Modular cell design for rapid, efficient strain engineering toward industrialization of biology
Cong T Trinh1,2,3 and Brian Mendoza1,3

Transforming biology into an engineering practice has great potential to shape the industrialization of biology that will drive rapid development of novel microbial manufacturing platforms. These platforms will be capable of producing a vast number of sustainable industrial chemicals at scale from alternative renewable feedstocks or wastes (e.g., biomass residues, biogas methane, syngas, CO2) without harming the environment. The challenge is to develop microbial platforms to produce targeted chemicals with high efficiency in a rapid, predictable, and reproducible fashion. This paper highlights recent progress in rational design of heterologous pathways for combinatorial biosynthesis of a large space of chemicals and modular cell design for rapid strain engineering.

Addresses
1 Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, United States
2 Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, United States
3 Bioenergy Science Center (BESC), Oak Ridge National Laboratory, Oak Ridge, United States

Corresponding author: Trinh, Cong T (ctrinh@utk.edu)

Introduction
Underlying the rapid advancement of technological innovation during the 20th century was the harnessing of petroleum for its diverse library of potential products. Recent research and socioeconomic developments have revealed the dangers of relying on fossil fuels as a singular source of specialty fuels and chemicals [1]. Chemicals derived from engineered microorganisms have been lauded as a promising sustainable alternative to the thousands of chemicals derived from petroleum [2]. While some success stories of biosynthesized molecules achieved commercialization in recent years [3], the complexity of constructing these microbial cell factories is surpassed by the vastness of the biochemical space open to exploration [4]. For biosynthesized products to compete with current solutions economically and in a variety of applications, research efforts must endeavor to achieve an understanding of the combinatorial space available, and devise an effective means for rapid strain engineering to produce these products based on nature’s inherent modularity and synergies.

In this paper, we explore the progress made in rational design of heterologous pathways and highlight the importance of engineering dynamic control of heterologous pathways coupled with the host cell metabolism to achieve improved pathway efficiency. We envision the development of modular cell design principles will enable rapid strain engineering for combinatorial biosynthesis of a large, sustainable chemical space in a plug-and-play fashion requiring minimum strain optimization cycles.

Probing the combinatorial space of biobased chemicals
Cellular metabolisms are diverse and complex, generating thousands of unique chemicals. Advancements in comparative genomics, systems and synthetic biology, and metabolic engineering have enabled researchers to access a multitude of biological parts to assemble heterologous pathways and start probing a combinatorial space of biobased chemicals (Figure 1). Characterized biological parts have been compiled into databases such as the Synthetic Biology Parts Registry (http://parts.igem.org/Main_Page), KEGG [5], and Biocyc [6] among others that continue to expand as new genomes are discovered. While these databases continue to add to the breadth of biological parts information, the minimal depth of quantitative knowledge (e.g., transcription rates, enzyme kinetics, etc.) is limited, and will be the next challenge to address as the databases evolve to encompass a greater understanding of parts and their interactions with cellular systems.

Recent achievements in harnessing biological parts for chemical biosynthesis include: (i) rewiring central metabolism for making non-natural bioplastics from sustainable feedstocks [7], (ii) redesigning fermentative pathways for combinatorial biosynthesis of unique esters with tunable carbon backbones [8,9], (iii) harnessing synthetic pathways of reverse beta-oxidation and non-decarboxylative Claisen condensation coupled with subsequent beta-reduction reactions for combinatorial biosynthesis of alcohols, dicarboxylic acids, hydroxyl acids, and lactones [10,11**], (iv) manipulating polyketide and isopenoid
Figure 1

(a) **Opportunities**

**Tools/Techniques**
- *In silico*
  - COBRA
  - BNICE
  - Vector NTI, GeneDesign, Gene Composer
- *In vitro*
  - Golden Gate Assembly
  - Gibson Assembly
  - Overlapping PCR
- *In vivo*
  - Stable plasmids
  - Homologous recombination
  - CRISPR/Cas
  - DNA transformation techniques

**Parts library**
- Promoters
- RBS
- Protein orthology
- Biosensors
- Regulators
- RNAs

**Feedstocks**
- Lignocellulosic biomass
  - C1 feedstocks (methane, CO2)
  - Syngas
  - Organic wastes

**Host organisms**
- Model organisms (e.g., *Saccharomyces cerevisiae*, *Escherichia coli*, *Bacillus subtilis*)
- Non-model organisms (e.g., *Clostridium thermocellum*, *thermoanaerobacter saccharolyticum*, etc.)

**Current Products**
- Primary Metabolites: 10^19
- Secondary Metabolites: 10^15
- Isomeric and Heterologous Combinations: 10^11

**Biochemical Space**
- Unexplored Biological Possibilities: >10^25

**Future Directions**
- Specialty chemicals including:
  - FAMEs/FAEEs
  - Pesticides
  - Antibiotics
  - Fuel Additives
  - Specialty polymers precursors
  - Food additives
  - Flavors and fragrances
  - CAI cleaning solvents
  - Adhesives
  - Lubricants
- Commodity chemicals including:
  - Synthetic palm oil
  - Aviation fuel blends
  - and more...

(b) **Challenges**

(i) Develop reusable genetic tools and parts to be compatible across species.
(ii) Exploit non-model organisms with novel phenotypes (e.g., assimilation of C1 substrates, lignocellulosic biomass degradation).
(iii) Rapidly develop modular microbial platform to synthesize biochemical space.

The power (a) and the daunting challenge (b) for microbial synthesis of the potential combinatorial space of specialty fuels and chemicals.

biosynthesis pathways to produce alcohol fuels [12], (v) engineering fatty acid biosynthesis for producing hydrocarbons [13–16], and (vi) rerouting amino acid biosynthesis pathways for making alcohols, drug precursors, and industrial chemicals [17,18]. These chemicals have broad applications related to health, energy, and the environment, but are sometimes difficult to synthesize by a conventional chemical method.

**Rational design of heterologous pathways**

Engineering heterologous pathways in a recombinant host can become very challenging especially as the pathway complexity increases as seen in the production of opioids in *Saccharomyces cerevisiae* [19**]. For assembling multiple parts into a heterologous pathway, it is critical to balance and optimize fluxes through not only the heterologous pathway to produce a desirable chemical but also the host’s native pathways to maintain good cell viability [20–22]. The challenge is how one can identify an efficient heterologous pathway and choose a proper combination and assembly of parts for the pathway to create optimal phenotypes (e.g., efficient production of desirable chemicals at high yields, titers, and productivities) without going through extensive screening. For instance, designing an optimal multi-gene pathway in a bacterial host can generate a vast space of solutions that depend on finding the best pairing of appropriate promoters, ribosome binding sites, terminators, gene orders, tunable intergenic regions, and orthologous genes. In recent years, a collection of computation-based techniques has been developed to assist heterologous pathway design and can be classified into three groups: pathway prediction, yield analysis, and parts identification.

**Pathway prediction** is an essential tool to identify all thermodynamically feasible routes and associated enzymes to produce desirable chemicals from the existing databases. Hatzimanikatis and coworkers first developed
the BNICE (Biochemical Network Integrated Computational Explorer) framework to identify novel heterologous pathways [23]. Currently, there are a variety of available tools for pathway prediction [24] with improved search algorithms to address the computational challenge in identifying heterologous pathways from large, putative biochemical reaction databases [25]. For instance, by employing both computational metabolic pathway search and analysis from KEGG, Zhang et al. interrogated the metabolic potential of *Escherichia coli* as a microbial platform capable of producing 1777 non-native chemicals, 279 of which have commercial use. Interestingly, more than 50% of these commercial products require a minimum of three heterologous reaction steps [26].

**Yield analysis.** Once a desirable heterologous pathway is formulated, the next important step is to evaluate the pathway efficiency. Yield analysis can be employed to evaluate a pathway’s potential [27] by considering thermodynamics, electron and carbon constraints for both the engineered heterologous pathway, and native metabolism of the host. Constraint-based metabolic network modeling such as flux balance or elementary mode analysis offers powerful tools for performing yield analysis with a large collection of useful software [28]. Recently, there has been a significant interest in engineering carbon-conserving pathways in the native host as they not only can potentially improve theoretical pathway yields but also reduce the CO₂ carbon footprint harming the environment. For instance, a native carbon-conserving pathway is the succinic acid-producing pathway that could yield 1 Cmol succinate/Cmol glucose equivalent from sugar fermentation. In contrast, the synthesis of biofuels (e.g., ethanol and butanol) can only yield up to 0.67 Cmol product/Cmol glucose equivalent where 33% of carbon is lost to CO₂ as waste. Recent discovery of the non-oxidative glycolysis (NOG) pathway opens new opportunities to potentially engineer heterologous pathways to produce chemicals with product yields greater than the theoretical limits [29,30]. Other unique pathways like NOG likely exist, waiting to be discovered to build alternative novel carbon-conserving pathways.

**Parts identification.** The next critical step in heterologous pathway design is to identify parts that can be assembled to achieve optimal phenotypes. Since these pathway components are often synthetic and heterologous, it becomes very challenging to select compatible parts that display desirable phenotypes without going through iterative optimization. Currently, parts selection is mostly trial-and-error and relies on high-throughput screening to generate feedback for the design-build-test-learn cycles of pathway engineering. Toward addressing this challenge, Farasat et al. has recently developed the SEAMAP framework to design heterologous pathways with balanced and enhanced metabolic fluxes by manipulating ribosome binding sites to fine-tune translation rates and narrow the experimental strain engineering space [31].

Once the heterologous pathway is designed with appropriate parts, a variety of tools and techniques exist for streamlined synthesis and assembly. One class of parts assembly techniques is based on homologous recombination including a large collection of in vitro [32–39] and in vivo [40–42] assembly methods. In parallel, the other popular class of parts assembly techniques is based on non-homologous recombination such as Golden Gate assembly [43] among many other options [44–49]. Combined, these techniques enable the creation of combinatorial libraries and large DNA fragments (>1 Mb) for constructing minimal cells [50]. Computational tools (Vector NTI [51], j5 [52], Gene Designer [53], GeneDesign [54], and Gene Composer [55]) make the assembly process very seamless nowadays.

**Pathway evaluation for design-build-test-learn cycles**

Ideally, the engineered heterologous pathways work consistently with design to achieve desired products at high yields, titers, and productivities. However, the performance is often suboptimal in practice due to metabolic flux imbalances caused by incompatibility of the heterologous pathways and the host (e.g., inefficient protein expression, enzyme stability, metabolic burden, and redox imbalances). Metabolic flux analysis coupled with OMICS data is a state-of-the-art technique to identify metabolic bottlenecks [56–58]. Once the bottlenecks are identified, the control of these pathways together with their interaction with the regulatory machinery of the host is essential to obtain viable yields through the design-build-test-learn cycles. Next, we highlight significant progress in balancing and optimizing metabolic fluxes through static and dynamic controls at both pathway and cell population levels.

**Balancing Act: Steady-state control.** To overcome reaction bottlenecks, various techniques can be employed to modulate reaction fluxes via environmental (temperature, pH, substrates) and genetic (transcription, translation) manipulation to adjust metabolite concentrations as well as enzyme stability, concentrations, activities, and localization [20]. Similar to modularity implemented in the realm of industrial design, the complexity of pathways can be modularized for efficient flux control leading to modular production of related chemicals [8,9,59–62]. Metabolic control theory has been developed, and should be of great value for guiding modular pathway engineering in the future [63,64].

**Juggling Act: Dynamic pathway control.** Fluctuations in the cellular environment, whether planned or inadvertent, can create metabolic imbalances in biological systems optimized for very particular conditions. Natural biological
systems have evolved with efficient sensor and regulator systems (e.g., the well-known lac operon of E. coli) so that they can effectively respond to environmental cues to maximize their fitness. Since heterologous pathways are often incompatible to the host, dynamic controllers engineered in these pathways would be highly advantageous to maximize product formation while maintaining healthy cells. Dynamic pathway control is the forefront of tool development, going beyond steady-state balance. Next, we highlight some recent advancement in harnessing biosensors and genetic circuit tools for dynamic pathway control.

To engineer dynamic control of any heterologous pathway, the critical element is to have controllable biosensors to construct the sensor-regulator system. Biosensors can be classified into many types including metabolic response transcription factors, two-component systems, regulatory RNAs, and protein allostery [65]. At pathway levels, these sensor-regulator systems are hard-wired into heterologous pathways for sensing intracellular metabolites and/or environmental cues (e.g., temperature, pH, light) to control metabolic fluxes via feedback and/or feedforward mechanisms [66,67]. At cell levels, they can be used to decouple growth and production phases by implementing transcriptional toggle switches [68,69]. A CRISPRi/a (interference and activation) system with RNA scaffolding, or a genetic switchboard using riboswitches, prove to be powerful tools to dynamically toggle between heterologous pathways [70,71]. Population quality control (PopQC) is also a useful genetic circuit to enforce the optimal performance of the heterologous pathway [72]. Because of growth competition fitness, heterogeneity in a cell population carrying heterologous pathways often causes low product production. To address this problem, Xiao et al. developed the PopQC strategy by dynamically coupling heterologous pathway flux with antibiotic selection to select for the most hyper-producing strain in the population during the chemical production phase. The team demonstrated a 3-fold improvement in production of fatty acids and tyrosine. Since specificity toward target metabolites is critical for engineered controllable biosensors of heterologous pathways, Taylor et al. took a de novo protein design approach to reengineer an allosteric transcription factor (αTF) LacI of E. coli to sense alternative ligands [73]. This approach is powerful to design new properties of abundant, native αTFs for dynamic pathway control.

**Genome and combinatorial engineering.** With the ability to create a large number of variants no longer a limiting step, focus is necessarily shifting toward improved methods of genome and combinatorial engineering. Instead of characterizing individual cells carrying pathway variants one by one, the entire population can be probed simultaneously. The approach is very powerful if one can screen or select for the desirable phenotype, that is, high-yield production of target chemicals. For instance, if a heterologous pathway produces a chemical that can emit light, chemical-hyperproducing strains in the cell population can be isolated via fluorescence-activated cell sorting (FACS) or solid plate screening [74,75]. If production of a chemical is coupled with growth, chemical-hyperproducing strains can be isolated based on simple growth selection. Deep sequencing can also be employed for rapid strain isolation [76]. Both TRMR [77] and TRACE [78] are powerful genome engineering tools to identify potential targets for manipulation; when coupled with MAGE [74] outperforming strains can be generated. While TRMR and MAGE rely solely on the homologous recombination machinery of the host for effective genome editing, recent advances in the CRISPR technology streamlines the genome editing process for rapid strain engineering by utilizing double strand breaks (DSB) and single strand nicks to improve homologous recombination efficiency and perform multiplexing [79]. Even though these methods are powerful to improve engineered phenotypes, they do not often generate the most optimal phenotypes due to incomplete sampling spaces and unforeseen native regulation.

**Challenges in heterologous pathway engineering**

One significant challenge in engineering heterologous pathways is to deal with non-model organisms because availability of genetic tools and parts libraries does not exist. Parts incompatibility is very common because cellular machinery and its regulation can vary greatly between different microbes. For instance, Clostridium thermocellum is one of the potential consolidated bioprocessing thermophiles that is very efficient and robust in degrading complex plant biomass [80]. While many attempts have been explored, no engineered E. coli or S. cerevisiae mutants reported to date could degrade biomass as efficiently as native C. thermocellum. It is of great interest to introduce heterologous pathways into C. thermocellum and rewire its metabolism for production of specialty biofuels and chemicals while exploiting its biomass-degrading machinery; however, reliable genetic tools and parts availability remain formidable bottlenecks. This challenge presents itself for each new organism and hence breaking these barriers will help industrialization of biology to exploit nature’s best. Advances in the CRISPR technology shows potential for manipulating non-model organisms for metabolic engineering applications [81,82].

One other significant challenge in engineering heterologous pathways is the host cell must be re-engineered in an iterative manner to produce different chemicals. This process is laborious and expensive (Figure 2a). Ideally, it is advantageous to develop a blueprint of the universal modular cell that, when combined with optimized exchangeable production modules, creates microbial
manufacturing platforms in a plug-and-play fashion for optimal production of desirable chemicals (Figure 2b).

**Modular cell design toward industrialization of biology**

Complexity has kept metabolic engineering confined to time-consuming validation and optimization via multiple design-build-test-learn cycles, despite best efforts to standardize and characterize parts [83]. The grand challenge is how to streamline pathway and strain engineering to rapidly explore the combinatory space of chemicals (Figure 1). While the high-throughput screening approach is powerful, it is not advantageous and proportionally scaled to deal with the large chemical space; a need for rational modular chassis design is required.

Trinh et al. has laid out the computational framework named MODCELL (modular cell) for designing modular cells that couple with a diverse class of production modules (i.e., heterologous pathways) [84**]. The MODCELL design principles are formulated such that the modular cell must be auxotrophic and contain the core metabolic pathways that are necessary but insufficient to support cell growth and maintenance under controllable physiological conditions (e.g., anaerobic conditions). The modular cell is designed to be auxotrophic by imposing cofactor imbalance and/or insufficient supply of precursor metabolites required for biosynthesis of biomass and targeted chemicals. To efficiently produce targeted chemicals, the modular cell must be tightly coupled with exchangeable production modules, auxiliary metabolic pathways designed to synthesize target chemicals. The tighter the coupling between the modular cell and production module, the faster the cell growth, substrate consumption, and desirable chemical production rate become. Here, the modularity of the design will enable rapid development of microbial platforms from the modular cell and exchangeable production modules in a plug-and-play fashion, whereas the metabolic coupling will provide powerful selection for production of targeted chemicals at high efficiency during both growth and non-growth associated phases. The MODCELL has shown promise through the demonstrated production of alcohol fuels [85,86,87] and combinatorial biosynthesis of esters [9**].

Modular cells can be constructed from a bottom-up approach where a minimal cell can be synthetically designed and constructed as a host [88–95]. Alternatively, native metabolism of the existing strains must be restricted based on the MODCELL design [84**]. The metabolic coupling
design in modular cells is an ideal chassis for dynamic pathway engineering as well as genome and combinatorial engineering. One could envision the modular cell being constructed with the production module(s) integrated into the host chromosome and transcriptionally controlled by an environmental signal (e.g., light, temperature, IPTG, nutrients, and/or a desirable chemical) in a controllable genetic circuit. This design can impose the auxotrophic characteristic of the modular cell and trigger cell growth and chemical overproduction only if environmental signals are received to activate the production module(s). An additional advantage is the MODCELL design can provide an alternative, secure strategy for microbial containment due to the auxotrophic characteristic of modular cells [96].

Conclusions

Developments in synthetic biology, as well as rapid discovery of novel organisms and their genes, enzymes, and regulatory mechanisms continually grow, enabling metabolic engineers to explore the boundless chemical space. Computational models and curated databases of biological parts play vital roles in driving systematic pathway design, and continued effort into standardizing and improving these sources of information is imperative. Research into building dynamic sensor-regulator devices is still in its infancy, and must be developed if we are to succeed in producing targeted chemicals at high yields, titers, and productivities. The need to develop modular chassis cells for tight coupling with optimized heterologous pathways can potentially minimize the iterative design-build-test-learn cycles. Further, high-throughput technologies will enable rapid selection of the best phenotype amongst the set of the rationally designed pathways. It is of our opinion that achieving the goals of these research areas will lead to the successful engineering of economical microbial factories for efficient production of specialty fuels and chemicals, replacing traditional petroleum-based products.

Acknowledgements

This research was financially supported in part by the NSF awards (NSF #1553250, NSF #1360867, NSF #1553250) as well as the DOE subcontract grant (DE-AC05-00OR22725) by the BioEnergy Science Center (BESC), the U.S. Department of Energy Bioenergy Research Center funded by the Office of Biological and Environmental Research in the DOE Office of Science.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


9. Layton DS, Trinh CT: Engineering modular ester fermentative pathways in Escherichia coli. Metab Eng 2014, 26:77-88. This study established a general platform for a modular fermentative ester biosynthesis.


This study successfully engineered the most complex heterologous pathway known to date for the opioid biosynthesis in S. cerevisiae. This pathway employed an expression of a set of 21 thebaine and 25 hydrocodone enzymes derived from plants, mammals, bacteria, and yeast itself.


This work presents an excellent overview on pathway and strain engineering design tools.


This study employed a computational biology approach to systematically analyze the metabolic feasibility of E. coli as a host for production of a large space of non-native chemicals.


This work presents the SEAMAP framework for designing heterologous pathways with optimized fluxes by manipulating ribosome binding sites.


This work presents the construction of a synthetic minimal cell with large DNA synthesis and assembly.


This heterologous hyper-producing approach allows for the directed evolution of metabolic pathways.

- **This study presents a comprehensive overview of different classes of biosensors that can be harnessed to build synthetic sensor-regulatory systems for engineering heterologous pathways.**
- **This paper established a multilayered activation/inhibition system in S. cerevisiae using a Cas9 master switch for controlled production of a family of violaceins.**
- **This paper developed a PopQC genetic circuit to dynamically select hyper-producing cells during chemical production phase by coupling the heterologous pathway with antibiotics selection.**