INTRODUCTION

The microbial generation of hydrocarbons, either from inexpensive plant biomass or directly from sunlight, could provide a renewable and carbon-neutral source of liquid transportation fuels, so long as the biomass is grown and harvested sustainably.1,2 The explosive advances in synthetic biology and metabolic engineering have dramatically accelerated efforts to engineer microbes for fuel production (Table 11.1). Thanks to an ever-decreasing cost of DNA synthesis and new, robust tools for DNA manipulation, our abilities to install whatever DNA sequence we can imagine into a widening range of host microbes are more powerful than ever, even up to the scale of installing whole genomes into cells.3 The sheer number of enzymes available in gene databases, and the incredible breadth of chemical transformations they catalyze, represents vast biosynthetic versatility that can be harnessed in simple cloning steps. These capabilities are already being applied to engineer microbial production of valuable chemicals, such as pharmaceuticals, fragrances, and vitamins. However, the current low cost of petrochemicals is a tremendous barrier to an economy-wide adoption of microbially produced bulk compounds such as fuels and plastics. Biologically generated commodity compounds will only be competitive with petrochemicals when inexpensive bioprocesses are finally developed. These processes will feature robust microbes capable of high product titers at near-theoretical yields. Achieving this ambitious goal will likely require every trick known to synthetic biology.

Synthetic biologists have already taken many of the steps towards this goal. Microbial production of biofuels with combustion properties similar to existing fuels has been repeatedly demonstrated at proof-of-concept levels.4–7 However, we still cannot underestimate the challenge of genetically rewiring the physiology and metabolism of microbes to reliably generate chemicals at a commodity scale. Our understanding of how synthetic DNA sequences will behave once they are inside a cell still severely lags behind...
our DNA synthesis abilities. In addition, the behavior of highly engineered microbes must be robust in the scale-up from shake-flask cultures to thousand-liter tanks. If we are to achieve the high titers required of biofuel production beyond the milligram per liter scales commonly demonstrated in laboratory production, and develop a process that whole economies can rely upon, our capabilities for biological design still need much improvement.

We describe here a selection of the methods available for engineering gene expression, enzyme function, and host cell physiology, many of which have been shown to be effective in improving the yield of a biologically produced compound (see Fig. 11.1 for an overview).

This chapter is organized into two parts, beginning with ‘Pathway design and optimization.’ We address issues and techniques for engineering the biofuel production pathway: the collection of enzymes that catalyze the chemical transformations that turn metabolites into fuels. The second part of the chapter describes issues and techniques for engineering the host organism responsible for expressing, feeding, and sustaining the biofuel production pathway.

### TABLE 11.1 Selected Examples of Synthetic Biology Principles for Biofuel and Biomaterial Production

<table>
<thead>
<tr>
<th>Mechanism of Control</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DNA</strong></td>
<td></td>
</tr>
<tr>
<td>Copy number</td>
<td>47, 41, 57, 49</td>
</tr>
<tr>
<td><strong>Transcription</strong></td>
<td></td>
</tr>
<tr>
<td>Promoter strength</td>
<td>51, 57, 53</td>
</tr>
<tr>
<td>Promoter timing</td>
<td>54</td>
</tr>
<tr>
<td>Operon organization</td>
<td>54a</td>
</tr>
<tr>
<td><strong>mRNA</strong></td>
<td></td>
</tr>
<tr>
<td>RNA half-life</td>
<td>62</td>
</tr>
<tr>
<td>Secondary structures</td>
<td>62</td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td></td>
</tr>
<tr>
<td>RBS strength</td>
<td>59</td>
</tr>
<tr>
<td>Codon usage</td>
<td>60</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td></td>
</tr>
<tr>
<td>Protein half-life</td>
<td>15</td>
</tr>
<tr>
<td>Protein optimization</td>
<td>7, 16, 34</td>
</tr>
<tr>
<td><strong>Metabolites and Products</strong></td>
<td>15</td>
</tr>
<tr>
<td>Biosensors</td>
<td>54, 54a</td>
</tr>
<tr>
<td>Allosteric regulation</td>
<td>29, 32</td>
</tr>
</tbody>
</table>
PATHWAY DESIGN AND OPTIMIZATION

Initial Pathway Design and Validation

The first step in the design of any biofuel production pathway is deciding which fuel compound will be produced using the pool of metabolic precursors available in the host organism. Many potential fuel compounds or their synthetic intermediates occur naturally, and their biosynthetic pathways have been determined. The suitability of biological compounds for use as fuel replacements is reviewed elsewhere. The next challenge is to assemble a biochemical pathway capable of synthesizing the target molecule using known enzymatic activities. The genes that encode the necessary enzymes can often be drawn from the vast library of publicly available sequences. Pathways to produce potentially any molecule can be assembled from the rich diversity of enzymatic activities that have evolved over billions of years. Entire biosynthetic pathways can be used as they are found in nature, or can be assembled piecemeal by incorporating enzymes from different sources. This diversity can be expanded yet further by engineering existing enzymes to generate nonnatural products. Biosynthetic pathways can also be constructed from catabolic pathways engineered to run in the reverse direction than their usual physiological role. This has been demonstrated using a reversed fatty-acid degradation pathway to produce butanol in high titers in *E. coli*. Construction of a biofuel pathway typically begins with selection of genes encoding enzymes that comprise the individual pathway components. The full pathway is assembled in one or more plasmids and expressed in the host organism of choice, which is often a well-characterized and genetically tractable organism such as *Escherichia coli* or *Saccharomyces cerevisiae*. Other organisms with different metabolic capabilities and toxicity tolerances that may be better suited to producing the desired compound can certainly be used as hosts, although they often present other challenges, most notably having fewer techniques for genetic manipulations. Production of the desired compound is usually assayed using an appropriate chromatographic method. As it is unlikely that the initial design of the pathway will produce high titer levels, production will be orders-of-magnitude lower than is needed in an industrial process. Nonetheless, an important milestone will have been achieved: proof that biological production of a fuel compound within a new host is possible. The brevity with which the initial stage of pathway design and validation is treated here is not intended to indicate its triviality or simplicity. It is a necessary first step, but an equally important challenge lies in increasing titers and yields closer to commercially viable levels.
Optimization of the Biofuel Pathway

The goal of pathway optimization is to maximize the efficiency of the process that synthesizes the desired fuel from existing metabolites in the host, which is determined by concentrations of precursors, intermediates, cofactors, and enzymes, as well as the catalytic activities of the enzymes, which are almost always subject to regulation on multiple levels, from gene expression to enzyme activity. This section will begin by addressing the optimization of the fundamental pathway components: the enzymes themselves.

PROTEIN ENGINEERING TO IMPROVE FLUX

If an enzyme is found to be limiting biofuel production, one straightforward approach to improve production is to screen a library of homologues to find a replacement for the bottleneck enzyme. This approach relies upon the sheer numbers of homologues often available within genetic databases that presumably span a wide range of properties, such as stability and catalytic rates. Often, an improved enzyme can be found, as has recently been demonstrated in work that achieved a high level of production of the biodiesel replacement bisabolene. A library of five bisabolene synthases from four plant species were cloned and screened for production titers, generating yields that varied by 100-fold, and one high-yielding (0.5 g/L) enzyme was discovered. However, homologues may not always be available or numerous (especially for certain classes of terpene cyclases), and high-yielding enzymes may not always be happened upon so fortuitously, especially for enzymes that have not evolved to produce high yields, as is often the case for secondary metabolites. A recent review argues that it is better to optimize production by adjusting the copy number, regulation, and enzymatic properties of the components of a pathway, rather than screening libraries of enzymes. While screening a library of enzymes might provide a coarse-grained search over a productivity landscape, protein and metabolic engineering methods are likely necessary to bring the pathway closer to a productivity maximum.

The lack of biochemical characterization of the vast majority of known enzymes is a natural consequence of the ease with which prospective genes can be found in sequence databases and cloned, compared to the difficulties and uncertainties inherent in protein overexpression and biochemical measurements. This is unfortunate, as knowledge of the properties of each enzyme within a pathway can better focus pathway optimization efforts. For instance, attempting to increase flux by increasing the copy number of a pathway component will be ineffective if the enzyme is nonfunctional (not expressed, or lacking necessary post-translational modifications), poorly functional (not folding properly), or functional but permitting a lower flux than is optimal for production of the biofuel (inhibition by a competitor or allosteric regulator). Of course, the enzymes selected for a pathway must be compatible with the host, a property referred to as ‘functional composibility.’ An enzyme is more likely to function in the production host if the organism from which it is obtained shares a similar intracellular chemical environment with the new host (such as pH or growth temperature), though the sensitivity of enzymes to these factors likely varies widely, and often compatibility cannot be known until expression is attempted. In particular, genes transplanted between kingdoms may suffer compatibility problems. These issues may be overcome by engineering the protein itself to improve its ‘in vivo’ properties, a term that refers collectively to the solubility, stability, selectivity, and activity of an enzyme.

All enzymes need to fold into their native three-dimensional structure before they can function. Proteins often misfold when expressed in a foreign host (or even when overexpressed in the original host). When a pathway enzyme does not fold, it can either be replaced or modified to improve its folding. Yoshikuni et al. found that the enzyme γ-humulene synthase (HUM), a sesquiterpene synthase from the gymnosperm Abies grandis, demonstrated low activity when transplanted into E. coli (15). Investigations found that very
little of the expressed HUM (~15%) resided in the soluble fraction of the cell, suggesting that much of the enzyme was misfolding or was insufficiently solubilized. An adaptive evolution approach that involved replacing amino acid residues in HUM with glycine and proline residues improved the solubility of the enzyme (~60% in the soluble fraction), resulting in a 220-fold increased production of sesquiterpines. A biochemical analysis of wild-type and mutant HUM enzymes revealed that the catalytic properties ($k_{cat}$ or $K_m$) of the mutant enzyme remained the same, implying that the production improvement was achieved by increasing concentration of folded, active enzyme. Other methods adapted from the protein overexpression field may help to solubilize an unfolded enzyme,\(^{16,17}\) and thus help improve final product titer by increasing the flux it catalyzes.

Many enzymes require post-translation modifications for activity, e.g. phosphopantethenylation.\(^{18-20}\) Often the proteins responsible for catalyzing these modifications need to be transplanted from their host organism, along with the pathway enzyme, if the modification activity is absent from the host or incompatible with the pathway enzyme. Other accessory proteins whose activities are absent in the host cell may also need to be present, and their activity levels optimized, in order to demonstrate activity. For instance, a key intermediate step in the microbial production of 1,3-propanediol using \textit{E. coli} is the dehydration of glycerol to 3-hydroxypropionaldehyde by a glycerol dehydratase gene \textit{dhaB1-3}.\(^{21}\) The enzyme can also be deactivated by glycerol, requiring reactivation by a glycerol dehydratase reactivate enzyme.\(^{22}\) An engineered strain capable of converting glucose to 1,3-propanediol on an industrial scale uses glycerol dehydratase and its reactivating factors from \textit{Klebsiella pneumonia}. Other examples of accessory proteins are the cytochrome P450 reductases (CPR) that are required to reduce the heme group in cytochrome P450s.\(^{23}\) Because commonly used platform organisms for biosynthesis lack CPR enzymes, CPR activity often needs to be imported with the P450 genes, and itself optimized to improve titer.\(^{14}\)

Flux through a pathway may also be modulated by feedback regulation of the enzymes by metabolites or other small molecules. Feedback regulation can strongly affect enzymatic activity through product or allosteric inhibition. This mode of regulation reacts to metabolite concentration far quicker than either translational or transcriptional control,\(^{24,25}\) and maintains the concentrations of intermediates within a narrow range to prevent osmotic stress or other toxicity resulting from over-accumulation of small molecules.\(^{26}\) Negative feedback regulation on an enzyme decreases its activity, effectively lowering the concentration of the enzyme. When production pathways contain allosterically regulated enzymes, these enzymes can become the bottleneck if the regulating metabolite reaches inhibitory levels. Thus, engineering feedback-resistant enzymes often enables higher production. Because our knowledge of allosteric inhibition and similar modes of regulation by metabolites has not kept pace with our vast catalogue of genes,\(^{27}\) it seems likely that careful biochemical measurements of pathway enzymes would reveal such feedback mechanisms.

Relieving feedback inhibition in the biosynthetic pathway of 2-keto acids,\(^{28}\) a production intermediate in the biosynthesis of amino acids, enabled workers in the Liao group to greatly increase their titers of alcohols. Work by Shen and Liao\(^ {29}\) found that production of threonine, a precursor to 2-ketobutyrate, was a bottleneck in their alcohol production pathway. The first two steps in threonine production from aspartate are catalyzed by the enzyme aspartate kinase/homoserine dehydrogenase (ThrA), which is inhibited by the downstream product threonine, an effect first discovered in 1976.\(^ {30}\) A feedback-resistant ThrA enzyme was cloned from a threonine-overproducing \textit{E. coli} strain and expressed under control of an inducible promoter. Expressing the feedback-resistant ThrA resulted in a three- to four-fold increase in the titers of both 1-propanol and 1-butanol.\(^ {29}\)
The abrogation of feedback inhibition can be achieved not only by mutations at inhibitor-binding allosteric sites, but also by simply relocating the enzyme to a different compartment within the cell. For instance, removing the leader peptide sequence of an *E. coli* thioesterase TesA prevents its export into the periplasmic space and confines it to the cytoplasm, where it depletes the pool of fatty acid-acyl carrier protein, a repressor of fatty acid biosynthesis. This resulted in accumulation of fatty acids. Overexpression of leaderless TesA in *E. coli* provided high levels of intracellular fatty acids and accelerated the production of fatty acid ethyl esters, a biodiesel.

The catalytic activity of the pathway enzymes may also be a limiting factor, as not all enzymes exhibit a fast turnover rate, particularly when they are members of a pathway that has not evolved to generate large amounts of its product, or if the enzyme is functionally promiscuous (generates multiple products). For instance, enzymes involved in terpene biosynthesis generate several products from a common precursor, a property that tends to correlate with having low catalytic activity. As a result, the flux through the pathway to the terpene products is often low. Leonard et al. screened active site mutants of the enzyme levopimaradiene synthase (LPS) from *Gingko biloba* to improve the product titer of diterpenoids produced by a factor of 10. A colorimetric screen of random mutants of the prenyl transferase enzyme GGPPS, which catalyzes the preceding step and provides LPS with a substrate, generated candidate GGPPS enzymes that produced higher yields of geranylgeranyl diphosphate (GGPP). By coupling the best-performing GPPSS mutant with the LPS mutant, the authors succeeded in increasing the diterpenoid yield by 18-fold, a result obtained using protein engineering alone.

**INCREASING FLUX BY TAILORING ENZYME LOCALIZATION OR COMPARTMENTALIZATION**

Pathway flux can also be increased without directly engineering higher catalytic activity. Enzymes that catalyze sequential reactions can be localized in close proximity to each other by means of artificial scaffolds, or by fusing enzymes that catalyze sequential reactions. This approach is inspired by examples found in native organisms of multiple enzyme activities being colocalized in one complex or organelle. Because the concentration of the product of an enzyme will be higher near its active site, placing an enzyme that reacts with that product nearby takes advantage of the higher local concentration to increase flux through the following reaction. Colocalization of pathway enzymes may be particularly useful when either the enzymes within a pathway cannot be optimized further, or an intermediate metabolite is toxic to the host cell, or is a substrate of a competing enzyme within the cell. This approach was validated using several enzyme components of the mevalonate pathway. The enzymes AtoB, HMGS, and HMGR were joined on an artificial scaffold coexpressed with the pathway enzymes, and the stoichiometry of the enzymes on the scaffold was adjusted by modifying the numbers of individual binding sites of each enzyme on the scaffold. The best scaffold resulted in a 77-fold increase in titer over the nonscaffolded pathway, while further enabling lower enzyme expression. Even simple protein fusions without a scaffold have demonstrated increased production. Fusing the enzymes FPP synthase and farnesene synthase resulted in a modest increase in farnesene production, compared to free enzymes, and eliminated formation of the side product farnesol.

Colocalization of proteins within a microcompartment, a strategy used by bacteria, is a promising avenue of research into improving pathway efficiency. Carboxysomes, a microcompartment found within cyanobacteria, increase the carbon fixation rate of RubisCo by colocalizing carbonic anhydrase enzymes, which produces CO$_2$ from bicarbonate, into a protein shell with RubisCO, which condenses the CO$_2$ with ribulose 1,5-bisphosphate to produce 3-phosphoglycerate. Because RubisCO has a low affinity for CO$_2$, the increased
concentration of bicarbonate inside the microcompartment enables a higher flux than RubisCO might otherwise achieve. Other compartments, such as the Pdu and Eut (1,2-propanediol and ethanolamine bacterial microcompartments, respectively) are thought to benefit their hosts by blocking the diffusional loss of toxic intermediates produced by the pathway into the cytoplasm. The protein shell of the microcompartment may also act as a regulation point by controlling the entry and efflux of precursors and products, enabling metabolite-level regulation on these pathways. The details of protein targeting into the microcompartments are becoming clear, and promising applications for harnessing microcompartments for biofuel production are as yet unexplored.

CENTRAL DOGMA ENGINEERING: HOW MANY ENZYMES AND AT WHAT TIME?

The ideal starting point in optimizing a biofuel production pathway is with biochemically characterized enzymes that are known to be capable of providing flux at a level needed for high production rates in vivo. The next step is to design a genetic system that expresses the optimal level of each enzyme at the correct time. Unfortunately, just as it is difficult to perform biochemical measurements of pathway enzymes, it is also extremely difficult to know a priori how much enzyme is needed in vivo for an optimum pathway, and to precisely engineer those protein levels. The use of in vitro assays, such as cell-free approaches that enable each pathway component to be titrated individually, can inform optimization. One goal of synthetic biology is to be able to engineer protein expression systems that robustly produce enzymes at a predetermined level, and at precise times. A simplistic viewpoint that approximates flux as a product of enzyme concentration, substrate concentration, and catalytic rate would assume that more enzymes would produce higher flux. Following this logic, the optimal solution for maximum flux might be to engineer the highest possible enzyme concentrations. Unfortunately, enzyme activities are often feedback-regulated by metabolites, and high concentrations of intermediates can be toxic to the cell. Furthermore, expressing a high level of pathway protein may be detrimental for other reasons, such as squandering transcription and translation machinery generating superfluous levels of protein. It has become clear that enzyme levels within a biofuel pathway must be carefully balanced. Metabolic control theory predicts that the best approach to maximizing productivity is to modify each component of a synthetic pathway in a way that increases fluxes but maintains intermediate concentrations near their physiological levels. Considered an extremely difficult undertaking when first articulated, advances in synthetic biology and protein engineering, which enable the careful tuning of enzyme levels and catalytic activities, may make this approach more tractable, if still challenging.

From Gene to mRNA: Modulating Transcription of Pathway Genes

The concentration of a protein can be expected to increase as the concentration of the mRNA transcripts that encode the protein increases, although this is not always a simple relationship (as reviewed in), and there is little to no correlation between mRNA levels and protein levels at the single-cell level. Fortunately for the bioengineer, factors that determine mRNA levels, such as DNA copy number and promoter strength, usually have predictable effects on protein expression, and can be readily modulated using straightforward genetic methods.

Typically when demonstrating biosynthesis of a fuel, production pathways are encoded on plasmids. Plasmid copy number can be easily modulated by using plasmids with various origins of replication that confer low- or high-copy numbers. This can considerably simplify efforts to manipulate gene copy number and resulting protein concentrations, which can identify enzymes that are expressed at insufficient levels. This approach was used to
determine the enzyme whose low concentration limited the productivity of the amorphadiene production pathway in *E. coli*. The complete pathway for amorphadiene production was encoded on a low-copy plasmid and transformed into *E. coli*. Higher-copy plasmids that encoded each individual pathway enzyme were subsequently introduced. A plasmid encoding the mevalonate kinase gene (MK) increased amorphadiene yields, suggesting that concentration of the MK enzyme limited production. This approach has also been used to improve production of the amorphadiene precursor mevalonate. Increasing the copy number of rate-limiting enzyme tHMGR relieved the growth inhibition caused by accumulation of a precursor metabolite, relieving toxicity and enabling higher mevalonate levels. Additional copies of a gene may be placed on the same plasmid as the rest of the pathway; however, this may lead to gene loss through recombination and is best avoided.

Plasmid-based expression of pathways may not be well-suited for high-titer production of biofuels. The metabolic requirements of plasmid replication impose on the host cell a metabolic burden, which can scale with plasmid copy number. Plasmids are autonomous genetic elements that may not be distributed evenly upon cell division. If expression of a plasmid-based biofuel pathway decreases cell growth rate or viability, individuals with fewer copies of the plasmid will rapidly out-compete the others, leading to a population with low overall productivity. Techniques that transfer the pathway genes to the chromosome itself avoid this problem; however, as the chromosome is present in very low copy numbers, this may place a very low limit on the pathway copy number that can be introduced. A recently developed innovation, chemically inducible chromosomal evolution (CIChE), enables the introduction of moderate-copy numbers of pathways onto the chromosome (~40). Stability of the repeated genes is ensured via knockout of the native recombinase gene *recA*.

Biofuel production genes are commonly expressed from well-characterized promoters induced by the addition of an exogenous chemical, such as anhydrous tetracycline or IPTG. This provides a simple means with which to modulate the strength of expression, as increasing inducer concentration should result in higher rates of transcription. While the use of promoters driven by exogenously added chemical inducers is very useful in the early stages of pathway demonstration, the price of typical inducers prohibits their use on an industrial scale. Constitutive promoters with a wide range of promoter strengths have been developed, enabling finely tuned enzyme levels, without the need for chemical inducers. Constitutively active promoters can enable an engineer to modulate the transcript levels of either the whole biofuel pathway, or each individual enzyme can be controlled by a specific constitutive promoter. However, constitutive expression of pathway enzymes may present a metabolic burden on the cell, presenting an opportunity for escape mutants to evolve and contribute to pathway instability.

A more sophisticated approach to controlling pathway expression is to use biosensors (see Box 11.1) to couple transcription of pathway enzymes to the concentration of specific metabolites or other cellular or environmental conditions. This takes advantage of control mechanisms evolved to dynamically respond to metabolite levels. An early demonstration of using biosensors to achieve dynamic metabolic control coupled transcription of key pathway genes of a lycopene biosynthesis pathway with the concentration of acetyl phosphate resulted in improved titers of the product in *E. coli*. The expression of the rate-controlling enzymes of lycopene synthesis was placed under control of a promoter native to *E. coli* that responds to acetyl phosphate, a metabolite that increases during periods of excess glycolytic flux. Thus, the lycopene pathway was activated when acetyl phosphate levels were high, which indicated an imbalance between carbon influx and carbon consumption. This increased the production of lycopene by 10-fold above what was achieved by
controlling expression of those genes using a chemically inducible promoter. Furthermore, growth retardation associated with overexpression of one of the rate-limiting proteins was not seen, despite a higher protein expression level. These improvements were attributed to a better coordination of lycopene biosynthesis with the metabolic state of the cell, resulting in better diversion of flux away from acetate into lycopene. Finally, because production is coupled to glycolytic excess, the extra carbon is channeled into lycopene rather than the toxic overflow product acetate, reducing the production of this waste by 66%.

A naturally occurring biosensor of an intermediate in the synthesis of the biodiesel fatty acid ethyl esters was recently used to drive expression of genes that further modify the intermediate into the final biofuel product. In *E. coli*, the transcription factor FadR responds to fatty acyl-CoAs and free fatty acids by relieving inhibition of downstream genes. The expression of an ethanol production pathway and a wax ester synthase gene responsible for esterifying the fatty acids with ethanol were placed under control of FadR. These genes were not transcribed until sufficient levels of fatty acyl-CoA had accumulated, preventing not only wasteful expression of downstream proteins before they had substrates to act upon, but also accumulation of the toxic ethanol to high levels. In addition, this type of regulation allows the cell to coordinate the levels of the reactants; the transcriptional output of the biosensor increases with increasing concentrations of acyl-CoA, which in turn drives faster production of ethanol and the wax ester synthase. With this technology it was possible to push FAEE levels up three-fold to 1.6 g/L and to 28% of the theoretical maximum yield.

A recent study that demonstrated the production of a precursor to the important drug Taxol in *E. coli* is another example of optimization of a pathway by combining copy number modulation with promoter strength. Ajikumar et al. split their pathway into two parts (referred to as the ‘upstream module’ and the ‘downstream module’), and varied copy number and promoter strength of each module independently. By measuring production levels, they mapped out a production landscape as a function of expression level of both the upstream and downstream pathways. Tellingly, production levels did not monotonically increase with increasing expression of either half of the pathway, but rather exhibited localized peaks that indicate improved balancing of the two halves of the pathway.
FROM mRNA TO PROTEIN: TUNING TRANSLATION OF PATHWAY PROTEINS

The overall process of how an mRNA transcript is translated into protein is well-understood. However, making accurate predictions of how much protein will be generated from a given transcript are very difficult. Because fluxes depend strongly on pathway enzyme abundances, especially for rate-limiting enzymes, modulating protein translation from mRNA can help to fine-tune enzyme expression. Translation is also indirectly controlled by mRNA transcript stability. We will focus on efforts to engineer two of the phases of translation (initiation and elongation) in the context of biosynthesis pathways.

It is thought that the rate-limiting step in protein production from a transcript is transcription initiation, a complex process largely determined by binding of the ribosome RNA to the ribosome binding site (RBS), a sequence upstream of the initiation codon. This binding step is controlled by several factors: the presence of secondary structures within the mRNA that occludes the RBS and reduces the availability of RBS to the ribosome; and the equilibrium binding strength of the RBS to the ribosomal RNA. Because these factors are highly dependent on the mRNA sequence upstream and downstream of the RBS, a single RBS will likely result in widely varying translation initiation rates for different surrounding sequences. This may complicate efforts to precisely tune protein levels across a biofuel production pathway.

An elegant method for predicting and controlling RBS-determined translation initiation in E. coli has been developed. It estimates protein expression levels using a thermodynamic model for translation initiation. The model uses an mRNA folding algorithm to calculate the free energies involved in disrupting mRNA secondary structure occluding the RBS, as well as the binding energy of the RBS with ribosomal RNA. Using the RBS calculator, the authors were able to predict protein levels that result from a variety of different RBS sequences reasonably well. Furthermore, their model enables forward engineering of desired expression levels by generating RBS sequences with predicted expression levels for specific proteins. The authors confirmed that RBS sequences that result in high expression activity for a given CDS do not universally apply to other proteins. Using a simple genetic switch, the authors demonstrate the utility of their model using RBS sequences generated from their calculator.

The rate of protein elongation can also be engineered. The most common approach to increasing protein yield from a transcript is to optimize the codon usage within each gene for its new host by replacing codons that occur rarely in the host coding sequences with more commonly used codons. This is thought to speed translation by using codons that correspond to tRNA species that are more abundant in the host organism. Codon optimization of pathway genes has been shown to lead directly to increased protein expression, which in turn results in higher yields. Redding-Johanson and coauthors used transcriptomics and proteomics to understand the rate-limiting steps of an amorphadiene production pathway in E. coli. They found that two proteins in the pathway, mevalonate kinase (MK) and phosphomevalonate kinase (PMK), were expressed at very low levels. By codon-optimizing the genes encoding MK and PMK, the authors were able to increase their protein abundances by about two-fold each, resulting in a two-fold increase in titer. However, the presence of the codon-optimized MK and PMK sequences resulted in a dramatic decrease in the levels of the other pathway enzymes encoded within the operon. The addition of a transcriptional terminator and a second promoter sequence restored the levels of the proteins, and led to a 2.5-fold further increase in amorphadiene production, a total production increase of five-fold over the original operon design.

The previous study also demonstrates the dangers of treating individual genes transcribed within a common operon as noninteracting entities. Because of poorly-understood and context-dependent phenomena such as folding and degradation of the mRNA transcripts,
efforts to optimize the translation level of an individual gene may affect expression of the
other pathway genes in ways that are difficult or impossible to predict. This relationship
between mRNA secondary structure and its effect upon both mRNA stability and translation
has been exploited to create *E. coli* mutants with varying protein concentrations. To
demonstrate this principle, a combinatorial library was generated in which different hairpins
were introduced into the intergenic regions between the three genes responsible for
producing mevalonate *atoB, HMGS*, and *thmGR*. This secondary structure led to expression
differences for each gene in different mutants. With this strategy it was possible to identify
levels of these proteins that lead to the production of seven-fold more mevalonate than the
starting strain. A very recent, more comprehensive model of gene expression takes into
account the kinetics of mRNA folding and degradation, and uses RNA devices (aptazymes
and ribozymes) to tune these parameters directly. RNA devices were placed within the
5′-untranslated region of mRNA to modulate stability of the transcript. This enabled reliable
static and dynamic control over the expression of the downstream gene over a wide range.

HOST ENGINEERING

The organism that hosts the biofuel production pathway will, of course, heavily
influence the behavior of the biofuel production pathway. In addition to expressing and
maintaining the biofuel synthesis pathway, the host organism supplies the metabolic
precursors and enzyme cofactors. However, no organism has evolved to generate chemicals
on a commodity scale for our consumption. Organisms have evolved to address goals (such
as self-replication) that are often at odds with engineering efforts; thus, the engineer must
consider the host physiology and biochemistry if high titers are to be obtained. This section
will review approaches to modifying production hosts that result in improved titers of final
products. For the sake of brevity, we will consider only rational approaches rather than
methods that employ mutagenesis or random, undirected perturbations that are then
screened for improved titers. The reader should consider that we are still quite ignorant of
many details of cellular physiology and metabolism, so undirected approaches often
produce completely unexpected targets for improvements.

KNOCKING OUT COMPETING PATHWAYS

The starting materials of biofuels are typically precursors to building blocks that will
ultimately become components of the cell (fatty acids, amino acids), or secondary
metabolites involved in other processes (isoprenoids). An engineered biofuel pathway will
compete with native pathways for these resources. Removing or attenuating competing
pathways through genetic manipulations have been repeatedly demonstrated to improve
product titers. For instance, in anaerobic conditions in the absence of an electron acceptor,
*E. coli* will use pyruvate as an electron acceptor in mixed acid fermentation, generating
lactate, acetate, and succinate. Pathways used to generate higher alcohols, such as butanol,
use pyruvate as a carbon source. The deletion of key enzymes required for mixed acid
fermentation relieves competition of the biofuel for pyruvate and acetyl-CoA, increasing
yield of the desired final product. Furthermore, as these pathways compete for the
electron carrier NADH, deletion of the pathways also increases the availability of NADH for
alcohol production.

Other deletions that have no direct effect on precursor pools may improve titer by
eliminating the generation of substrates that might act as a competitive inhibitor of an
enzyme along the biofuel production pathway. The product of *ilvD* generates a compound
that acts as a competitive inhibitor for one of the enzymes involved in producing butanol
from a 2-keto acid. Deletion of the *ilvD* resulted in a three-fold production improvement.
Gene deletions may also increase product yield by eliminating pathways that catabolize
biofuel precursors that are present in excess. Steen et al. improved the yield of fatty acids in
E. coli by removing the genes responsible for beta-oxidation of free fatty acids that had accumulated as a result of their earlier interventions. This provided a three- to four-fold increase in fatty acid titer.\textsuperscript{65}

A caveat for gene deletion attempts is that cellular metabolism can be highly robust to gene deletions, compensating for knockouts by increasing flux to a metabolite through other pathways that may affect the biofuel production pathway.\textsuperscript{66} Another caveat is that the deletion of competing pathways may paradoxically decrease the final titer of the desired product, possibly indicating that the competing pathways may in fact generate energy or cofactors that are required for cell growth. Deleting pathways necessary for the generation of an essential metabolite will result in auxotrophy. In this case, rather than deleting the competing pathway, it may be better to decrease the expression of the pathway.\textsuperscript{67} One can imagine a compromise between attenuation and complete elimination of a competing but necessary pathway, such as engineering competing pathways to reduce flux or shut off after the growth phase is completed, enabling precursors to flow entirely into biofuel production. This principle was demonstrated by down-regulating the first gene in the biosynthesis pathway for ergosterol in a strain of yeast producing artisemic acid. Ergosterol is required for survival, and decreasing production without completely eliminating it minimized flux into this competing pathway leading to an increase in artisemic acid titer.\textsuperscript{68,69}

**INCREASING PRECURSOR CONCENTRATION BY UP-REGULATING INPUT PATHWAYS**

Along with deleting competing pathways, the availability of precursor metabolites can be increased by up-regulating the native pathways that produce them. Higher levels of 1-butanol were engineered by overexpressing native E. coli genes *ilvA* and *leuABCD*, increasing the pool of 2-ketobutyrate and directing it towards the synthesis of 2-ketovalerate, precursors of 1-butanol.\textsuperscript{29} This method also uncouples enzyme expression from native transcriptional negative feedback, as highlighted in a production improvement of 2-methyl-1-butanol.\textsuperscript{70} However, overexpressing native genes may not increase flux as much as desired due to allosteric regulation of the enzymes themselves, as discussed above. Furthermore, if a pathway enzyme is inhibited by its own substrate, increasing precursor pools may not increase the flux as quickly as without substrate inhibition.\textsuperscript{71}

**IN SILICO MODELING TO FIND DELETION AND UP-REGULATION CANDIDATES**

Because it can be extremely difficult to predict the effect of knockouts on metabolism, and even more difficult to experimentally screen a large library of knockout mutants for production increases, in silico modeling of cellular metabolism can suggest unintuitive candidates for gene deletion that may increase titers. This approach has been used to improve the production of succinate in E. coli.\textsuperscript{72} In silico methods can also be used to not only pick gene deletion candidates, but also native gene candidates for up-regulation. Choi et al. used an in silico model to predict gene candidates for overexpression that would result in an increased titer of the isoprenoid lycopene.\textsuperscript{73} Impressively, most of the gene overexpression candidates selected resulted in increased yield. Combining several of the overexpression candidates into a single strain resulted in a yield increase of approximately five-fold over the wild-type strain. Gene deletion candidates were also chosen from an in silico model and combined with the amplification targets, resulting in further increases in yield. The authors acknowledge that not all suggested targets resulted in increased yield, potentially due to poorly-understood feedback regulation mechanisms. As our knowledge of post-translational regulation of native pathways grows and becomes incorporated into in silico models, their predictions will likely improve.
STRAIN ENGINEERING FOR INCREASED BIOFUEL TOLERANCE

Because many fuel candidates are toxic to microbial strains used in industrial processes, tolerance of a candidate host organism to the fuel, and to biosynthetic intermediates of that fuel, must be considered when selecting an appropriate production host. If the production host cannot tolerate high concentrations of the biofuel it is designed to make, it might be difficult to produce the compound in high titers. For instance, it makes little sense to use *E. coli* to produce ethanol, as its tolerance to this compound is limited to \(~4\%\), whereas *S. cerevesiae* can withstand far higher concentrations. There are many microbial species known to show high tolerance to hydrophobic solvents, but unfortunately, the more genetically tractable organisms typically show low tolerance to many biofuel candidates. Some exceptions include two promising biodiesel compounds, fatty acid ethyl esters and bisabolene, which have been demonstrated to inflict no toxicity to *E. coli*. Interestingly, there are recent reports of *E. coli* producing butanol above the concentrations where toxicity is seen. We speculate that biofuel tolerance in late growth phase and early stationary phase, where much of the biofuel is typically produced, might be higher than in the exponential growth phase, where toxicity studies are typically conducted. Furthermore, the most commonly used measure of tolerance (inhibition of growth) may not accurately report on the stability of biochemical pathways within the cell.

Increased tolerance can also be engineered by modifying the host organism, or equipping it with mechanisms to resist toxicity. An excellent review of the mechanisms of biofuel toxicity and techniques for engineering-increased tolerance has been written, but will be briefly sketched out here. Microbial species possess efflux pumps that can expel hydrophobic compounds and enable the cells to thrive in the presence of low levels of the toxic compounds. For instance, in *E. coli*, the AcrAB-ToIC efflux pump provides some tolerance to biofuel compounds. More effective efflux pumps from hardier microbial species (e.g. species isolated from oil fields) can be harnessed to increase tolerance in the host organism. This approach was used to demonstrate that expression of solvent-tolerance pumps can result in increased biofuel production. A library of efflux pumps was cloned from a variety of microbial species and expressed in *E. coli* in the presence of several fuel candidates. A competition experiment among strains expressing the various pumps revealed the pumps that bestowed the most tolerance to the fuels. While this study demonstrated the potential of using heterologous efflux pumps for increasing tolerance to biofuels, compatibility of the foreign pumps with the host may need to be adjusted through additional engineering efforts.

A metagenomic library of heterologous genomic DNA was used to increase the tolerance of *E. coli* to various toxins produced in biomass pretreatment. Microbial genomic DNA (gDNA) was isolated from peat bog soil and other environments with high concentrations of toxic inhibitors, and a phage library was made from the gDNA and infected into *E. coli*. Transfected cells that exhibited increased tolerance were obtained by imposing a selection for tolerance to inhibitory concentrations of the toxins. An investigation into the heterologous genes that bestowed higher tolerance suggested that increased tolerance may have been achieved by complementation of an *E. coli* enzyme affected by the toxins with a foreign enzyme that could tolerate the toxin. This study is particularly attractive as it provided a strong clue to the mechanism of toxicity of those particular poisons. While this method was demonstrated to engineer increased tolerance to biomass pretreatment toxins, there seems no obvious reason why it could not also be applied to engineering increased biofuel toxicity as well. It might also provide information as to the specific enzymes that are affected the most in the presence of biofuels, which are often assumed to be systemic.
FUTURE PROSPECTS

This chapter has discussed methods to improve yields of biofuel production pathways. There are many aspects of industrial biofuel production that we lack the space to discuss further. As discussed in the introduction, constraints on economic margins require that the yields of fuel from biomass must be extremely close to the theoretical yield. While impressive yields quite close to theoretical are beginning to appear in the literature, results obtained in laboratory shake flasks, or even small fermentation vessels, rarely scale predictably to large-scale fermentation processes. Researchers often tailor their engineered microbes for laboratory conditions rather than realistic industrial situations. For instance, production on industrial scales will likely occur in anaerobic environments, due to the difficulty in aerating thousand-liter or larger vessels. Also, production results are often (though fortunately not always) reported using rich media, which are far too expensive to use at industrial scales. A biofuel pathway or a host organism optimized to produce biofuel in rich, oxygenated medium may not achieve high yields in minimal medium, in an anaerobic environment. While proof-of-concept results are certainly useful to demonstrate that production in a new host is indeed possible, redesigning a pathway to work at industrial scales often requires efforts that match or exceed those required to achieve the demonstration level. Furthermore, other advantages of laboratory-scale production that are taken for granted, such as lack of contaminants and absence of phage, will probably not apply to industrial scales. These are issues that can be addressed through synthetic biology.

Other complications of biofuel production need to be considered. The sugar stream that the biofuel production host will be converting into biofuels will most likely be the hydrolysate of a high-yielding lignocellulosic feedstock such as switchgrass. Most biofuel produced at the laboratory scale is done with glucose or glycerol; in contrast, plants are composed of five- and six-carbon sugars. Cells growing on purified glucose may have different metabolite pools than cells growing on such complex carbon mixtures, because these sugars enter central carbon metabolism through different routes. It may be necessary to refine the biofuel synthesis pathway to accommodate these metabolic perturbations. In addition, it will probably be necessary to modify the host organism so all sugars are consumed simultaneously rather than sequentially. This is necessary because it leads to more efficient biofuel production and more complete sugar conversion. The host organism can also be engineered to use actual plant biomass, rather than biomass hydrolysate, in order to further reduce processing costs. Given these large obstacles, production of economically viable biofuel may seem incredibly daunting. We anticipate that with applications of both existing and as-yet undiscovered synthetic biology tools, advanced biofuels will become an integral part of the energy economy.

References
28. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as

29. Shen CR, Liao JC. Metabolic engineering of


SECTION III
Applications in Synthetic Biology

42. Jones KL, Kim SW, Keasling JD. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab Eng. 2000;2:328–338.


