HHS Public Access

Author manuscript

*Technology (Singap World Sci)*. Author manuscript; available in PMC 2016 July 21.

Published in final edited form as:

*Technology (Singap World Sci)*. 2016 March ; 4(1): 9–14. doi:10.1142/S233954781640001X.

# High-throughput evaluation of synthetic metabolic pathways

**Justin R. Klesmith**[1] and **Timothy A. Whitehead**[2,3]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[2]Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, MI 48824, USA

[3]Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

## Abstract

A central challenge in the field of metabolic engineering is the efficient identification of a metabolic pathway genotype that maximizes specific productivity over a robust range of process conditions. Here we review current methods for optimizing specific productivity of metabolic pathways in living cells. New tools for library generation, computational analysis of pathway sequence-flux space, and high-throughput screening and selection techniques are discussed.

### Keywords

Metabolic Engineering; Protein Engineering; Synthetic Biology; Deep Mutational Scanning; Computational Optimization

## Introduction

Microorganisms have the potential to produce many chemicals of use to society[1,2]. In some cases, production from heterologous microorganisms is more sustainable than purifying the chemical from natural sources. Examples include harvesting Pacific yew trees or Chinese wormwood for taxol[3] or the anti-malarial artemisinin[4], respectively. Additionally, the ability to create renewable and sustainable biofuels and biochemicals is increasingly attractive given concerns about climate change and peak oil[5].

An organism producing a desired product may not exist, or a given strain may not be suitable for required economical processing conditions. Because of this, reconstructed pathways are often implanted into chassis microorganisms[5]. Some of these pathways include those specific for biofuels (ethanol[5], isobutanol[6], 1-butanol[7], 1,4-butanediol[8]), polymer monomers (polylactic acid[9], isoprene[10], 3-hydroxypropionic acid[11]) and pharmaceutically active ingredients (precursors for taxol[3] or opioids[12]). However, in many cases product toxicity or transport limits end titers, product recovery from aqueous fermentation broths is

Correspondence should be addressed to: T.A.W. (taw@egr.msu.edu).

**Author Contributions**: J.R.K. and T.A.W. wrote the review article.

inefficient, or the volumetric productivity is below that required for a cost-effective process. Combined, these limitations temper the promise of sustainable replacement of the palette of petrochemicals and naturally extracted specialty chemicals currently in use by society.

In particular, the specific productivities of most engineered metabolic pathways are far below that which is needed for industrial production. Some implanted pathways have limited flux because of substantial thermodynamic reversibility at key steps[13]. Additionally, pathway enzymes transplanted into heterologous hosts often have poor expression because of weak catalytic efficiency[14], poor protein solubility, or membrane targeting issues[12,15]. Host-specific problems include cofactor accessibility[16], siphoning of pathway intermediates, intermediate toxicity and post-translational flux regulation of key precursors[12,17,18]. Furthermore, the performance of a specific engineered metabolic pathway may differ between host strains[19], media formulations[12], temperatures and oxygen conditions[20].

A grand challenge in the field of metabolic engineering is the accurate and efficient identification of a pathway genotype that maximizes specific productivity over a robust range of process conditions. Attempts to improve specific productivity have largely focused on screening individual pathway enzymes for activity or balancing gene expression by testing libraries of elements like promoters and ribosome binding sites (RBS)[21–24]. Error-prone polymerase chain reaction (PCR) mutagenesis of pathway enzymes has also been used to find activity-improving mutations[17,18]. However, pathway optimization by total enumeration becomes unwieldy, as balancing activity at multiple nodes leads to a combinatorial explosion. Consider a plasmid-encoded pathway composed of a series of expression elements (e.g. promoters, RBS, terminators) and pathway gene variants (Fig. 1a). A pathway library comprised of a single enzyme of average length[25] driven by 10 alternative promoter and 10 alternative RBS sequences can be covered by testing $10^2$ variants. A library containing the above gene expression genotypes with all possible single non-synonymous mutations to the enzyme now contains $6 \times 10^5$ variants. Testing the same number of variants using a two-enzyme pathway requires a theoretical coverage of $3.6 \times 10^{11}$ variants, which is too large a sequence space to cover under most conditions. The combinatorial problem only becomes more acute with more pathway enzymes. To partially circumvent this combinatorial intractability, modular pathway design has been used to partition individual enzymatic steps into reaction groups. Then, expression of the resulting modules are balanced, not individual enzymes[12,26–28] (Fig. 1b). Alternative ways to explore sequence-flux space include computational predictions from small training sets[29–31] (Fig. 1c) or high-throughput screening or selection techniques[32–36] (Fig. 1d).

The focus of this review is on new technologies that identify highly productive and robust synthetic metabolic pathways. This review will not cover continuous evolution[37], whole-genome engineering[38] or computational pathway design[39] — the interested reader can find excellent reviews on some of these topics elsewhere[40–43]. We begin by describing pathway evaluation of isogenic cultures. We next describe computational approaches to predict high-performing genotypes with a limited training set of pathway variants. Next, we consider high-throughput methods to assess metabolic pathways, including population-based screens or selections. New enabling techniques for DNA library construction, sequencing and evaluation will be described throughout.

## Evaluation of Isogenic Cultures

One way to evaluate pathway variants is through the use of isogenic cultures. In a typical set-up, a library containing a combinatorial library of expression elements or enzyme variants is created, and clonal variants are tested individually. Lu *et al.* optimized a xylose fermentation pathway in *Saccharomyces cerevisiae* by shuffling promoters of various strength in front of each pathway enzyme[44]. Different promoter combinations were made and tested individually for ethanol productivity and enzymatic activity. Solomon *et al.* tested different expression levels of glucokinase (*glk*) and galactose permease (*galp*) to enable glycolytic uptake in *Escherichia coli* independent of the phosphotransferase system[45]. Carbon flux was modulated by controlling expression of *glk* and *galp* under control of synthetic constitutive promoters. Juminaga *et al.* constructed a pathway for L-tyrosine production in *E. coli* MG1655 by modifying plasmid copy numbers, promoter strength, gene codon usage, and placement of genes in operons[46]. The best pathway variant had a volumetric productivity of 55 mg L-tyrosine/L/hr, representing 80% of the theoretical yield. Ajikumar *et al.* optimized a pathway for overproduction of taxadiene, a key taxol precursor[3]. The authors used a modular approach by separating the pathway into two operons, with one containing the methylerythritolphosphate pathway and the other containing genes encoding the downstream terpenoid-producing enzymes. The promoter strength in front of each operon was systematically varied and taxadiene product measured. Notably, the taxadiene production landscape was highly non-linear in response to operon expression.

Similar isogenic approaches can be used to engineer key rate-determining enzymes or transporters in implanted metabolic pathways. Zhang *et al.* used site-directed mutagenesis of active site residues of the enzymes KivD and LeuA[47]. Fermentations of *E. coli* harboring pathways with different combinations of KivD/LeuA variants were tested for quantification of desired alcohol products. Leonard *et al.* generated combinatorial mutations in the downstream enzymes geranylgeranyl diphosphate synthase and levopimaradiene synthase to tune the selectivity and increase the productivity of levopimaradiene production in *E. coli*[48]. The best strain had a maximum volumetric productivity of 7.3 mg levopimaradiene per L per h. Lee *et al.* improved xylose utilization in *S. cerevisiae* by directed evolution of xylose isomerase[49]. After three rounds of error-prone PCR and screening they isolated a mutant with a 61-fold improvement in aerobic growth rate and an eight-fold improvement in ethanol production and xylose consumption. Screening pathway variants is not only limited to enzymes. Young *et al.* demonstrated the tunability of yeast sugar transporters through a combination of motif-based design and saturation mutagenesis[50]. This approach was used to identify xylose-specific fungal molecular transporters, which when expressed improved xylose utilization by *S. cerevisiae*.

## Computational Predictions Using Empirical Training Sets

Adjusting the right balance of enzyme specific activities within a pathway is crucial as the fitness cost of protein expression[51], catabolism of pathway intermediates, and off-product reactions can all lower specific productivities. While there have been many admirable attempts to forward engineering biological systems and parts[22,24,52–54] and analytical equations describing pathway flux have been formulated[55], tuning metabolic pathways is

largely still an empirical exercise. Because of this, predictive computational models have been used to predict high productivity portions of sequence-flux space given sparse flux datasets resulting from testing isogenic cultures. Lee *et al.* used a linear regression model trained on empirical data to relate enzyme expression levels to product titers in a violacein biosynthetic pathway[29]. This simple model could accurately predict promoter combinations resulting in the production of violacein or one of the three alternative products. Another approach to computationally model and improve pathway performance is to correlate targeted proteomics and metabolite data. George *et al.* generated isopentenol pathway variants with differing promoters, operon organization, and codon-usage[30]. They then used HPLC and LC-MS to quantify glucose, organic acids, and pathway intermediates and used selected reaction monitoring mass spectrometry to quantify all proteins in their pathway. Spearman rank correlations were calculated from values of protein area and metabolite concentrations. Based on these relationships, individual variants were reconstructed and tested in time-course experiments to test model predictions. While this method may not capture complex regulatory interactions like feedback inhibition, other methods like $^{13}C$ metabolic flux analysis studies are more than capable of doing so[56,57]. In one example, Feng *et al.* tested different xylose reductase, xylitol dehydrogenase and xylulose kinase variants in a yeast xylose pathway and used $^{13}C$ metabolic flux analysis to determine if the different cofactor requirements of the enzyme variants had any effect on growth and/or ethanol production[58]. They found that production of ethanol was unaffected by the cofactor requirements of the xylose pathway. However, the cofactor balanced xylose pathway allowed growth under a wider variety of conditions. Farasat *et al.* developed a sequence-expression-activity mapping method to find optimal expression conditions with desired activity for a carotenoid biosynthetic pathway[31]. In a first step, an RBS calculator is used to make a library that spans a large range of protein expression space. Next, a subset of the library is tested for activity and used as a training set for a computational model. A new library is then constructed with targeted expression within a narrow window specified by the model. Zelcbuch *et al.* performed an iterative assembly of three fluorescent reporters, each with an associated RBS, into an operon[59]. This initial search reduced the expression search space for a balanced astaxanthin pathway. In a clever approach, they were able to haplotype the individual non-local RBS sequences included within the operon by sequencing a downstream barcode built using iterative restriction and ligation steps.

## Enabling DNA Construction Methods

New genetic modification methods like DNA Assembler[60], Golden Gate assembly[61], Gibson cloning[62], SLIC[63], site-specific recombination or VEGAS assembly[64] enable efficient construction of pathway variants with an array of different enzymes, promoters and RBS sequences. Smanski *et al.* utilized Gibson cloning[62] and Golden Gate assembly[61] to refactor the *Klebsiella oxytoca* nitrogen fixation gene cluster[32] by systematically varying the expression levels of individual genes in the complete 16-gene pathway. Performance of their clusters was assessed by RNA-seq for expression levels and nitrogenase activity assays. The best of the 122 full-length pathways tested resulted in recovery of 57% of the wild-type activity. Layton and Trinh used Gibson cloning to make ester fermentative pathways in *E. coli*[33]. Their modular design of their pathway allowed quick replacement of RBS and

promoter sequences. Oliver *et al.* improved 2,3-butanediol production in cyanobacteria by using SLIC to swap different RBS sequences in front of each pathway enzyme[34]. Colloms *et al.* used serine integrase site-specific recombination to rank gene order and RBS sites for a more efficient production of violacein and lycopene[35]. Du *et al.* used *S. cerevisiae* native homologous recombination to swap promoters of various expression strength in front of relevant genes[65]. This was used to improve xylose and cellobiose utilization pathways. Kim *et al.* used a similar approach to balance the flux of a xylose utilizing pathway for biofuel production[36]. Importantly, the optimal pathway was strongly dependent on both the host genotype but also the sugar composition of the growth medium. Latimer *et al.* combinatorially tested promoters of the eight-gene pathway for xylose utilization in *Saccharomyces cerevisiae*[20]. Library plasmids were made with Golden Gate assembly. Similar to results above, they found that the enrichment of specific yeast promoters in their library after selection was dependent on the number of genes expressed, the culture media conditions, and the cofactor dependence of the enzymes.

## Alternative High-Throughput Screening Methods

Many of the above examples utilized medium-throughput plate-based screening or a growth based selection in order to sort variants. There have been recent developments to employ fluorescence-activated cell sorting (FACS) or microfluidic sorting technology in cases without an observable growth phenotype. For example, Wang *et al.* cultured xylose-consuming strains in droplets and microfluidic sorting based on the fluorescence of oxidized extracellular metabolites[66]. Michener *et al.* utilized FACS to screen improved variants of caffeine demethylase using a designed RNA biosensor[67]. The RNA biosensor is a combination of ribozyme and aptamer located in the 3′ UTR of a fluorescent reporter gene. When the aptamer is bound to a desired ligand, the ribozyme misfolds leading to lower RNA cleavage rates and increasing the fluorescent output. Tang *et al.* utilized FACS to screen for *E. coli* clones with enhanced triacetic acid lactone (TAL) production using a an engineered TAL fluorescent reporter[68]. Jha *et al.* used a FACS screen to identify *E. coli* clones with increased enzymatic production of 3,4 dihydroxy benzoate[69]. In these above examples, the limitation is developing a fluorescent reporter that is coupled to intracellular concentrations of a target metabolite.

## Population-Based Measurements

One limitation of high-throughput screening is the inability to haplotype a unique pathway sequence to an output phenotype. Typically, only a few “winners” of the selection are sequenced. This is sub-optimal for two reasons. First, the winner variants depend strongly on the exact selection or screening conditions used, and so a selection must be repeated for each change of fermentation condition or host genotype. This limitation precludes use of multi-objective optimization (Pareto optimization) techniques into pathway redesign. Recent contributions from the Salis lab show the utility of Pareto optimization to improve translation elongation rates and mRNA stability of a single construct across multiple bacterial species[70,71]. Secondly, high-throughput methods do not allow coverage of a complete sequence-flux space for even moderate-length pathways, and losing crucial genotypic information of the pathway makes it impossible to use the powerful computational

analyses and prediction tools that have been demonstrated for low-throughput pathways. We envision population-based measurements that can more thoroughly search sequence-flux space and also identify Pareto optimal genotypes that are robust to different processing conditions (Fig. 2).

Recent advances in deep sequencing technology allow the ability to track tens of thousands of pathway variants in a high throughput screen. Most of such methods rely upon "barcoding" individual cells with a short unique identifier DNA sequence (Fig. 2a). A growth selection is performed, and these populations are deep sequenced at the barcode locus. The change in frequency of an individual barcode can be related to the fitness of that unique variant[72,73]. While in principle such techniques could be used to track individual metabolic pathway variants, most demonstrations have been for studies on evolution. Smith *et al.* developed a barcode sequencing method (Bar-seq), which they validated by performing growth selections of a mixed culture containing yeast deletion strains[74]. The barcode abundance after selection was determined for each deletion strain by deep sequencing the entire population. More recently, Levy *et al.* barcoded 500,000 lineages of *Saccharomyces cerevisiae* and used a growth selection to track time-dependent changes in fitness among the population[75]. Chubiz *et al.* introduced FREQ-Seq, a method to barcode and determine allele frequencies from a mixed population[72]. FREQ-Seq was used to map seven variants of the enzyme Tet(X2), conferring tetracycline resistance, in ten different evolving populations[76].

Frequency analysis of variants within a population can be used to assess if a single variant improves, reduces, or has no effect on function. This approach has been used for evaluation of yeast translation initiation sites[77] and bacterial promoter strengths[21] by coupling these upstream elements to fluorescent reporter proteins. Subsequently, populations are sorted by FACS. In fact, massively parallel sequence-function mapping is now commonplace in determining the sequence effects on function for proteins[78]. For example, this methodology has been used to improve the affinity of engineered protein binders to Influenza[79]. The question remains how to leverage impressive deep sequencing technology to improve implanted metabolic pathways.

In principle, high-throughput sequence-function mapping can be used to determine metabolic pathways supporting higher or lower flux, provided that it is coupled to a selectable phenotype like growth. A new approach called FluxScan was recently reported[80] that maps the sequence determinants of flux in living cells (Fig. 2b). First, a selection is designed to allow growth if and only if flux is routed through the implanted pathway. A mutational library is then created and transformed into the strain of interest. After a growth selection is performed for 4–10 generations, the entire population is deep sequenced and compared with the population before selection. The frequency change of each variant can be calculated and converted to a flux value. To demonstrate this method, Klesmith *et al.* determined the effect of flux for over 8,000 single-point mutants in a pyrolysis oil catabolic pathway[80]. One designed pathway incorporating 15 beneficial mutations identified from FluxScan supported a 15-fold improvement in growth rate on levoglucosan, a chief pyrolysis oil constituent.

One significant technical challenge with FluxScan and related deep sequencing approaches is the inability to cover the complete length of metabolic pathways: current long read lengths of the Illumina platform are approximately 300 bp, whereas full operons can exceed 10 kb. One method to escape this limitation is to sequence small contiguous regions of sequence (a gene tile) able to fit on a single read. This “tiling” is then repeated along the length of the entire gene encoding sequence[81]. Other potential solutions to extend deep sequencing include coupling a pre-defined barcode sequence to a given pathway variant using clever DNA construction approaches[32,59] or to utilize the next generation of long-read, highly accurate haplotype sequencing technologies[82,83].

The ideal outcome of any experiment should be to find Pareto optimal pathway sequences that are robust and transferrable to any process condition. Single isogenic culturing conditions are not suited for this task as each pathway variant would individually have to be tested under each process condition to determine the resulting phenotype. Therefore, high-throughput population-based measurements are more capable to resolve the fitness of each sequence variant under each process condition (Fig. 2c) provided that they are performed under diverse conditions. Sequence variants from current high-throughput genomic methods that originate from different process conditions highlight this open problem. Gall *et al.* used the SCALEs method to map the gene expression in *E. coli* that conferred an advantage in the presence of 1-naphthol[84]. They show that genes with enhanced expression depend on the type of culturing method used. Only 25% of clones that were reproducibly enriched in serial transfer cultures were similarly enriched in single batch cultures. Similarly, Warner *et al.* used TRMR and found differential gene expression that was dependent on the four different growth conditions under evaluation[85]. While these examples highlight gene expression on the genomic scale, the problem also applies to expression and gene variants on the individual pathway level. An example of this is from the aforementioned FluxScan study where the enrichment of individual mutations from the enzyme levoglucosan kinase was strongly dependent on the biophysical properties of the starting enzyme variant from each selection[80]. Being able to measure the pathway phenotype from different expression elements and individual gene variants under different growth conditions should allow the elucidation of pathway sequences that are optimal over a range of diverse conditions and determine why other sequences fail when these conditions change. If implemented, these high-throughput population-based measurements can help train computational models by providing empirical data and similarly computational models can help reduce the sequence search space for a more targeted population-based screen.

## Outlook

In the near future we believe that robust, high performing pathways can be efficiently identified. New DNA assembly technologies allow for construction of large libraries of pathway variants covering a large range of protein expressions and activities. The number of unique pathway variants that can be made far exceeds that which can be accurately validated using existing technology. In this review, we have covered current methods to reduce the search space.

There is no general method that can assess any metabolic pathway, as there are limitations to each of the main approaches. There are two practical limitations that must be surmounted. First, the relationship between gene expression and pathway flux is highly non-linear. Second, a specific genotype may only support high productivity in a narrow range of process conditions. We suggest that marrying computational modeling with empirical datasets resulting from population-based measurements will allow a more efficient discovery of Pareto optimal gene encoding, expression, or regulatory sequences.

## Acknowledgments

The authors acknowledge support from the National Science Foundation (Awards CBET-1254238, CBET-1236120).

## References

1. Keasling JD. Synthetic biology for synthetic chemistry. ACS Chem Biol. 2008; 3:64–76. [PubMed: 18205292]
2. Markham KA, Alper HS. Synthetic biology for specialty chemicals. Annu Rev Chem Biomol Eng. 2014; 6:35–52. [PubMed: 26083056]
3. Ajikumar PK, et al. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science. 2010; 330:70–74. [PubMed: 20929806]
4. Paddon CJ, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013; 496:528–532. [PubMed: 23575629]
5. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. Microbial engineering for the production of advanced biofuels. Nature. 2012; 488:320–328. [PubMed: 22895337]
6. Trinh CT, Li J, Blanch HW, Clark DS. Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. Appl Environ Microbiol. 2011; 77:4894–4904. [PubMed: 21642415]
7. Atsumi S, et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab Eng. 2008; 10:305–311. [PubMed: 17942358]
8. Yim H, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol. 2011; 7:445–452. [PubMed: 21602812]
9. Jung YK, Kim TY, Park SJ, Lee SY. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol Bioeng. 2010; 105:161–171. [PubMed: 19937727]
10. Lindberg P, Park S, Melis A. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. Metab Eng. 2010; 12:70–79. [PubMed: 19833224]
11. Borodina I, et al. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. Metab Eng. 2015; 27:57–64. [PubMed: 25447643]
12. Thodey K, Galanie S, Smolke CD. A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol. 2014; 10:837–844. [PubMed: 25151135]
13. Bond-Watts BB, Bellerose RJ, Chang MCY. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat Chem Biol. 2011; 7:222–227. [PubMed: 21358636]
14. Milo R, Last RL. Achieving diversity in the face of constraints: Lessons from metabolism. Science. 2012; 336:1663–1667. [PubMed: 22745419]
15. Trenchard IJ, Smolke CD. Engineering strategies for the fermentative production of plant alkaloids in yeast. Metab Eng. 2015; 30:96–104. [PubMed: 25981946]
16. Avalos JL, Fink GR, Stephanopoulos G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. Nat Biotechnol. 2013; 31:335–341. [PubMed: 23417095]
17. Bastian S, et al. Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. Metab Eng. 2011; 13:345–352. [PubMed: 21515217]

18. DeLoache WC, et al. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol. 2015; 11:465–471. [PubMed: 25984720]
19. Paddon CJ, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013; 496:528–532. [PubMed: 23575629]
20. Latimer LN, et al. Employing a combinatorial expression approach to characterize xylose utilization in *Saccharomyces cerevisiae*. Metab Eng. 2014; 25:20–29. [PubMed: 24930894]
21. Kosuri S, et al. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc Natl Acad Sci USA. 2013; 110:14024–14029. [PubMed: 23924614]
22. Mutalik VK, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods. 2013; 10:354–360. [PubMed: 23474465]
23. Lee ME, DeLoache WC, Cervantes B, Dueber JE. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol. 2015; 4:975–986. [PubMed: 25871405]
24. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009; 27:946–950. [PubMed: 19801975]
25. Skovgaard M, Jensen LJ, Brunak S, Ussery D, Krogh A. On the total number of genes and their length distribution in complete microbial genomes. Trends Genet. 2001; 17:425–428. [PubMed: 11485798]
26. Sheppard MJ, Kunjapur AM, Wenck SJ, Prather KLJ. Retro-biosynthetic screening of a modular pathway design achieves selective route for microbial synthesis of 4-methyl-pentanol. Nat Commun. 2014; 5:5031. [PubMed: 25248664]
27. Tseng HC, Prather KLJ. Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. Proc Natl Acad Sci USA. 2012; 109:17925–17930. [PubMed: 23071297]
28. Zhou K, Qiao K, Edgar S, Stephanopoulos G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. Nat Biotechnol. 2015; 33:377–383. [PubMed: 25558867]
29. 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. [PubMed: 24038353]
30. George KW, et al. Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production. Biotechnol Bioeng. 2014; 111:1648–1658. [PubMed: 24615242]
31. Farasat I, et al. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol Syst Biol. 2014; 10:731. [PubMed: 24952589]
32. Smanski MJ, et al. Functional optimization of gene clusters by combinatorial design and assembly. Nat Biotechnol. 2014; 32:1241–1249. [PubMed: 25419741]
33. Layton DS, Trinh CT. Engineering modular ester fermentative pathways in *Escherichia coli*. Metab Eng. 2014; 26:77–88. [PubMed: 25281839]
34. Oliver JWK, Machado IMP, Yoneda H, Atsumi S. Combinatorial optimization of cyanobacterial 2,3-butanediol production. Metab Eng. 2014; 22:76–82. [PubMed: 24412567]
35. Colloms SD, et al. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. Nucleic Acids Res. 2014; 42:e23. [PubMed: 24225316]
36. Kim B, Du J, Eriksen DT, Zhao H. 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. [PubMed: 23183982]
37. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. Nature. 2011; 472:499–503. [PubMed: 21478873]
38. Zeitoun RI, et al. Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. Nat Biotechnol. 2015; 33:631–637. [PubMed: 25798935]
39. Henry CS, Broadbelt LJ, Hatzimanikatis V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng. 2010; 106:462–473. [PubMed: 20091733]

40. Medema MH, van Raaphorst R, Takano E, Breitling R. Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol. 2012; 10:191–202. [PubMed: 22266781]
41. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. Mol Syst Biol. 2009; 5320
42. Woolston BM, Edgar S, Stephanopoulos G. Metabolic engineering: Past and future. Annu Rev Chem Biomol Eng. 2013; 4:259–288. [PubMed: 23540289]
43. Goldsmith M, Tawfik DS. Directed enzyme evolution: Beyond the low-hanging fruit. Curr Opin Struct Biol. 2012; 22:406–412. [PubMed: 22579412]
44. Lu C, Jeffries T. Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. Appl Environ Microbiol. 2007; 73:6072–6077. [PubMed: 17693563]
45. Solomon KV, Moon TS, Ma B, Sanders TM, Prather KLJ. Tuning primary metabolism for heterologous pathway productivity. ACS Synth Biol. 2013; 2:126–135. [PubMed: 23656436]
46. Juminaga D, et al. Modular engineering of L-tyrosine production in *Escherichia coli*. Appl Environ Microbiol. 2012; 78:89–98. [PubMed: 22020510]
47. Zhang K, Sawaya MR, Eisenberg DS, Liao JC. Expanding metabolism for biosynthesis of nonnatural alcohols. Proc Natl Acad Sci USA. 2008; 105:20653–20658. [PubMed: 19064911]
48. Leonard E, et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci USA. 2010; 107:13654–13659. [PubMed: 20643967]
49. Lee SM, Jellison T, Alper HS. Directed evolution of xylose isomerase for improved xylose catabolism and fermentation in the yeast *Saccharomyces cerevisiae*. Appl Environ Microbiol. 2012; 78:5708–5716. [PubMed: 22685138]
50. Young EM, Tong A, Bui H, Spofford C, Alper HS. Rewiring yeast sugar transporter preference through modifying a conserved protein motif. Proc Natl Acad Sci USA. 2014; 111:131–136. [PubMed: 24344268]
51. Bienick MS, et al. The interrelationship between promoter strength, gene expression, and growth rate. PLoS ONE. 2014; 9:e109105. [PubMed: 25286161]
52. Brewster RC, Jones DL, Phillips R. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. PLoS Comput Biol. 2012; 8:e1002811. [PubMed: 23271961]
53. Mutalik VK, et al. Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods. 2013; 10:347–353. [PubMed: 23474467]
54. Guimaraes JC, Rocha M, Arkin AP, Cambray G. D-Tailor: Automated analysis and design of DNA sequences. Bioinformatics. 2014; 30:1087–1094.
55. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci USA. 2013; 110:10039–10044. [PubMed: 23630264]
56. McAtee AG, Jazmin LJ, Young JD. Application of isotope labeling experiments and 13C flux analysis to enable rational pathway engineering. Curr Opin Biotechnol. 2015; 36:50–56. [PubMed: 26319894]
57. Young JD. 13C metabolic flux analysis of recombinant expression hosts. Curr Opin Biotechnol. 2014; 30:238–245. [PubMed: 25456032]
58. Feng X, Zhao H. Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via 13C metabolic flux analysis. Microb Cell Fact. 2013; 12:114. [PubMed: 24245823]
59. Zelcbuch L, et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res. 2013; 41:e98. [PubMed: 23470993]
60. Shao Z, Zhao H, Zhao H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 2009; 37:e16. [PubMed: 19074487]
61. Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden Gate shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS ONE. 2009; 4:e5553. [PubMed: 19436741]
62. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–345. [PubMed: 19363495]

63. Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat Methods. 2007; 4:251–256. [PubMed: 17293868]
64. Mitchell LA, et al. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. Nucleic Acids Res. 2015
65. Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res. 2012; 40:e142. [PubMed: 22718979]
66. Wang BL, et al. Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. Nat Biotechnol. 2014; 32:473–478. [PubMed: 24705516]
67. Michener JK, Smolke CD. High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. Metab Eng. 2012; 14:306–316. [PubMed: 22554528]
68. Tang SY, et al. Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. J Am Chem Soc. 2013; 135:10099–10103. [PubMed: 23786422]
69. Jha RK, Kern TL, Fox DT, Strauss CEM. Engineering an Acinetobacter regulon for biosensing and high-throughput enzyme screening in *E. coli* via flow cytometry. Nucleic Acids Res. 2014; 42:8150–8160. [PubMed: 24861620]
70. Kushwaha M, Salis HM. A portable expression resource for engineering cross-species genetic circuits and pathways. Nat Commun. 2015; 6 Article No.: 7832.
71. Ng CY, Farasat I, Maranas CD, Salis HM. Rational design of a synthetic Entner– Doudoroff pathway for improved and controllable NADPH regeneration. Metab Eng. 2015; 29:86–96. [PubMed: 25769287]
72. Chubiz LM, Lee MC, Delaney NF, Marx CJ. FREQ-Seq: A rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. PLoS ONE. 2012; 7:e47959. [PubMed: 23118913]
73. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. Nat Methods. 2015; 12:203–206. [PubMed: 25559584]
74. Smith AM, et al. Quantitative phenotyping via deep barcode sequencing. Genome Res. 2009; 19:1836–1842. [PubMed: 19622793]
75. Levy SF, et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015; 519:181–186. [PubMed: 25731169]
76. Walkiewicz K, et al. Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. Proc Natl Acad Sci USA. 2012; 109:21408–21413. [PubMed: 23236139]
77. Noderer WL, et al. Quantitative analysis of mammalian translation initiation sites by FACS-seq. Mol Syst Biol. 2014; 10:748. [PubMed: 25170020]
78. Fowler DM, Fields S. Deep mutational scanning: A new style of protein science. Nat Meth. 2014; 11:801–807.
79. Whitehead TA, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol. 2012; 30:543–548. [PubMed: 22634563]
80. Klesmith JR, Bacik JP, Michalczyk R, Whitehead TA. Comprehensivesequence-flux mapping of a levoglucosan utilization pathway in *E. coli*. ACS Synth Biol. 2015; 4:1235–1243. [PubMed: 26369947]
81. Kowalsky CA, et al. High-resolution sequence-function mapping of full-length proteins. PLoS ONE. 2015; 10:e0118193. [PubMed: 25790064]
82. Stapleton, JA., et al. Haplotype-phased synthetic long reads from short-read sequencing. bioRxiv. 2015. doi: http://dx.doi.org/10.1101/022897
83. Hong L, et al. BAsE-Seq: A method for obtaining long viral haplotypes from short sequence reads. Genome Biol. 2014; 15:517. [PubMed: 25406369]
84. Gall S, Lynch MD, Sandoval NR, Gill RT. Parallel mapping of genotypes to phenotypes contributing to overall biological fitness. Metab Eng. 2008; 10:382–393. [PubMed: 18789396]

85. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol. 2010; 28:856–862. [PubMed: 20639866]

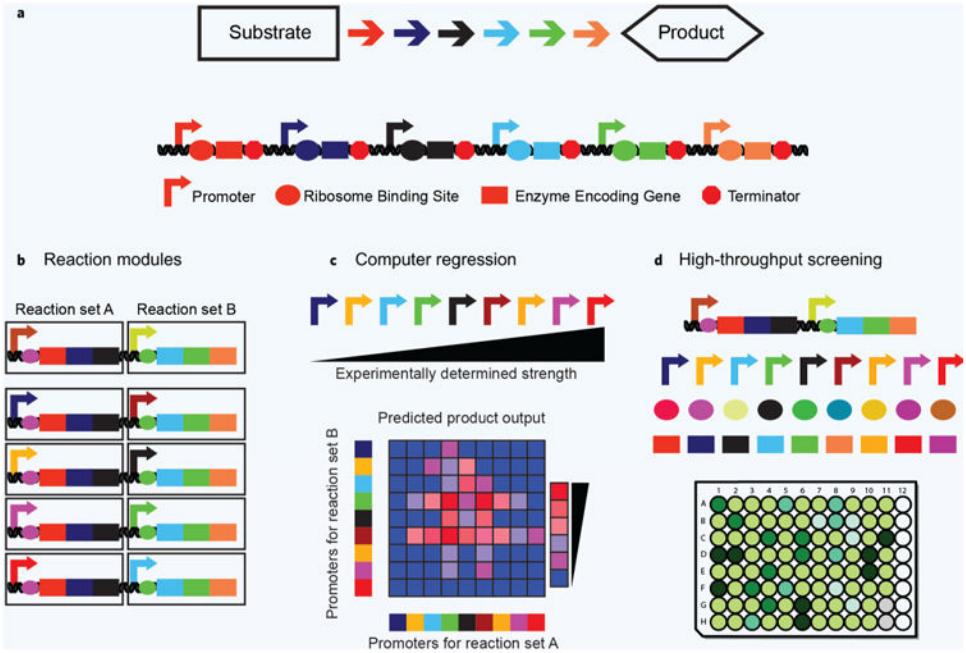


**Figure 1.**

There are three main strategies to sample sequence-flux space in metabolic pathways. (**a**) A model pathway. Each arrow represents a different reaction chemistry used to convert an initial substrate into a value-added product. Expression elements like promoters and ribosome binding sites (RBS) facilitate expression of the enzyme-encoding gene sequences. (**b**) Individual reaction chemistries are grouped into modules. Expression of these modules is varied, reducing the combinatorial search space. (**c**) Given a training set of different sequences with a given output, models can be harnessed to predict optimal pathway expression levels. (**d**) Advanced DNA assembly methods can be used to create unique pathway variants that are then assessed using high-throughput screens or selections.

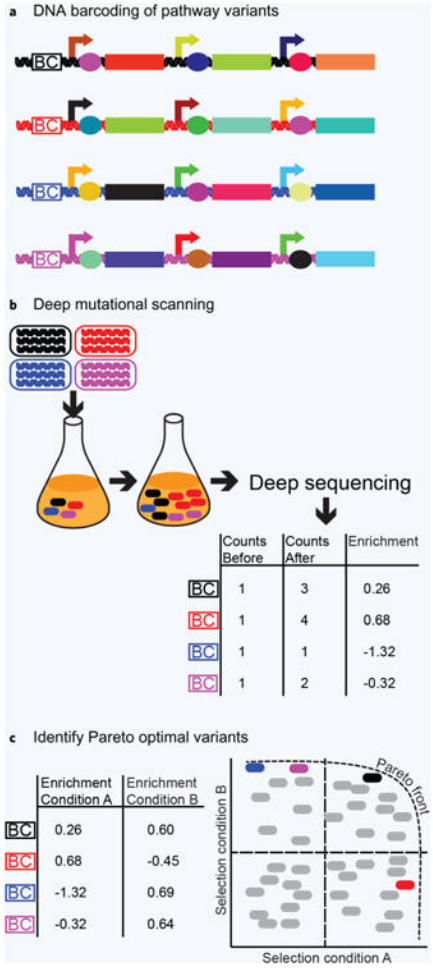


**Figure 2.**

Population-based measurements of pathways enable thorough search of flux space. (**a**) DNA barcoding methods allow long DNA constructs to be uniquely identified by short sequences. (**b**) Deep mutational scanning quantifies the enrichment of individual DNA variants after a selection. The enrichment relating the change in frequency of an individual variant can be related to specific productivity. (**c**) Comparing the fitness of individual variants between different selection conditions allows one to find Pareto optimal pathway sequences. This enables the identification of pathways supporting high specific productivity under robust processing conditions.