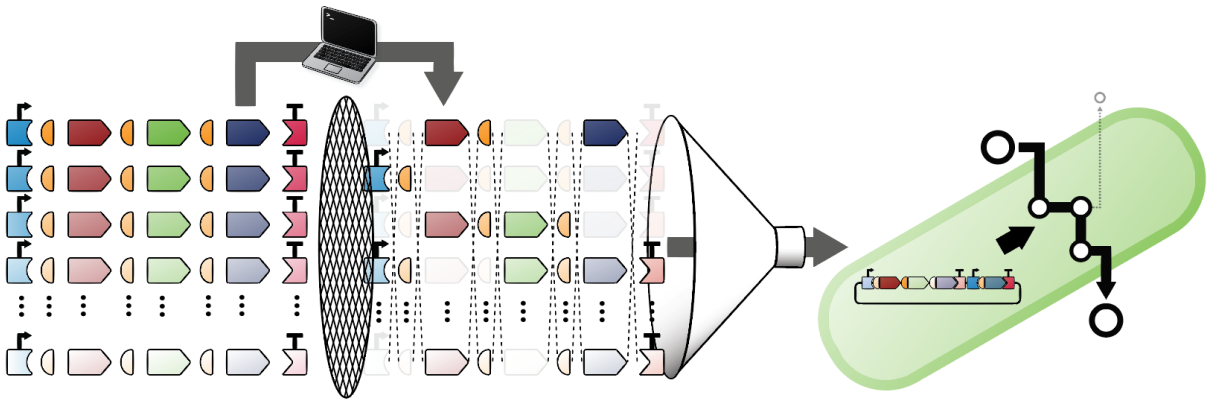
Affiliations: [1] Department of Biosystems Science and Engineering, ETH Zurich
Mattenstrasse 26, CH-4058 Basel, Switzerland

Contact information:
Phone: +41 61 387 32 09
* Correspondence to: sven.panke@bsse.ethz.ch ; markus.jeschek@bsse.ethz.ch

Abstract

Elimination of metabolic flux imbalances in microbial cell factories is an important part in the establishment of viable biotechnological production processes. However, due to the high complexity of cellular metabolism, the limited *a priori* knowledge about the majority of production pathways and a lack of forward design standards, metabolic engineers strongly rely on empirical screening methodologies to achieve the required improvement of cell behavior. Combinatorial pathway engineering provides an interesting tool to identify global solutions for intricate pathways, but methods for the reduction of combinatorial library size are inevitably required to restrict the experimental effort to an affordable size. Here we review recent advances from this field by scrutinizing commonly applied diversification methods and highlighting crucial strategies for the minimization of experimental effort.

Graphical abstract



Highlights

- Combinatorial pathway optimization for microbial production is highlighted
- Frequently applied methods for diversification are critically scrutinized
- Important heuristics for the reduction of experimental effort are introduced
- Recent studies applying combinatorial pathway optimization are summarized

Funding sources

This work was supported by the NCCR Molecular Systems Engineering of the Swiss National Science Foundation and the European Commission [project ST-FLOW (289326)].

Introduction

Metabolic engineering relies on the capability to precisely manipulate cellular metabolism in a targeted manner to achieve a desired system behavior, which is usually associated with the fabrication of economically relevant products such as fuels, commodity chemicals or pharmaceuticals [1,2]. The term “engineering” implies the inherent goal of this field, namely the design of biological systems according to predefined specifications at ever increasing precision [1-3]. Installation of efficient pathways strictly based on forward design rules, however, remains infeasible, mainly due to a significant lack of *a priori* knowledge about the target pathway and the intricate orchestration of cellular metabolism with its thousands of components and innumerable interactions [4,5]. As a consequence pathway flux, in particular within newly introduced heterologous or artificial metabolic routes, is inherently imbalanced, which can lead to a variety of obstacles including accumulation of (toxic) intermediates, side product formation, metabolic burden and ultimately growth inhibition and/or low product yield, to name but a few [4-6,7•].

Addressing the aforementioned imbalances to enable and improve target pathway flux and eventually create microbial cell factories for industrial application represents a major challenge for metabolic engineers [5]. Classically this has been done by identifying major bottlenecks in the initial pathway design and subsequently removing these individually by sequential optimization campaigns, thereby gradually optimizing pathway performance (“de-bugging”, “de-bottlenecking”) [7•,8]. However, these efforts require substantial *a priori* knowledge about pathway topology as well as suitable techniques to quantify relevant intermediates [7•,8,9••]. Moreover, due to the sequential optimization procedure holistic interactions within the immediate pathway as well as with overall host metabolism are neglected and globally optimal solutions are unlikely to be identified in this manner [9••].

More recently, novel approaches for pathway optimization have become feasible due to significantly reduced costs for commercial DNA synthesis [10], reliable *in vitro* and *in vivo* techniques for DNA assembly [11,12] as well as technologies for efficient and precise genome manipulation in a multiplexed manner [13-15]. These proceedings allow for the creation of variant libraries in which several pathway elements are diversified simultaneously (Figure 1) and thus facilitate combinatorial optimization of pathways (please note that different interpretations of “combinatorial” pathway optimization can be found in literature; for the definition used in this review please refer to Glossary Box 1). The latter can be performed with little prior knowledge about the pathway allowing for pragmatic and goal-oriented optimization schemes, and could in fact be capitalized upon for pathway characterization and establishment of pathway models [5].

<Glossary Box 1>

Definition of "combinatorial" pathway optimization

For the purpose of this review we define the term "*combinatorial*" as a scenario in which more than one variable or component of a system is varied concomitantly (i.e. multivariate optimization). In the context of pathway engineering this boils down diversifying two or more genetically encoded entities that directly or indirectly influence the pathway of choice. Varied parts can comprise regulatory elements (e.g. promoters, ribosomal binding sites, terminators, RNase cleavage sites etc.), coding sequences (CDSs) and chassis (hosts), and variation can occur both for relative and absolute part abundance (e.g. different mRNA levels) as well as part identity (e.g. different genes homologues).

This variation of several parts at the same time leads to the spanning of a multidimensional search space (reflected by the corresponding variant library, see Figure 1), which can be subsequently navigated by appropriate screening heuristics in order to identify mutants with improved traits. Therefore "combinatorial pathway optimization" should be clearly distinguished from optimization procedures in which several pathway components are optimized individually in a sequential manner ("de-bottlenecking", Figure 1). Moreover, we would like to emphasize that combinatorial pathway optimization can also be performed on non-genetic variables (e.g. reaction conditions etc.), which is undoubtedly an important part of process optimization but shall not be discussed in this review.

<\Glossary Box 1>

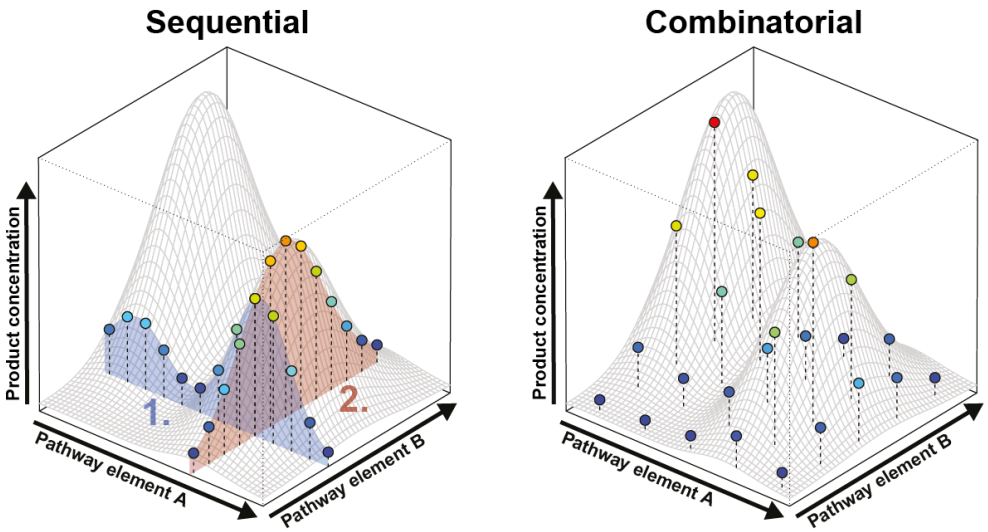


Figure 1: Sequential and combinatorial pathway optimization. In sequential pathway optimization (left) a pathway element is varied individually (blue curve) until a local optimum is identified, which is then fixed as a starting point for the optimization of a second element (red curve) and so forth. By contrast, in combinatorial optimization (right) multiple elements are varied simultaneously which allows to systematically screen the multidimensional space for an optimum (red point) that is not accessible by successive heuristics. Points indicate 25 experimental measurements for both strategies.

In this review we summarize recent developments in the field of combinatorial pathway engineering. We discuss important diversification strategies as well as heuristics to cope with the problem of combinatorial explosion, which are indispensable to keep the experimental

effort manageable, on the basis of selected recent studies. We intend to provide useful guidelines for pathway optimization rather than claiming that available strategies are exhaustively covered.

Strategies for diversification

In order to create genetic diversity within the respective pathways of interest different basic strategies can be distinguished (Figure 2), which tackle i) the identity and properties of utilized coding sequences (or their encoded gene products), ii) the expression level of the involved genes by variation of gene dosage as well as transcriptional or translational engineering, and iii) combined strategies that tackle the pathway on different levels concomitantly. These strategies shall be elucidated here on the basis of recent metabolic engineering studies, which are also summarized in Table 1.

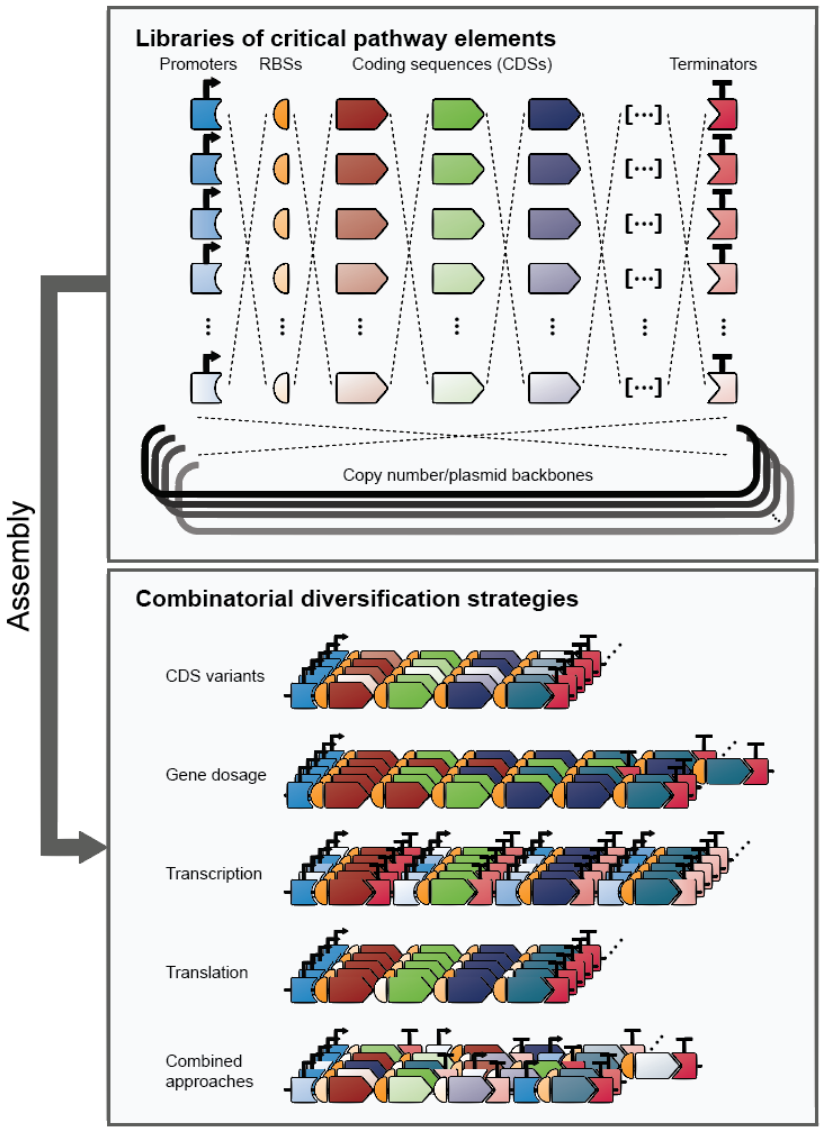


Figure 2: Diversification strategies for combinatorial pathway optimization. In order to generate diversity, libraries of critical pathway elements (e.g. plasmid backbones, promoters, RBSs, coding sequences, terminators etc.) are assembled to yield a suitable combinatorial library. Consequently, variability can be introduced on different levels of the pathway (e.g. gene copy number, transcriptional and translational regulation, coding sequence) and frequently several levels are tackled simultaneously in combined approaches.

Variation of coding sequences

Two basic strategies exist for the variation of coding sequences within a pathway: the first employs different structural or functional gene homologues, whose corresponding enzymes are known (or suspected) to catalyze the respective reaction steps [16,17•,18,19]. In the absence of suitable candidates for a desired reaction metagenomics libraries can be exploited to identify appropriate biocatalysts. This strategy was for instance used to graft xylose utilization into *Saccharomyces cerevisiae* by re-combining potential enzymes of the responsible three-step pathway from more than 20 different organisms [16]. Feher and co-workers compellingly demonstrated that computer-aided design can be used to identify suitable gene homologues from enzyme and pathway databases at a comparably low experimental screening effort, which enabled combinatorial optimization of flavonoid production in *Escherichia coli*. This led to the retrieval of several strains capable of pinocembrin production, albeit at low initial titers, which could be improved 17-fold to up to 24.1 mg L^{-1} in a second round of computational optimization [17•].

The second strategy uses random or targeted mutagenesis to diversify coding sequences and ultimately enzyme activity and function, as exemplified by the simultaneous error-prone diversification of three genes to improve lycopene production [20] and the directed mutagenesis of multiple sites in three genes involved in heme biosynthesis [21] in *E. coli*. Unfortunately, both of these studies do not solely tackle the CDSs but also regulatory elements (promoters and ribosomal binding sites) of the involved genes, and consequently the reasons for the observed improvements remain elusive and may well be attributed to mere changes in relative and/or absolute expression levels rather than functional amino acid substitutions.

After all, studies, in which multiple coding sequences are simultaneously varied, remain complicated and therefore scarce. On the one hand, preparation of combinatorial homologue libraries is difficult since physical acquisition of large numbers of genes from different organisms is either experimentally complicated (e.g. for cloning from the natural hosts) or costly (for commercial gene synthesis). On the other hand, diversification of multiple CDSs by mutagenesis, while comparably simple in preparation, very rapidly leads to combinatorial library sizes that exceed any available screening methods.

Engineering of expression levels

Setting up a balanced pathway with a high flux towards a desired product requires optimization of the relative and absolute expression levels of involved genes. This fine-tuning can be

achieved on different levels, which are listed below and which differ in important characteristics that should be considered upon selection of the optimization strategy to be applied.

Gene dosage

A commonly applied strategy in metabolic engineering to influence the number of enzyme copies per cell is the variation of the dosage of the corresponding genes. The latter ranges from one-copy systems, in which the genes of interest are integrated into the host genome or single-copy vectors (e.g. bacterial/yeast artificial chromosomes), to plasmids with multiple hundred copies per cell. A frequently applied procedure is the combination of plasmids with different copy numbers to balance the expression of different pathway modules (for modularization of pathways see paragraph below) in *E. coli* [22,23•] and *S. cerevisiae* [24]. Alternatively, gene copy number diversification has been achieved by multiple chromosomal integration of expression cassettes as a result of antibiotic selection pressure, which has been used to improve production of carotenoids and isobutanol in yeast [25,26].

However, the number of available orthogonal plasmid systems with different copy numbers is low and likewise chromosomal integration is limited to a small number of copies per genome due to recombination problems. Consequently, combinatorial optimization of expression levels based on gene copy number remains difficult and laborious. In order to overcome the somewhat “static” nature of the gene dosage, a plasmid system that allows for conditional adjustment of the gene copy number has been developed for *E. coli* [27].

Transcriptional engineering

The most widely exploited strategy for expression level fine-tuning is based on alteration of gene transcript levels and has mainly been done by direct engineering of promoters (i.e. transcription efficiency) [28]. Alternatively, introduction of hairpins and RNase sites into the mRNA (i.e. transcript stability) [29,30] as well as engineering of the global transcription machinery by mutation of the σ^{70} factor [31] have been suggested. However, the two latter strategies have not been followed up extensively in recent years whereas promoter engineering remains a crucial tool for pathway engineering.

To this end several promoter libraries have been created that allow spanning of several orders of magnitude in expression levels [7•,28,32-35,36••,37,38] and numerous studies have applied promoters or promoter libraries for the combinatorial optimization of pathways in *E. coli* [7•,20-22,35,38,39••,40-42,43•] and yeast [24,34,36••,44,45]. Particularly noteworthy recent examples include the optimization of violacein production [7•,36••] and lycopene biosynthesis [24] in *E. coli* as well as xylose and cellobiose utilization in *S. cerevisiae* [34,44].

It should be mentioned that promoter engineering represents a highly versatile method since it allows expression both in conditional (i.e. inducible) as well as constitutive manner and can

be combined with other strategies for expression level and/or CDS variation as shown in several studies [20,21,24,35,39••,43•]. Drawbacks of promoter engineering include the necessity to express additional proteins (repressors, activators) for inducible systems, which can impose a substantial additional "work-load" on the host cell in particular when multiple different systems are required to allow differential expression of several pathway elements (e.g. for combinatorial optimization). Constitutive promoters are therefore frequently used to avoid this additional complexity, but the latter can lead to growth inhibition and do not allow to separate biomass formation from the actual production phase, which is an important feature in many industrial processes.

Translational engineering

A relatively young trend in metabolic engineering is the targeted manipulation of translation in order to alter expression levels. In bacteria this is mainly done by engineering of the RBS, which contains the well-known Shine-Dalgarno sequence upstream of the start codon. Engineering of the RBS has several advantages including accessibility of a wide range of expression levels by few base changes, individual adjustment of genes in polycistronic operons and applicability in a wide range of prokaryotic hosts [9••]. Several recent studies apply RBS libraries for combinatorial optimization of pathways in *E. coli* [9••,13,18,21,23•,35,39••,43•,46-48,49••,50] and other prokaryotic hosts [49••,51]. The small number of bases that need to be altered to achieve large dynamic ranges of expression levels makes RBS engineering a practicable approach for multiplexed genome engineering as initially demonstrated by Wang and coworkers [13] and later by other groups [49••,50]. Moreover, Gill and colleagues have recently suggested a method that allows to map genomic mutations in a multiplexed manner, which represents an interesting technology to track diversity within mutant populations [50,52].

A very compelling attribute of RBSs as engineering targets is the availability of predictive biophysical models that allow relative adjustment of expression levels by forward design [53-56]. Based on these works a number of studies have appeared that use *in silico* designed RBSs (or libraries thereof) for combinatorial pathway engineering [9••,18,39••,43•,46-48,49••]. Despite these rapid developments in translational pathway engineering in the past years, the available methods for non-bacterial hosts such as yeasts remain scarce. The 5' untranslated region (5' UTR) has been shown to have significant influence on expression levels in *S. cerevisiae*, which can be used to influence gene expression, for instance by minimization of secondary structures according to thermodynamic models or by construction of suitable aptamers and riboswitches in the 5' UTR of the mRNA [57-59]. However, the systematic exploitation of the 5' UTR for combinatorial pathway optimization remains to be demonstrated.

Combined and integrated approaches

Importantly, the aforementioned approaches to create diversity in CDSs and expression levels must be viewed upon as complementary rather than mutually exclusive for combinatorial pathway optimization. This is reflected by several studies that simultaneously integrate different methods for diversity creation thereby achieving substantial improvements for the respective applications [18,21,22,24,35,39••,43•]. The combinatorial refactoring of a 16-gene pathway for nitrogen fixation from *Klebsiella oxytoca* in *E. coli* achieving 57 % of wild-type activity by combining multiple approaches including variation of promoters, RBSs, terminators and gene order represents a compelling example [39••]. Following up on these works Woodruff and coworkers recently introduced a pipeline for the multiplexed assembly of large combinatorial and sequence-perfect (i.e. no mutations) operon structures from universal pools of composite parts by a barcode based dial-out PCR method [43•]. This allowed for the rapid combinatorial optimization complex multicomponent systems as demonstrated on the examples of nitrogen fixation and the design of complex genetic circuits.

However, it should be mentioned that introducing diversity on multiple levels inevitably leads to large theoretical library sizes, which cannot be exhaustively screened by any available experimental method, and consequently smart strategies for search space reduction need to be applied to keep the screening effort manageable (see below).

Man versus statistics – strategies for the reduction of experimental effort

A major constraint of combinatorial pathway optimization is the number of permutations that need to be screened for their behavior in order to identify the optimal (or best) clone for a desired application. The latter increases linearly with the number of different variants to be tested per pathway component and exponentially with the number of different pathway components to be optimized [9••]. This problem, commonly referred to as “combinatorial explosion”, renders full factorial searches in many cases infeasible, even for high-throughput screening methodologies, and therefore strategies to reduce the experimental effort are inevitably required [7•,9••,36••,49••,60]. Experimental effort can be saved on different stages throughout the optimization process, which is schematically illustrated in Figure 3, and important recent studies that apply respective techniques to improve combinatorial screening efficiency are listed in Table 1.

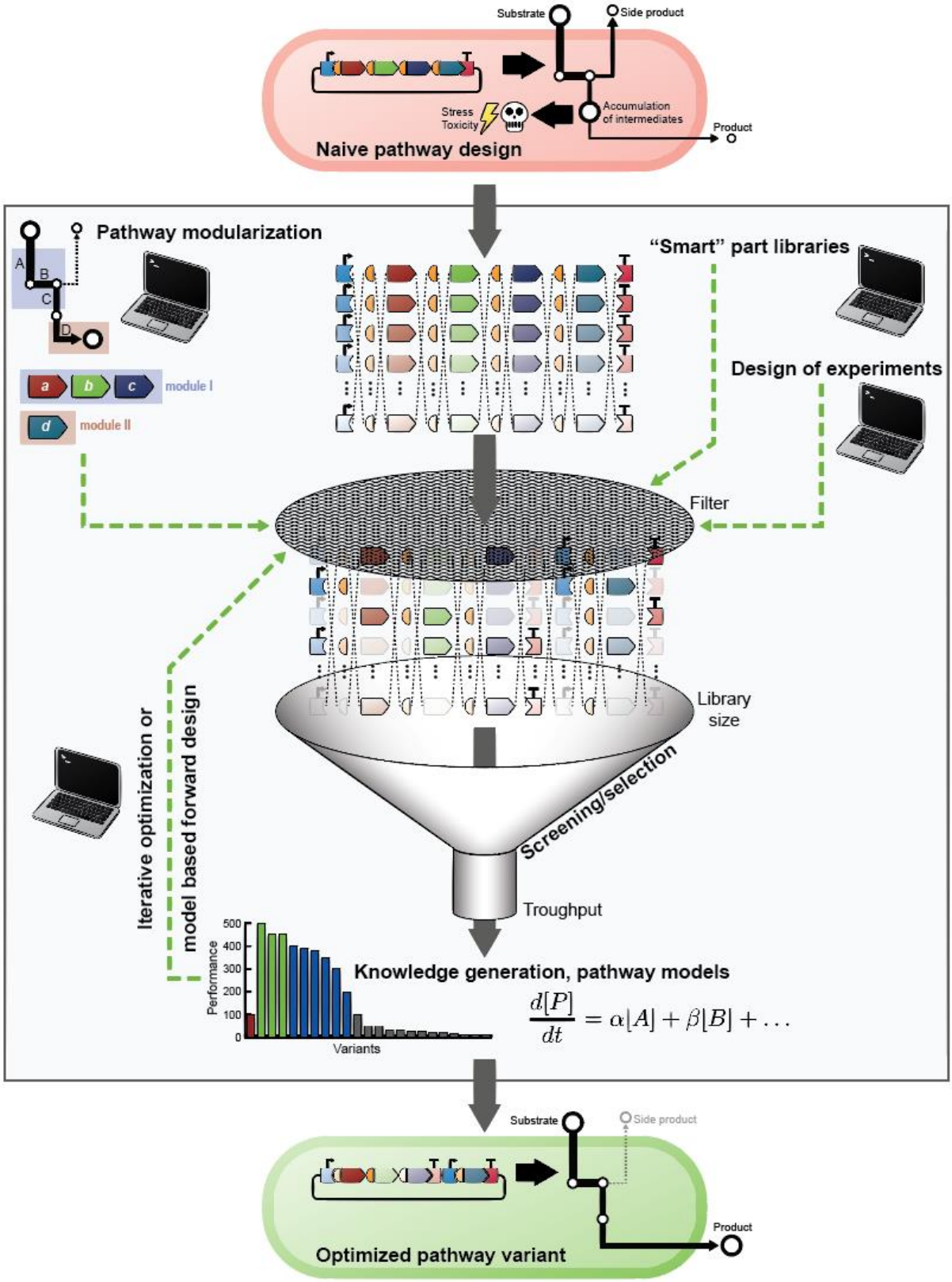


Figure 3: Schematic workflow of combinatorial pathway optimization. Different rational approaches such as pathway modularization, "smart" part library construction and design of experiments (DoE) can be used to reduce the initial combinatorial library size and/or minimize the number of required experiments without extensive *a priori* knowledge about the pathway. The results from the subsequent screening/selection campaign can be harnessed to collect knowledge about the pathway, which in turn can be used to improve library design for iterative optimization cycles and even model-based forward design of optimal constructs. As a consequence combinatorial production landscapes can be efficiently searched and optimized pathway variants are obtained at a minimized collective effort.

A commonly used “trick” is the merging of several enzymatic reaction steps to pathway modules, which are treated as “black boxes” that are only defined by the fluxes coming in and out of the module. This simplification of the pathway topology can subsequently be used by optimizing the different pathway modules at large without the need to fine-tune single enzymatic steps individually [8]. Following early works of Ajikumar *et al.* [22] modularization strategies such as “multivariate modular metabolic engineering” (MMME) (for in-depth reviews about pathway modularization please refer to [5,8,61]) have been widely exploited for combinatorial pathway engineering in the past years (e.g. [7•,20,22,23•,24,62-64]). For instance, fatty acid synthesis was improved in *E. coli* (20-fold in shake flask cultivation) by combinatorial optimization of three modules thereby drastically reducing the effort for the engineered 15-step pathway resulting in a maximum titer of 8.6 g L^{-1} [23•]. Importantly, modularization is not a trivial procedure since the choice of module nodes (i.e. metabolites between two modules) is crucial for effective optimization [5]. Ideally, this should be done based on rational knowledge from experimental data and/or *in silico* flux models, which can help to identify crucial bottlenecks in the pathway and define meaningful modules.

Another strategy is based on the design of libraries with a high content of meaningful options for each diversified pathway component. These “smartly condensed” libraries ensure that the number of options to be tested per pathway component (“resolution”) remains limited. This has been done using pre-characterized sets of parts that allow uniformly discretizing the multidimensional expression level space. This was demonstrated for promoters [34,36••,39••,65••] and RBSs [23•,39••,46]. Importantly, pre-characterization should be done individually for all CDSs due to the strong context dependency of regulatory elements [9••,53]. Similarly, highly condensed libraries of CDS variants can be produced from large sets of homologues based on metadata from enzyme and pathway databases by computer-aided design [17•]. Recent works by Farasat and coworkers [49••] as well as our group [9••] describe algorithms for the predictive *in silico* design of degenerate RBS sequences, which allows covering several orders of magnitude in expression levels by only testing few genetic variants. Despite significant differences in the computational approaches in these two studies, both enable facile and efficient searches through multidimensional expression level spaces without the need of anterior characterization of RBSs as demonstrated on the test-beds of neurosporene and violacein production in *E. coli*, which were improved to up to 517 µg $(g_{DCW} \cdot h)^{-1}$ [49••] and over 91% purity [9••], respectively.

Finally, different heuristic procedures can (and must) be applied to keep the overall experimental effort limited. These include iterative optimization cycles, which are hierarchical in the sense that results from the preceding iteration determine the design of the next round of screening. Using a set of pre-characterized promoters, Lee and colleagues elegantly demonstrated that “bold” under-sampling (< 3%) of a five-dimensional expression level space

for violacein production in yeast is sufficient to train a linear regression model with good predictive properties allowing to efficiently obtain various desired phenotypes with altered product spectrum [36••]. Similarly, strategies of initial “coarse” sampling of production landscapes and subsequent fine-tuning of the pathway have been successfully applied [9••,41,49••]. To this end, different techniques of design of experiment (DoE) have recently been used in order to identify critical pathway parameters and drastically reduce the number of experiments required for optimization [7•,65••]. Zhou *et al.* used DoE to combinatorially diversify six genes and three media conditions to characterize bottlenecks and optimize performance of a pathway for 6-aminocaproate 5-fold (from 9 mg L^{-1} to 48 mg L^{-1}) [65••] and Xu and coworkers used a statistical DoE approach to establish a model with strong predictive power for the 5-step production pathway for violacein (3.2-fold improvement in shake flasks and up to 1.31 g L^{-1} in a bioreactor) [7•].

Future prospects

With this survey, we intended to elucidate the field of combinatorial pathway optimization by highlighting commonly applied techniques to introduce combinatorial genetic diversity into pathways as well as strategies to reduce the subsequent experimental screening effort required to obtain variants with improved behavior.

Due to the ever-decreasing cost of *de novo* DNA synthesis, the appearance of more and more sophisticated DNA assembly technology and the ongoing standardization of genetic parts, the bottleneck in the field of metabolic engineering will continue shifting from physical acquisition of potentially interesting pathway variants (“limited by construction”) towards their highly parallelized and automated testing by screening or selection (“limited by evaluation”). Moreover, metabolic systems of higher complexity will become subject to systematic engineering including large pathways, intricate metabolic networks, and eventually entire organisms. As a consequence, one can foresee a strongly increasing demand for improved high-throughput screening methodologies and, more importantly, for heuristic methods, such as those introduced towards the end of this review, which facilitate efficient navigation of the vast combinatorial space associated with these complicated systems in a rational manner. This will inevitably require a well-orchestrated consortium of methods for rational pathway modularization and genetic part generation, design-of-experiments and flux modeling to name but a few, as well as iteration as a basis for knowledge generation, which will in turn manifest in the form of new and improved designs. It should be emphasized that these aforementioned techniques for the reduction of experimental effort must be applied in a complementary rather than a competitive manner in order to enable metabolic engineers of the future to successfully navigate the enormous complexity of metabolic systems and to tap the vast potential of biological entities.

Table 1: Selected recent studies about combinatorial pathway optimization.

Product/Application	Host(s) [vector]	Combinatorial diversification target(s)	Key features/remarks	Reference
Xylose utilization Ethanol production	*S. cerevisiae* [plasmid]	CDS homologues	3-step pathway optimized using CDS homologues	[16]
Pinocembrin	*E. coli* [plasmid]	CDS homologues	4-step pathway optimized with retrosynthetic CAD tool that identifies promising candidates from enzyme databases	[17•]
Isobutanol/mevalonate production	*S. cerevisiae* [genome]	Copy number	Multiple chromosomal integration of expression cassettes by selective pressure, optimization of 5-step and 8-step pathway	[26]
Xylose/cellobiose utilization Ethanol production	*S. cerevisiae* [plasmid]	Promoter	3-step and 2-step pathway, pre-characterized promoters	[34]
Violacein	*S. cerevisiae* [plasmid]	Promoter	Linear regression models based on random undersampling, 5-step pathway	[36••]
Xylose utilization Ethanol production	*S. cerevisiae* [plasmid]	Promoter	3-step and 8-step pathway, pre-characterized promoters	[44]
6-aminocaproic acid	*E. coli* [plasmid]	Promoter	Design of experiments, predictive model, 6-step pathway	[65••]
Violacein	*E. coli* [plasmid]	Promoter	Statistical design of experiments, predictive model, 5-step pathway	[7•]
Lycopene	*E. coli* [genome]	RBS	24 RBSs simultaneously targeted with degenerate oligonucleotides	[13]
Carotenoid	*E. coli* [plasmid]	RBS	7-step pathway, pre-characterized RBSs, iterative assembly, genotyping barcode	[46]
Neurosporene	*E. coli* [plasmid+genome]	RBS	Predictive RBS library design, 3-step pathway, iterative optimization	[49••]
Violacein	*E. coli* [plasmid]	RBS	Predictive design of uniform RBS libraries, 3-step pathway, iterative optimization	[9••]
Taxadiene	*E. coli* [plasmid+genome]	Copy number, Promoter	10-step pathway divided into two modules	[22]
Fatty acids	*E. coli* [plasmid]	Copy number, Promoter, RBS	15-step pathway divided into three modules, expression is optimized on different levels	[23•]
Nitrogen fixation	*E. coli* [plasmid]	Promoter, RBS, terminator, gene order	full refactoring of 16-step native pathway, hierarchical cluster assembly	[39••] ([43•])
Lycopene	*E. coli* [plasmid]	RBS, CDS homologues, gene order	PCR-free assembly, 4-step pathway	[18]
Heme	*E. coli* [plasmid]	Promoter, RBS, CDS mutants	PCR-based multi-site mutagenesis, 3-step pathway	[21]

References and recommended reading

1. Keasling JD: **Manufacturing molecules through metabolic engineering**. *Science* 2010, **330**:1355-1358.
2. Nielsen J, Keasling JD: **Engineering cellular metabolism**. *Cell* 2016, **164**:1185-1197.
3. Way JC, Collins JJ, Keasling JD, Silver PA: **Integrating biological redesign: Where synthetic biology came from and where it needs to go**. *Cell* 2014, **157**:151-161.
4. Pitera DJ, Paddon CJ, Newman JD, Keasling JD: **Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli***. *Metab Eng* 2007, **9**:193-207.
5. Biggs BW, De Paepe B, Santos CNS, De Mey M, Ajikumar PK: **Multivariate modular metabolic engineering for pathway and strain optimization**. *Curr Opin Biotechnol* 2014, **29**:156-162.
6. Dai ZJ, Nielsen J: **Advancing metabolic engineering through systems biology of industrial microorganisms**. *Curr Opin Biotechnol* 2015, **36**:8-15.
7. Xu P, Rizzoni EA, Sul SY, Stephanopoulos G: **Improving metabolic pathway efficiency by statistical**
• **model-based multivariate regulatory metabolic engineering**. *ACS Synth Biol* 2017, **6**:148-158.
In this work Xu and colleagues use promoter engineering and statistical design of experiment to optimize a five-step enzymatic route to the pigment violacein in E. coli. On the basis of a fractional combinatorial library a predictive model for violacein production is established at comparably low experimental effort.
8. Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G: **The future of metabolic engineering and synthetic biology: towards a systematic practice**. *Metab Eng* 2012, **14**:233-241.
9. Jeschek M, Gerngross D, Panke S: **Rationally reduced libraries for combinatorial pathway**
•• **optimization minimizing experimental effort**. *Nat Commun* 2016, **7**:11163.
The authors introduce a computational method to produce predictive RBS libraries that uniformly span large expression level ranges. The algorithm designated RedLibs is used for the optimization of multi-protein systems firstly for fluorescent proteins and subsequently in a two-step iterative optimization of violacein biosynthesis in E. coli.
10. Kosuri S, Church GM: **Large-scale *de novo* DNA synthesis: technologies and applications**. *Nat Methods* 2014, **11**:499-507.
11. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome**. *Proc Natl Acad Sci U S A* 2008, **105**:20404-20409.
12. Casini A, Storch M, Baldwin GS, Ellis T: **Bricks and blueprints: methods and standards for DNA assembly**. *Nat Rev Mol Cell Biol* 2015, **16**:568-576.
13. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution**. *Nature* 2009, **460**:894-898.
14. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**:816-821.
15. Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, Hsu PD, Wu XB, Jiang WY, Marraffini LA, *et al.*: **Multiplex genome engineering using CRISPR/Cas systems**. *Science* 2013, **339**:819-823.
16. Kim B, Du J, Eriksen DT, Zhao HM: **Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels**. *Appl Environ Microbiol* 2013, **79**:931-941.
17. Feher T, Planson AG, Carbonell P, Fernandez-Castane A, Grigoras I, Dariy E, Perret A, Faulon JL:
• **Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering**. *Biotechnol J* 2014, **9**:1446-1457.
In this work Feher and colleagues introduce the retro-synthetic computational tool RetroPath to construct pathways for user-specified target compounds of choice in silico. RetroPath filters large enzyme and pathway databases to identify suitable candidates and allows to

drastically reduced the number of homologues to be experimentally tested, which is demonstrated on the example of pinocembrin biosynthesis in E. coli.

18. Zhang SS, Zhao XJ, Tao Y, Lou CB: **A novel approach for metabolic pathway optimization: Oligo-linker mediated assembly (OLMA) method**. *J Biol Eng* 2015, **9**:23.
19. Zhao SJ, Jones JA, Lachance DM, Bhan N, Khalidi O, Venkataraman S, Wang ZT, Koffas MAG: **Improvement of catechin production in *Escherichia coli* through combinatorial metabolic engineering**. *Metab Eng* 2015, **28**:43-53.
20. Lv XM, Gu JL, Wang F, Xie WP, Liu M, Ye LD, Yu HW: **Combinatorial pathway optimization in *Escherichia coli* by directed co-evolution of rate-limiting enzymes and modular pathway engineering**. *Biotechnol Bioeng* 2016, **113**:2661-2669.
21. Jin P, Kang Z, Zhang JL, Zhang LP, Du GC, Chen J: **Combinatorial evolution of enzymes and synthetic pathways using one-step PCR**. *ACS Synth Biol* 2016, **5**:259-268.
22. Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: **Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli***. *Science* 2010, **330**:70-74.
23. Xu P, Gu Q, Wang WY, Wong L, Bower AGW, Collins CH, Koffas MAG: **Modular optimization of**
- **multi-gene pathways for fatty acids production in *E. coli***. *Nat Commun* 2013, **4:**1409.
  Xu and coworkers tackle a 15-step pathway for fatty acid biosynthesis pathway in E. coli by re-casting it into three modules. Copy number, promoters and RBSs are combinatorially varied to optimize fatty acid production.
24. Chen XL, Zhu P, Liu LM: **Modular optimization of multi-gene pathways for fumarate production**. *Metab Eng* 2016, **33**:76-85.
25. Yuan JF, Ching CB: **Combinatorial engineering of mevalonate pathway for improved amorpha-4,11-diene production in budding yeast**. *Biotechnol Bioeng* 2014, **111**:608-617.
26. Yuan JF, Ching CB: **Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds**. *ACS Synth Biol* 2015, **4**:23-31.
27. Kittleson JT, Cheung S, Anderson JC: **Rapid optimization of gene dosage in *E. coli* using DIAL strains**. *J Biol Eng* 2011, **5**:10.
28. Alper H, Fischer C, Nevoigt E, Stephanopoulos G: **Tuning genetic control through promoter engineering**. *Proc Natl Acad Sci U S A* 2005, **102**:12678-12683.
29. Smolke CD, Carrier TA, Keasling JD: **Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures**. *Appl Environ Microbiol* 2000, **66**:5399-5405.
30. Pfleger BF, Pitera DJ, D Smolke C, Keasling JD: **Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes**. *Nat Biotechnol* 2006, **24**:1027-1032.
31. Alper H, Stephanopoulos G: **Global transcription machinery engineering: A new approach for improving cellular phenotype**. *Metab Eng* 2007, **9**:258-267.
32. Cox RS, Surette MG, Elowitz MB: **Programming gene expression with combinatorial promoters**. *Mol Syst Biol* 2007, **3:**145.
33. Ellis T, Wang X, Collins JJ: **Diversity-based, model-guided construction of synthetic gene networks with predicted functions**. *Nat Biotechnology* 2009, **27**:465-471.
34. Du J, Yuan YB, Si T, Lian JZ, Zhao HM: **Customized optimization of metabolic pathways by combinatorial transcriptional engineering**. *Nucleic Acids Res* 2012, **40:**e142.
35. Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai QA, Tran AB, Paull M, Keasling JD, Arkin AP, *et al.*: **Precise and reliable gene expression via standard transcription and translation initiation elements**. *Nat Methods* 2013, **10**:354-360.
36. Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE: **Expression-level optimization of a multi-**
- **enzyme pathway in the absence of a high-throughput assay**. *Nucleic Acids Res* 2013, **41**:10668-10678.
  In this study a pre-characterized promoters are used to optimize a five-enzyme route for violacein production in yeast. On the basis of random under-sampling a linear regression

model with good predictive powert is established, which allowed for the construction of a highly improved pigment producing variant.

37. Redden H, Alper HS: **The development and characterization of synthetic minimal yeast promoters**. *Nat Commun* 2015, **6**:7810.
38. Yu JH, Zhu LW, Xia ST, Li HM, Tang YL, Liang XH, Chen T, Tang YJ: **Combinatorial optimization of CO_2 transport and fixation to improve succinate production by promoter engineering**. *Biotechnol Bioeng* 2016, **113**:1531-1541.
39. Smanski MJ, Bhatia S, Zhao DH, Park Y, Woodruff LBA, Giannoukos G, Ciulla D, Busby M, Calderon
•• J, Nicol R, et al.: **Functional optimization of gene clusters by combinatorial design and assembly**. *Nat Biotechnol* 2014, **32**:1241-U1104.
A heterologous gene cluster for nitrogen fixation is re-factored in E. coli by constraint-based combinatorial design and assembly of up to 16 genes. Smanski and coworkers permute promoters, RBSs, terminators and the order of genes to establish nitrogen fixation in E. coli demonstrating pathway optimization at hithertho unmatched complexity.
40. Coussement P, Maertens J, Beauprez J, Van Bellegem W, De Mey M: **One step DNA assembly for combinatorial metabolic engineering**. *Metab Eng* 2014, **23**:70-77.
41. Freestone TS, Zhao H: **Combinatorial pathway engineering for optimized production of the anti-malarial FR900098**. *Biotechnol Bioeng* 2016, **113**:384-392.
42. Coussement P, Bauwens D, Maertens J, De Mey M: **Direct combinatorial pathway optimization**. *ACS Synth Biol* 2016.
43. Woodruff LB, Gorochowski TE, Roehner N, Mikkelsen TS, Densmore D, Gordon DB, Nicol R, Voigt
• CA: **Registry in a tube: multiplexed pools of retrievable parts for genetic design space exploration**. *Nucleic Acids Res* 2017, **45**:1553-1565.
Following up on the works of Smanski et al. [39••], Woodruff and colleagues introduce a pipeline for the multiplexed assembly of complex combinatorial and sequence-perfect operons structures from universal pools of composite parts by a barcode based dial-out PCR method as demonstrated on the test-bed of nitrogen fixation in E. coli.
44. Latimer LN, Lee ME, Medina-Cleghorn D, Kohnz RA, Nomura DK, Dueber JE: **Employing a combinatorial expression approach to characterize xylose utilization in *Saccharomyces cerevisiae***. *Metab Eng* 2014, **25**:20-29.
45. EauClaire SF, Zhang J, Rivera CG, Huang LL: **Combinatorial metabolic pathway assembly in the yeast genome with RNA-guided Cas9**. *J Ind Microbiol Biotechnol* 2016, **43**:1001-1015.
46. Zelcbuch L, Antonovsky N, Bar-Even A, Levin-Karp A, Barenholz U, Dayagi M, Liebermeister W, Flamholz A, Noor E, Amram S, et al.: **Spanning high-dimensional expression space using ribosome-binding site combinatorics**. *Nucleic Acids Res* 2013, **41**:e98.
47. Li YF, Gu Q, Lin ZQ, Wang ZW, Chen T, Zhao XM: **Multiplex iterative plasnnid engineering for combinatorial optimization of metabolic pathways and diversification of protein coding sequences**. *ACS Synth Biol* 2013, **2**:651-661.
48. Nowroozi FF, Baidoo EEK, Ermakov S, Redding-Johanson AM, Batth TS, Petzold CJ, Keasling JD: **Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly**. *Appl Microbiol Biotechnol* 2014, **98**:1567-1581.
49. Farasat I, Kushwaha M, Collens J, Easterbrook M, Guido M, Salis HM: **Efficient search, mapping,**
•• **and optimization of multi-protein genetic systems in diverse bacteria**. *Mol Syst Biol* 2014, **10**:731.
Based on a biophysical model for the prediction of RBS strength, Farasat and coworkers introduce an algorithm to generate RBS libraries in silico, which allow to span large ranges of expression levels. Applicability is demonstrated for expression level optimization in different organisms as well as on plasmid and genome level. Efficient searching of a multi-protein expression level space is shown on the combinatorial optimization of neurosporene biosynthesis in an iterative manner relying on sequence-expression-activity maps and a kinetic pathway model.

50. Zeitoun RI, Garst AD, Degen GD, Pines G, Mansell TJ, Glebes TY, Boyle NR, Gill RT: **Multiplexed tracking of combinatorial genomic mutations in engineered cell populations**. *Nat Biotechnol* 2015, **33**:631-637.
51. Oliver JWK, Machado IMP, Yoneda H, Atsumi S: **Combinatorial optimization of cyanobacterial 2,3-butanediol production**. *Metab Eng* 2014, **22**:76-82.
52. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG, *et al.*: **Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering**. *Nat Biotechnol* 2017, **35**:48-55.
53. Salis HM, Mirsky EA, Voigt CA: **Automated design of synthetic ribosome binding sites to control protein expression**. *Nat Biotechnol* 2009, **27**:946-U112.
54. Na D, Lee D: **RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression**. *Bioinformatics* 2010, **26**:2633-2634.
55. Seo SW, Yang JS, Kim I, Yang J, Min BE, Kim S, Jung GY: **Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency**. *Metab Eng* 2013, **15**:67-74.
56. Borujeni AE, Channarasappa AS, Salis HM: **Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites**. *Nucleic Acids Res* 2014, **42**:2646-2659.
57. Crook NC, Freeman ES, Alper HS: **Re-engineering multicloning sites for function and convenience**. *Nucleic Acids Res* 2011, **39**:e92.
58. Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, Weinberger A, Segal E: **Deciphering the rules by which 5 '-UTR sequences affect protein expression in yeast**. *Proc Natl Acad Sci U S A* 2013, **110**:E2792-E2801.
59. Redden H, Morse N, Alper HS: **The synthetic biology toolbox for tuning gene expression in yeast**. *FEMS Yeast Res* 2015, **15**.
60. Young E, Lee SM, Alper H: **Optimizing pentose utilization in yeast: the need for novel tools and approaches**. *Biotechnol Biofuels* 2010, **3**:24.
61. Boock JT, Gupta A, Prather K: **Screening and modular design for metabolic pathway optimization**. *Curr Opin Biotechnol* 2015, **36**:189-198.
62. Wu J, Liu P, Fan Y, Bao H, Du G, Zhou J, Chen J: **Multivariate modular metabolic engineering of *Escherichia coli* to produce resveratrol from L-tyrosine**. *J Biotechnol* 2013, **167**:404-411.
63. Wu J, Du G, Zhou J, Chen J: **Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy**. *Metab Eng* 2013, **16**:48-55.
64. Zhao J, Li Q, Sun T, Zhu X, Xu H, Tang J, Zhang X, Ma Y: **Engineering central metabolic modules of *Escherichia coli* for improving beta-carotene production**. *Metab Eng* 2013, **17**:42-50.
65. Zhou H, Vonk B, Roubos JA, Bovenberg RA, Voigt CA: **Algorithmic co-optimization of genetic**
•• **constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor**. *Nucleic Acids Res* 2015, **43**:10560-10570.
This study elegantly demonstrates the power of design of experiments: a six-step pathway for 6-aminocaproic acid production is balanced simultaneously with three media parameters. This is achieved by combinatorial promoter screening of a small subset of the theoretical library to identify critical pathway components and subsequently establishing a predictive model. The latter allowed for the straight forward design of a highly improved producer.