

Chapter 2

Gene Expression Engineering

Nicholas J. Morse and Hal S. Alper

Abstract Cellular systems can become platforms for chemical production. Over the last four years, over 50 biopharmaceuticals have been approved for production, ranging in scope from hormones, enzymes, fusion proteins, antibodies, and vaccines. However, each of these applications—whether chemicals or pharmaceuticals, requires both a host organism and tools to engineer pathways in this chosen organism. The cellular hosts for these processes range in scope and complexity to include bacterial systems like *Escherichia coli*, yeast systems like *Saccharomyces cerevisiae*, and a variety of mammalian cell systems. To accomplish these production goals, it is necessary to control gene expression (especially of heterologous genes and pathways). This chapter will evaluate methods for controlling gene expression in the context of heterologous genes, endogenous genes, pathway expression and provide insight into new paradigms for flux control through gene expression circuits. A focus of this chapter will be on the various synthetic tools available for gene expression control. Although these basic principles are broadly applicable to multiple organisms, the predominant focus of this chapter will be on microbial systems, particularly *E. coli* and *S. cerevisiae*.

2.1 Introduction

Cellular systems can become platforms for chemical production. In 2004 and 2007, the U.S. Department of Energy and Pacific Northwest National Laboratory defined a set of top value-added chemicals which could be derived from biomass (Werpy et al. 2004; Holladay et al. 2007). These molecules can be used as precursors for a

N.J. Morse · H.S. Alper (✉)

Department of Chemical Engineering, The University of Texas at Austin,
200 E Dean Keeton St. Stop C0400, Austin, TX 78712, USA
e-mail: halper@che.utexas.edu

H.S. Alper

Institute for Cellular and Molecular Biology, The University of Texas at Austin,
2500 Speedway Avenue, Austin, TX 78712, USA

wide range of industries, including transportation, textiles, food supply, environment, housing, recreation, and health. At the same time, as of 2014, there are over 200 approved biopharmaceutical products produced in cellular systems, accounting for well over \$100 billion in sales (Walsh 2014). Over the last four years, over 50 biopharmaceuticals have been approved for production, ranging in scope from hormones, enzymes, fusion proteins, antibodies, and vaccines. However, each of these applications—whether chemicals or pharmaceuticals, requires both a host organism and tools to engineer pathways in this chosen organism. The cellular hosts for these processes range in scope and complexity to include bacterial systems like *Escherichia coli*, yeast systems like *Saccharomyces cerevisiae*, and a variety of mammalian cell systems. To accomplish these production goals, it is necessary to control gene expression (especially of heterologous genes and pathways).

As a result, selecting the appropriate expression host, understanding how these cellular systems function, and optimizing production of these pharmaceuticals and chemicals have become a focus in the field of metabolic engineering. Of utmost importance is the ability to control gene expression. Thus, this chapter will evaluate methods for controlling gene expression in the context of heterologous genes, endogenous genes, pathway expression and provide insight into new paradigms for flux control through gene expression circuits. A focus of this chapter will be on the various synthetic tools available for gene expression control. Although these basic principles are broadly applicable to multiple organisms, the predominant focus of this chapter will be on microbial systems, particular *E. coli* and *S. cerevisiae*.

2.2 Selecting Host Organisms and the Need for Heterologous Expression

The choice of host organism is often a major and first deciding factor in process optimization. Successfully importing heterologous pathways into organisms requires a delicate balance between the pathway of interest, the required pathway precursor availability, and the capacity for strong overexpression of foreign DNA. Bacterial systems, for example, have the ability to overexpress high quantities of proteins and enzymes (Liu et al. 2013a; Makrides 1996); however more complex biopharmaceuticals, which require more sophisticated posttranslational modification, are often expressed in higher eukaryotic systems (Walsh 2010, 2014; Sanchez and Demain 2012). Nevertheless, a wide range of cells are currently being explored for these broad applications, including bacterial cells, yeast cells, mammalian cells, insect cells, plant cells, and cell-free systems (Sanchez and Demain 2012). We highlight the three major classes of organisms chosen for metabolic engineering and chemical/protein production here.

First, the most popular choice for production especially in proof-of-concept experiments is the bacterium *E. coli*. The well-characterized genome and wide

range of available synthetic tools make *E. coli* one of the most commonly used organisms for heterologous protein expression (Liu et al. 2013a; Sanchez and Demain 2012; Terpe 2006; Chou 2007; Rosano and Ceccarelli 2014). Rapid growth, high cell density fermentations, inexpensive media requirements, and a large range of available expression vectors make *E. coli* a prominent choice as a metabolic engineering host (Terpe 2006). For example, *E. coli* has been engineered to produce biofuel components like butanol and propanol, organic acids like lactic acid and succinic acid, amino acids like threonine and tryptophan, sugar alcohols like xylitol and mannitol, and a variety of drugs and polymers (Chen et al. 2013).

Yeast cells, on the other hand, have many of the advantageous of *E. coli* but provide the ability to perform eukaryotic posttranslational modifications. Much like their bacterial counterparts, yeast cells are economical, can reach high cell densities, and have been successful at producing high protein titers (Celik and Calik 2012). In addition, yeasts have the ability to perform posttranslational modifications like glycosylation, lack pyrogens, and viral inclusions which would be harmful in biopharmaceuticals, produce chaperonins that assist in protein folding, and have a higher tolerance to fermentative conditions (Sanchez and Demain 2012; Celik and Calik 2012; Liu et al. 2013b). As a result, yeasts are seeing an increase in use for the production of value-added chemicals. Their tolerance for acidic conditions makes yeast cells useful for the production of muconic acid, itaconic acid, and ricinoleic acid (Liu et al. 2013b; Blazeck et al. 2014a). In addition, they are seeing an increasing role in the large-scale biosynthesis of lipids and drug precursors (Ro et al. 2006; Blazeck et al. 2014b).

The last of the most commonly used expression systems for heterologous protein production are higher eukaryotic such as mammalian cells—especially for applications of biopharmaceutics as these cells have the capacity to secrete properly folded and glycosylated therapeutic proteins (Martinez et al. 2012; Nielsen 2013; Dalton and Barton 2014; Zhu 2012). Chinese hamster ovary (CHO) cells, for example, are among the most commonly used eukaryotic cells in biopharmaceuticals because of their ability to produce human-like proteins (Sanchez and Demain 2012; Martinez et al. 2012; Nielsen 2013). However, unlike their microbial counterparts, much less is known about their genetics, often limiting their uses to the expression of a single, protein product instead of an entire pathway.

Nevertheless, the wild-type cell rarely produces many of these valued metabolites at a high concentration. For example, *S. cerevisiae* cannot natively produce muconic acid from carbohydrates because of its lack of the necessary enzymes. However, by introducing and expressing the necessary genes for these enzymes inside the cell, the production of valuable precursor chemicals, like muconic acid, could be achieved in cellular expression systems (Curran et al. 2013a). As another example, the production of salvianic acid A in *E. coli* required both new enzymes to be introduced into the cell and certain native protein production to be stopped (Yao et al. 2013). In both cases, the cell may not produce the enzymes needed to convert intermediates in a pathway into the final product. Instead, these deficiencies were fulfilled by expressing heterologous proteins that can perform the needed function inside the host cell.

Thus, engineering entire pathways has two parts. One is the modification of the native cell machinery in order to force flux through a given pathway or prevent the degradation of an intermediate. The other is the expression heterologous proteins that are needed to perform the reactions the cell cannot natively catalyze. The next couple sections of this chapter will focus on the common approaches to performing both of these functions.

2.3 Modifying the Expression of Native Genes

Wild-type cells possess a cascade of regulatory elements that control the native transcriptome within the cell. However, most engineering efforts will force a cell to deviate from its typical, wild-type behavior, requiring genetic rewiring to do so. Thus, a major focus of gene expression control involves the modification of native gene expression. While the tools for both native and heterologous expression (described in the next sections) are similar, the rationale is quite different.

In contrast to heterologous expression, when modifying the expression of a native gene, it is important to realize that this gene, before any modifications, has some innate expression level inside the cell. The goal of engineering efforts is to modify (and potentially remove regulation) from this native expression pattern. If the relative amount of protein from the modified gene *increases* from the innate levels, it is said that the gene is overexpressed. Often times, this can be accomplished by either introducing multiple copies of the gene into the cell or modifying the regulatory region of a gene to swap in higher strength promoters and terminators (Nielsen 2013; Da Silva and Srikrishnan 2012; Redden et al. 2014). By this definition, nearly all heterologous protein expression is said to be a type of overexpression since the cell did not natively produce it prior to its introduction.

Alternatively, gene expression can be changed through a gene knockdown, which *decreases* gene expression relative to the innate levels. This can be accomplished by a total knockout of the gene, silencing the expression in the cell, or knocking down expression through methods such as RNA interference systems (Redden et al. 2014; Suess et al. 2012; Crook et al. 2014; Giaever and Nislow 2014) or promoter replacements. Essential genes that would otherwise inhibit growth of the cells (Giaever and Nislow 2014; Giaever et al. 2002; Winzeler et al. 1999) can be modified with knockdown techniques that can enable the elicitation of favorable phenotypes (Crook et al. 2014). Thus, unlike heterologous expression, a more delicate balance of expression is required when rewiring endogenous genes.

Yet, simply modifying the expression of native genes can be quite powerful and can lead to massive shifts in metabolic flux toward different products, thereby producing interesting or valuable products and phenotypes. For example, lipogenesis is an innate cellular process which converts intermediate products of sugar metabolism into fatty acids and triglycerides. By altering the expression of native enzymes, it is possible to rewire the flux through this pathway to make cells producing upwards of 90 % lipid content (Blazek et al. 2014b).

Therefore, there is a need for altering gene expression—whether to create heterologous function or to enable rewiring of native function. Either way, a suite of synthetic tools is required to accomplish these goals. Thus, the remaining sections of this chapter will first cover how to introduce a new gene into the cell through either a plasmid-based expression system or a genomic integration. Next, strategies for increasing or decreasing the gene expression through approaches such as promoter engineering will be addressed. Following, we will look at regulation at the translational level in order to control the net protein production from expression. Finally, we will address emerging, sophisticated methods for flux control through gene expression control. Examples of applications to metabolic engineering will be addressed throughout.

2.4 Expression of Multiple Copies of a Gene Versus Higher Strength Promoters

DNA editing (either for heterologous pathways or other modifications) is typically performed in one of two ways: plasmid vectors or through genomic integrations. Plasmids are circular, double-stranded DNA molecules that replicate and express autonomously from the cell's chromosome. Therefore, they are useful in applications with only a handful of genes or as a tool for rapid prototyping and assessing the expression of a gene in a cell (Da Silva and Srikrishnan 2012; Madyagol et al. 2011). However, since multiple copies can be maintained in a cell at any one time, plasmid burden has been shown to affect the cell growth (Karim et al. 2013) and long-term stability is a challenge in bioprocessing. Genomic integrations, on the other hand, are typically used for applications requiring more stable and tight regulation on the expression of these genes (Da Silva and Srikrishnan 2012; Madyagol et al. 2011). Single or multiple copies can be inserted into the cell's genome at a time. Therefore, the relative strength of these genes will depend on the promoter and transcriptional regulation at their respective locus.

2.4.1 *Plasmids*

Plasmids are the most common, facile way to transfer genetic information into an organism (Berlec and Strukelj 2013). These elements are characterized based on their stability of replication, copy number in a cell, and segregation into daughter cells (Berlec and Strukelj 2013). These characteristics are controlled by the plasmid's replicon, promoters, selection markers, multiple cloning sites, and fusion protein tags (Rosano and Ceccarelli 2014). As such, a wide range of different combinations of these elements has been made, allowing for flexibility in the expression of a gene.

Autonomously replicating pieces of DNA have a replicon which contains both the origin of replication and the *cis*-acting control elements used to control the plasmid copy number in the cell (Rosano and Ceccarelli 2014; del Solar and Espinosa 2000). In *E. coli*, the three most commonly used origins of replications (*ori*) are the ColE1, the p15A origin, and the pSC101 origin (Rosano and Ceccarelli 2014). ColE1 is an origin of replication that has both a high-copy derivative and a low-copy derivative. The pUC plasmid series uses the high-copy (500–700 copies per cell) derivative of the origin of replication; the low-copy derivative (15–60 copies per cell) is used in the pMB1 plasmid series (Rosano and Ceccarelli 2014; Berlec and Strukelj 2013; Bolivar et al. 1976; Lee et al. 2006; Liang et al. 1999; Minton 1984; Sorensen and Mortensen 2005). However, since these plasmids use origins of replication from the same family, they compete for the same replicative machinery thus preventing maintenance of more than one unique plasmid (del Solar et al. 1998; Camps 2009). To create a multiple plasmid system, plasmids using origins of replications from different families are often used. For example, plasmids containing the p15A origin of replication can be used in combination with ColE1 plasmids when needing to express multiple plasmids in the same cell (Rosano and Ceccarelli 2014; Berlec and Strukelj 2013; Chang and Cohen 1978; Guzman et al. 1995; Nordstrom 2006). Additionally, other low-copy plasmids, such as the pSC101 series (≤ 5 copies per cell), are useful when the expressed protein is inhibitory or toxic to the cell (Stoker et al. 1982; Wang and Kushner 1991). Selection for cells maintaining the plasmids are then made through a variety of antibiotic gene markers. In *E. coli*, this is most readily done using antibiotic resistance genes such as ampicillin, chloramphenicol, kanamycin, or tetracycline (Rosano and Ceccarelli 2014; Berlec and Strukelj 2013). However, antibiotic-free plasmids have also been used to prevent the high cost of antibiotics at the industrial scale (Rosano and Ceccarelli 2014; Chen 2012; Goh and Good 2008; Hägg et al. 2004; Kroll et al. 2009, 2010, 2011; Peubez et al. 2010; Voss and Steinbuchel 2006; Zielenkiewicz and Ceglowski 2001).

Yeast plasmid copy number is also regulated by its origin of replication, much like *E. coli* plasmids. However, many of the common yeast plasmids act as shuttle vectors between *E. coli* and yeast, so they contain two origins of replication: one for yeast and one for *E. coli* (Da Silva and Srikrishnan 2012; Redden et al. 2014). Yeast origins of replications are usually either one of two types. The 2μ origin is often used for high-copy plasmids (≥ 10 copies per cell), while an autonomously replicating sequence and a centromeric sequence (ARS/CEN) are often used for low-copy plasmids (1–2 copies per cell) (Celik and Calik 2012; Da Silva and Srikrishnan 2012; Redden et al. 2014; Clarke and Carbon 1980). Yeast selection markers, unlike their *E. coli* counterparts, commonly use auxotrophic markers for selection. Yeast strains auxotrophic for leucine (*LEU2*), uracil (*URA3*), histidine (*HIS3*), lysine (*LYS2*), or tryptophan (*TRP1*) can carry auxotrophic markers on their plasmids as a selection for cells with the plasmid (Da Silva and Srikrishnan 2012; Redden et al. 2014). Drug resistance markers on vectors, such as resistance to Geneticin (*kanMX4*), nourseothricin (*natNT2*), and hygromycin (*hphNT1*), have

also been used successfully for yeast plasmid systems for negative selections (Taxis and Knop 2006). However, much like *E. coli*, multiple plasmids containing the same origin of replication become burdensome to the cell (Futcher and Carbon 1986; Mead et al. 1986). As with *E. coli*, stability of these plasmids and the need for certain strains (e.g., industrial yeast strains that may not have the appropriate auxotrophy) are self-limiting when considering most industrial scale production needs.

One of the major challenges of a plasmid-based system is the metabolic burden of maintaining large copy numbers of plasmids inside the cell. The production of taxadiene, for example, saw that plasmid copy number for each expression module was a key determinant in the final titer (Ajikumar et al. 2010). The production of polyphosphate and lycopene demonstrate that high-copy plasmids caused such a burden on cells that low-copy plasmid alternatives actually increased overall concentration (Jones et al. 2000). And the biosynthesis of amorphaadiene saw an improvement when modifying the plasmid copy number in the optimization of their pathway (Anthony et al. 2009). In each of these cases, plasmid copy number was attributed to a higher metabolic burden in the cell, leading to slower growth rates, lower cell density, and decreased processing. Instead, engineering a metabolic pathway requires a balance between gene expression and plasmid burden. Plasmid burden has been decreased by introducing expression cassettes into the chromosome of the host organism in a titrated fashion, and gene expression has been controlled through promoter engineering and translational optimization. This will be examined in the sections that follow.

2.4.2 Genomic Integration

The alternative to plasmid-based expression systems is genomic integration. While chromosomal modifications are considered more laborious compared to plasmids, genomic integrations offer greater stability over successive generations of growth and better control over expression. The most common approach to genomic editing is through the use of the cell's native DNA double-stranded breaks repair pathways, such as homologous recombination and nonhomologous end joining, followed by subsequent selection. Moreover, with the recent advent of CRISPR-Cas9 systems, targeted and efficient genomic integrations have become possible (Sander and Joung 2014).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas proteins are a family of nucleases discovered in prokaryotes which specifically cleave DNA of a certain sequence. The most popularly used of these, protein 9 from the *Streptococcus pyogenes* CRISPR system (Cas9), is advantageous for its ability cause a double-stranded DNA based on the sequence of RNA bound to it, therefore promoting the above repair pathways (Sander and Joung 2014). Interchanging this RNA has allowed targeted integrations and knockouts in a wide range of hosts,

including bacterial cells (Jiang et al. 2013), yeast cells (DiCarlo et al. 2013), and mammalian cells (Wu et al. 2014) via both homology direct and nonhomologous end joining pathways. As more *Cas* proteins are discovered, it is anticipated that this field of genomic manipulation will grow.

Other methods for genomic integrations, such as the use of recombinases and transposases, are also available for both bacterial and yeast hosts. Such methods promote the cleavage and rearrangement of DNA, often based on specific recognition sequences. In *E. coli*, the λ bacteriophage Red system or the prophage Rec system can introduce DNA flanked by short regions homologous to the host genome (Madyagol et al. 2011; Datsenko and Wanner 2000). The Cre-Lox system, shown to work in both prokaryotic and eukaryotic systems, induces recombination at a specific recognition sequence (Sauer 1987). Or the use of transposases, such as the Tn7 transposase in *E. coli* (Silva-Rocha and de Lorenzo 2014) or Ty1 and delta (δ) elements in *S. cerevisiae* (Da Silva and Srikrishnan 2012; Genbauffe et al. 1984), facilitate DNA movement into the genome.

2.5 Promoters

Optimal gene expression levels are critical to the success of both heterologous and endogenous pathway engineering efforts. In this vein, transcription of the gene to RNA (controlled by promoters) is the first step of this process. Therefore, understanding and controlling the regulation of transcription allow one to edit the rate of expression. The most common approach to change the rate of expression is by changing the promoter sequence driving the gene(s) of interest. Promoters are responsible for recruiting the necessary transcription machinery and initiating transcription elongation. Many different motifs are contained within a promoter sequence, and the complexity of these elements tends to scale with the complexity of the organism (more complex in higher eukaryotes). Together, these various elements aid the RNA polymerase to find promoter regions and open up the DNA to prepare it for transcription (Feklistov 2013). Promoters are typically categorized as either constitutive or inducible, depending on their activity and may be derived from endogenous, heterologous, or synthetic sources. Constitutive promoters are generally considered “on” under most conditions, while inducible promoters require a stimulus to change the mode of expression.

2.5.1 Constitutive Promoters

A strong emphasis has been placed in the field on the identification of strong, constitutive promoters (typically one of the first to be used to prototype a pathway of interest). For the case of bacteria, while several endogenous, constitutive

promoters can be selected (usually ribosomal in nature), the vast majority of high strength promoters are based on phage sequences. There are several constitutive and native promoters in *E. coli* which have been studied for their expression levels under a variety of conditions (Singh 2014). As an example, Liang et al. characterized seven of these promoters: the *spc* ribosomal protein operon promoter P_{spc} , the β -lactamase gene promoter P_{bla} , the P_L promoter of phage λ , the replication control promoters P_{RNAI} and P_{RNAII} , and the P1 and P2 promoters of the *rrnB* ribosomal RNA operon (Liang et al. 1999). However, despite strong expression from each of these elements, there is still some interdependence on growth conditions.

Therefore, constitutive promoters derived from phages have been introduced into *E. coli* with great success. The T7 promoter, for example, is extremely successful in *E. coli* expression systems as it can lead to the production of target protein in excess of 40–50 % of the total cell protein (Baneyx 1999). To accomplish this, a T7 RNA polymerase must be introduced and expressed separately (some bacterial strains already have this in the genome). However, T7 RNA polymerase is unique in that it only recognizes T7 promoters, allowing for very specialized transcription and orthogonality between the native cell's transcription machinery. As such, this has become a very common expression system in bioprocesses (Rosano and Ceccarelli 2014; Berlec and Strukelj 2013; Baneyx 1999). Moreover, with advances in directed evolution, further orthogonal and distinct T7 promoter-polymerase pairs can further expand the scope of this system (Ellefson et al. 2014).

For the case of yeast, it is not possible to take motivation from phage, and thus most strong, constitutive promoters are those isolated from the yeast's genome (Da Silva and Srikrishnan 2012; Redden et al. 2014). Many of the native promoters characterized and used in yeast systems originate from the glycolytic pathway and have been shown to range in expression. As an example, Sun et al. characterized fourteen constitutive promoters, including the *ADH1*, *TEF1*, *TEF2*, and *GPD* promoters in the context of the most popular yeast expression vectors mentioned previously (Sun et al. 2012; Mumberg et al. 1995). Likewise, Partow et al. characterized seven additional yeast promoters, including the *TEF1*, *ADH1*, *TPI1*, *HXT7*, *TDH3*, *PGK1*, and *PYK1* promoters (Partow et al. 2010). In these characterizations, the relative strengths of the promoters were determined by measuring the amount of protein production enabled by each promoter sequence (Da Silva and Srikrishnan 2012; Redden et al. 2014; Sun et al. 2012; Partow et al. 2010). Unlike the bacterial system, though, these promoters are bounded to the levels of native transcription in yeast as they are derived from native promoters. Several reviews have been published that summarized the various strengths of common promoters for metabolic engineering applications (Da Silva and Srikrishnan 2012; Redden et al. 2014). However, synthetic approaches to promoter engineering are working to expand the toolbox of available promoters beyond those found natively in the cell.

2.5.2 *Synthetic Promoters*

Endogenous promoter sequences can be sufficient for certain applications of pathway engineering. However, their use in larger constructs with multiple genes is challenging because expression can be limited, there is a limited set of strong promoters, these elements may be subject to latent endogenous regulation, and their homology to the genome can prove to be unstable, especially in cells that perform homologous recombination (Dehli et al. 2012). Since the balancing of gene expression is important for pathway engineering (Ajikumar et al. 2010; Li et al. 2013), an increasing set of synthetic promoters and promoter libraries has begun to emerge. We briefly highlight a few of these approaches here.

Synthetic promoters can be designed in several ways. In one approach, error-prone polymerase chain reaction (PCR) mutagenesis is performed on a lead promoter sequence to introduce mutations at the promoter sequence level, thus leading to expression variation. This approach is generalizable across hosts and allows for new promoters to be identified with either increases or decreases in relative expression level (Redden et al. 2014; Alper et al. 2005; Nevoigt et al. 2006; Rajkumar and Maerkl 2012; Blazeck and Alper 2013). Typically, most members of the promoter library will be lower expression than that of the starting, lead promoter. However, libraries with a wide range of expression are of high utility when balancing enzymatic levels in a pathway as it is difficult to a priori determine the optimal expression level. Thus, these synthetic promoter libraries can be coupled with screening to identify the combinations leading to the best phenotype (Alper et al. 2005). Additionally, other synthetic promoter libraries have been used to find promoters that have better regulation than their wild-type counterparts (Nevoigt et al. 2006; Nevoigt et al. 2007). Such efforts have allowed minimal, synthetic promoters to be designed which improve upon native promoters or induce under culture conditions (Redden and Alper 2015).

The other approach to synthetic promoters is through rational design. In these instances, there is even less homology between the resulting synthetic promoter and endogenous promoters. For the case of yeast, within a promoter sequence, it is possible to dissect two major components: a core promoter and an upstream activating sequence (UAS). The core promoter typically contains the TATA box or other necessary transcriptional elements. The UAS, on the other hand, contains binding sites for transcription factors that aid in the recruitment of RNA polymerase to the DNA, but cannot initiate transcription by itself. In many cases, these two elements can be identified separately. Thus, a “hybrid promoter” approach is one in which the UAS sequence from one promoter is stitched to the core promoter of another. Using such an approach, high expression promoters have been made by combining highly active UAS elements to core promoters (Blazeck et al. 2011, 2012). Other promoter characteristics, like nucleosome occupancy, have also been engineered to make high-strength synthetic promoters (Curran et al. 2014). These results have led to pure de novo design of synthetic promoters that contain very little, if any, homology to endogenous elements. However, there is still much to be

learned about the components that influence promoter strength, so the field of rationally designed promoters has room to expand the available toolbox for metabolic engineers.

2.5.3 *Constitutive Versus Inducible Systems*

High-strength, constitutive promoters which are always in the “on” state are extremely useful, especially for the first design of heterologous pathways. However, in order for an expression cassette to be successful at the industrial scale, the promoter must be both highly expressed and tightly regulated (Makrides 1996). In some cases, this desired regulation is constitutive and the promoters described above can suffice. However, the overexpression and accumulation of recombinant and heterologous protein in a cell can often be detrimental to the cell’s productivity (Chou 2007). As such, it is often desirable to introduce tight regulation such that protein production can be turned on and off as desired. In order to do this, inducible promoters that are highly expressed have a low level of basal transcription, are transferable across strains, and have cheap and simple induction need to be used (Makrides 1996; Berlec and Strukelj 2013).

In *E. coli*, a wide variety of inducible promoters is available that can be induced via temperature (Chao et al. 2004; Wang et al. 2012), pH (Makrides 1996), or carbon source (Guzman et al. 1995). The most common and widely studied of these are those promoters derived from the bacterial *lac* operon (Graumann and Premstaller 2006). The *lac* operon and its promoter are known for their ability to respond to lactose or lactose analogs such as isopropyl- β -D-thio-galactoside (IPTG) and induce expression. However, the native *lac* promoter is relatively weak and becomes repressed in the presence of glucose, so derivatives, such as *lacUV5*, have been designed that are less sensitive to catabolites (Silverstone et al. 1970). Elements of this promoter have then been used in combination with the strong T7 RNA polymerase promoter, such as in the pET expression system, to achieve strong, inducible expression of a gene (Rosano and Ceccarelli 2014; Berlec and Strukelj 2013; Makoff and Oxer 1991). However, other expression systems, such as the pBAD system based on the arabinose inducible promoter from the *araBAD* operon, are gaining popularity because the transcription rate scales with arabinose concentration (Guzman et al. 1995). This promoter is also repressed by glucose, though. Lastly, the strong λ promoter P_L has also been successfully used in a similar way by making a tryptophan inducible system (Mieschendahl et al. 1986).

Yeast systems also have a variety of inducible promoters used in expression applications, the *GAL* system of promoters being the most common (Johnston 1987). *GAL1*, *GAL7*, and *GAL10* are three promoters which exhibit induction in the presence of galactose instead of glucose (Bassel and Mortimer 1971). As such, these promoters have become very commonly used since galactose induction is relatively cheap and easy for cell culturing—however, such an element is not desirable when biomass sugars are used (i.e., when glucose will be present).

However, other yeast inducible promoters can also be used. The native gene *CUP1* from *Saccharomyces cerevisiae* is inducible by copper ions (Labbe and Thiele 1999). The promoter for *ADH2* is repressed in the presence of glucose (Price et al. 1990) and the *MET3* and *MET25* promoters are regulated by the presence of methionine (Sangsoda et al. 1984; Cherest et al. 1987). Yet, despite the availability, the *GAL* promoters remain the most commonly used because of their foldexpression change, their response time, and their ease of use in laboratory settings (Adams 1972). Thus, cheap, inducible, non-glucose repressed promoter elements need to be developed for fungal systems.

2.6 Translational Level Regulation

Transcriptional control by promoters can determine the level of overexpression or down regulation at the transcriptional level. After this point, additional bottlenecks/controls exist at the translational level. There are two primary points of regulation at the translational level: translation initiation and translation elongation. Translation initiation is regulated by the ability for tRNA, initiation factors, and rRNA to recognize and bind the messenger RNA and begin the translation process (Jackson et al. 2010). Translational elongation is regulated by the flux of the ribosome unit through the open reading frame (ORF) of the gene and the speed at which it can polymerize the polypeptide chain (Gorgoni et al. 2014). Steps can be taken to optimize both of these steps and increase the net expression of a gene. In addition, it is possible to manipulate mRNA half-life through a variety of techniques.

2.6.1 Translation Initiation

The first step in regulating translation occurs at the translation initiation level. Initiation occurs by first binding the ribosomal subunits and recruiting the necessary machinery to begin translation. In bacteria, most ribosomal binding is determined by the Shine–Dalgarno sequence on the 5' untranslated region (UTR) of the mRNA transcript (Shine and Dalgarno 1975; Malys 2012). This sequence, spaced between 4-18 nucleotides upstream of the start codon, is responsible for recruiting and binding the ribosome near to the start codon. The efficiency of this sequence to recruit ribosomes has been shown to have an almost 1000-fold effect in protein expression (Malys 2012; Curry and Tomich 1988). As such, tools that predict these ribosomal binding sites (RBS) can be used to increase expression values. The RBS Calculator, for example, offers a way to predict and optimize RBS sequences in bacterial systems (Salis 2011). However, RBS strength is still influenced by the neighboring DNA sequence, so a strong Shine–Dalgarno sequence is not guaranteed to maximize translation initiation.

Ribosome binding in yeast, on the other hand, occurs further upstream of the start codon and recognizes the 5' cap structure on mRNA transcripts. The ribosome then scans the transcript until it recognizes the start codon and Kozak sequence where it begins translation (Muller and Trachsel 1990; Yoon and Donahue 1992). In yeast, the optimum Kozak sequence had been identified to be 5'-AAUAAUGG-3' for translation initiation (Hamilton et al. 1987). There have been studies, such as in the production of metallothionein III, where using an optimal consensus sequence in yeast has been shown to increase production (Wang et al. 1998). However, mutational analysis of this sequence showed that alternative sequences were sufficient for expression (Cigan et al. 1988), and because of the proposed scanning mechanism of the ribosome, the start codon closest to the 5' end of the transcript is still usually favored (Kozak 2002).

2.6.2 Translation Elongation

Once translation initiates, the rate of expression will become limited by the rate at which ribosomes are able to move through the open reading frame (ORF) of the gene. Although this rate can be attributed to a couple factors, the most commonly engineered and optimized involves the availability of tRNA. Within the cell, the respective levels of each tRNA vary for each codon. This bias and availability of the correct tRNA for each codon can therefore become a limiting reagent in elongation (Rosano and Ceccarelli 2014; Gorgoni et al. 2014; Kane 1995). So far, there have been two common methods to combat this. One has been to codon optimize the genes to match the host organism. Due to the degeneracy of the genetic code (i.e., multiple tRNA coding for the same amino acid), targeted mutations in the DNA sequence can be used to swap rarer tRNA codons to those coding for more abundant tRNA without a change in amino acid sequence (Sorensen and Mortensen 2005; Gorgoni et al. 2014). Many algorithms exist to predict optimal codon optimization; however, these methods do not guarantee increased expression, and may lead to mRNA instability. An alternative approach has been to insert plasmids that increase the relative expression levels of the rare tRNA and increase their availability (Gorgoni et al. 2014; Kane 1995). Codon-optimized genes can be ordered for a variety of organisms online whereas strains overexpressing rare tRNA are available when codon optimization is not the preferred method of choice or fails to increase expression.

However, despite proper translation optimization, protein expression is still not guaranteed to increase. For example, the heterologous expression of P450 and P450 reductase in yeast saw an increase in protein production upon translation optimization (Batard et al. 2000; Gustafsson et al. 2004). Other systems, such as those used in the production of amorpha-14:0 diene, saw little to no effect on protein output from codon bias optimization (Westfall et al. 2012). Therefore, the effect of

translation optimization tends to be very transcript specific and reliant on multiple factors, such as relative transcript levels and mRNA structures (Gustafsson et al. 2004; Welch et al. 2009; Shah et al. 2013).

2.6.3 Terminator Design

The final method to increase the net rate of translation is to have more mRNA transcript available for the ribosomes. In the prior section, we described how transcript levels could be increased by using higher strength promoters and higher gene copy numbers. While these methods work to increase the production of transcript, the stability of these transcripts also determines their overall concentration in the cell. The selection of terminator can be used to influence net mRNA stability. Specifically, it had been found that mRNA stability changes depending on the terminator sequence and termination pathway (Abe and Aiba 1996). It was also seen that the termination efficiency varied greatly from one terminator sequence to another (Cambray et al. 2013). Therefore, one of the last ways to boost protein production is done by swapping the terminator sequence to a higher strength terminator (Yamanishi et al. 2013; Curran et al. 2013b). High strength terminator sequences have been shown to increase the net protein production by increasing the stability of the mRNA transcript (Yamanishi et al. 2013; Curran et al. 2013b; Yamanishi et al. 2011; Ito et al. 2013) and thus lead to more efficient pathways. Therefore, terminator sequences complement high-expression cassettes by leading to higher mRNA availability and bigger bursts of protein production with lower overall net transcriptional load. The development of terminator libraries and synthetic sequences is still very nascent and there is still much research to be done on the termination mechanism in cells. Thus, this is a growing area of synthetic biology research.

2.7 Synthetic Biology Tools

Many of the common bottlenecks to protein expression and pathway optimization have been traditionally addressed by stitching together a series of biological parts as described above. Specifically, elements including the plasmid, promoter sequence, ribosomal binding site, and terminator sequence are all parts that, when placed together, control the expression of a gene. However, as the library of these parts expands, interesting synthetic tools and paradigms are being developed that aid in the development of more sophisticated expression cassettes and control gene expression. Although this field is still relatively young, we will look at some of the synthetic biology tools currently available for optimizing gene expression pathways.

2.7.1 Logic Circuit Design

Gene cassettes and their parts are expected to have robust expression and tight regulation. For example, many natural biological processes rely on the ability to accurately time gene expression in response to the environmental conditions. There are some gene control elements capable of doing this, such as the inducible promoters described above, which allow expression to be modulated in response to a single environmental trigger. However, as pathways become more complicated, there is a need to make expression cassettes that have a variety of responses to multiple triggers. To do this, engineers have started developing and programming cellular gene circuits into cells in response to complex environmental patterns (Brophy and Voigt 2014).

The logic signaling circuits used in cells are often thought to be similar to the Boolean logic gates used in programming (Morris et al. 2010). For example, the enhancer regions used in gene circuits contain the binding sites for transcription factors that either silence or activate the expression of a gene (Amit 2012). Combinations of these binding sequence motifs make it possible to make Boolean logic AND gates and OR gates which activate transcription in response to which signals are present (Brophy and Voigt 2014; Morris et al. 2010; Ramalingam et al. 2009). In bacteria, the LacI, TetR, and CI repressors were some of the first to be used in the development of logic gates (Brophy and Voigt 2014; Ramalingam et al. 2009). However, the range of inputs has been expanded to include signals from pH, sugar content, metabolites, ligands, light, chemical pulses, or signaling proteins (Brophy and Voigt 2014; Morris et al. 2010; Wang and Buck 2012). Therefore, as this field expands, it will become possible to automate the production of these genetic circuits (Nielsen et al. 2016) and control each gene in a pathway based on the unique signals present (Brophy and Voigt 2014). Such control is expected to help optimize pathway expressions and increase regulation. Moreover, these approaches can lead to more versatile cells that can perform multiple functions all controlled by stimuli.

2.7.2 Synthetic Operons

Another approach to controlling the regulation of a series of genes has been through synthetic operons (sometimes referred to as “refactoring”). Bacterial cells have been studied for a number of years because of their ability to co-transcribe a number of genes as part of the same operon (Okuda et al. 2007). In an operon, multiple genes are transcribed onto the same mRNA and then translated in order to obtain multiple protein products from the same mRNA transcript (and thus controlled by a singular promoter element). The *lac* operon, for example, is one of the most well studied operons in *E. coli* and all bacterial systems (Lewis 2005, 2011). In it, multiple genes are transcribed in response to a single repressor protein that binds upstream of them

on the DNA. Therefore, the use of these operons has started to grow in popularity for pathway engineering approaches.

Controlling the expression of multiple genes in a pathway can be challenging because the native regulation in the cell can be different or complicated for each gene. Synthetic operons allow one to circumvent this regulation by introducing these genes in operons that use well-characterized and regulated parts (Temme et al. 2012). For example, synthetic operons make it possible to refactor entire gene clusters onto a single operon that can be induced by different signals (Temme et al. 2012). The organization of these genes on the operon, the strength of the ribosome binding site, and posttranslational regulation can then be used to modulate the relative expression levels of each gene individually (Lim et al. 2011; Levin-Karp et al. 2013; Agnew and Pflieger 2011). This way, reactions can be coupled together and controlled with regulation that is well known and characterized for entire pathways (Temme et al. 2012; Lu and Ellington 2014; Matsumoto et al. 2011). This approach can be used to activate secondary metabolite production and remove regulation found in native pathways. In this regard, these approaches enable complete, synthetic control of a pathway.

2.7.3 *Synthetic Feedback Loops*

The balance of expression is important for the engineering of metabolic pathways, particularly when toxic or unproductive intermediates exist. For example, microbes can be used for biofuel production, but the fuels themselves are often toxic to cell growth (Dunlop et al. 2010). In traditional approaches, gene expression would have to be balanced through promoter engineering in order to prevent an overproduction of these toxic compounds. However, these approaches are laborious and cannot always account for fluctuations in environmental conditions.

Therefore, synthetic feedback loops have been used to add robustness to pathways and increase expression control. In a feedback loop, synthetic parts are designed that autoregulate their own expression in order to prevent toxic buildup in a cell. This results in a dynamic control over pathways rather than the static control afforded by constitutive promoters. When the expression of a protein increases its own expression, it is called a positive feedback loop; when a protein down regulates its own expression, it is a negative feedback loop. In the biofuel example, feedback from biofuel drove the expression of an efflux pump to keep the level of biofuel inside the cell from being prohibitive to growth (Dunlop et al. 2010; Harrison and Dunlop 2012). However, other times feedback loops are used to control transcriptional noise. Negative feedback loops allow gene expression to stay relatively static in response to environmental conditions that would otherwise skew their expression (Holtz and Keasling 2010). Other times, positive feedback loops allow biosensors to be more sensitive to external conditions by overexpressing when sensing metabolites (Kobori et al. 2013). Lastly, a combination of positive and negative feedback loops is used in synthetic circuit parts, such as oscillators, which

allow cellular dynamics to be studied (Singh 2014; Brophy and Voigt 2014; Fung et al. 2005; Chen and Arkin 2012; Stricker et al. 2008; Elowitz and Leibler 2000; Atkinson et al. 2003; Yokobayashi et al. 2002; Gardner et al. 2000). However, the overarching idea of synthetic feedback loops is the ability to control a robust expression system that can adapt to environmental changes (either outside or inside the cell). These approaches can enable a flexible pathway that can respond to perturbation.

2.7.4 Metabolic Engineering Using Synthetic Biology Tools

Many of the synthetic biology tools listed above have applications in pathway and metabolic engineering by themselves. However, the combination of multiple approaches from the above has expedited and advanced metabolic engineering in a variety of applications. The production of taxadiene, an intermediate for the anti-cancer drug Taxol, is one example (Ajikumar et al. 2010). In this study, expression modules of a native pathway and a heterologous pathway were created using synthetic operons with a variety of promoter strengths. By collecting data under a variety of conditions, including plasmid-based and genomic-based expression, promoter strengths, and operon organization, they used multivariate analysis to obtain a 15,000-fold improvement in the production of taxadiene (Ajikumar et al. 2010). Whereas many studies would test these conditions separately, the use of multivariate analysis greatly improved the influences each variable also had on each other.

Another example is one in which the yield of biofuel production was increased threefold (Zhang et al. 2012). In this case, the balance of an intermediate metabolite in the biofuel production pathway was a key to maximizing the overall yield. Thus, these researchers used a biosensor that could change the protein production of the pathway in response to the intermediate's level. The pathway itself was split into multiple, synthetic modules that were then optimized underneath different promoter controls until an optimal expression level was obtained. Feedback loops were then used to increase and decrease gene expression levels and push metabolic flux into the production of fatty acids, increasing production to 28 % theoretical yield (Zhang et al. 2012).

These two examples emphasize that importance of finding the proper balance of enzymes needed for production. In both of these cases, varying the expression of gene modules, in combination with feedback loops controlling the level of intermediates, was able to achieve optimal production. As such, these synthetic tools offer very powerful advantages to a metabolic engineer, but there is still much to be discovered.

2.8 Conclusions

Many of the challenges in metabolic engineering come from the ability to properly balance the necessary proteins and enzymes in a pathway. It is often difficult to ab initio predict the optimal expression level for multiple genes in a pathway, so engineering has been largely empirical. However, throughout this chapter, we presented many of the tools available to a metabolic engineer in order to achieve fine control over the level of gene expression. Together, these various parts were designed to control the transcriptional level either at the transcription initiation phase or in the posttranscription and translation level.

The advances of synthetic biology, though, are certainly helping to create sophisticated control systems that can autoregulate pathways. Such systems and techniques are providing the promise of fully dynamic control in the cells, thus establishing a step toward continuous process optimization. The rapid screening of synthetic and pathway libraries is increasing the speed at which pathways can be optimized for overall productivity. Therefore, as we move forward and these tools are developed, we will further have our control over metabolism and develop better ways to rewire cells into cellular factories.

References

- Abe H, Aiba H (1996) Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. *Biochimie* 78(11–12):1035–1042
- Adams BG (1972) Induction of galactokinase in *Saccharomyces cerevisiae*: kinetics of induction and glucose effects. *J Bacteriol* 111(2):308–315
- Agnew DE, Pflieger BF (2011) Optimization of synthetic operons using libraries of post-transcriptional regulatory elements. *Methods Mol Biol* 765:99–111
- Ajikumar PK et al (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330(6000):70–74
- Alper H et al (2005) Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* 102(36):12678–12683
- Amit R (2012) Towards synthetic gene circuits with enhancers: biology’s multi-input integrators. *Subcell Biochem* 64:3–20
- Anthony JR et al (2009) Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab Eng* 11(1):13–19
- Atkinson MR et al (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 113(5):597–607
- Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* 10(5):411–421
- Bassel J, Mortimer R (1971) Genetic order of the galactose structural genes in *Saccharomyces cerevisiae*. *J Bacteriol* 108(1):179–183
- Batard Y et al (2000) Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. *Arch Biochem Biophys* 379(1):161–169
- Berlec A, Strukelj B (2013) Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol* 40(3–4):257–274

- Blazeck J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* 8(1):46–58
- Blazeck J et al (2011) Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl Environ Microbiol* 77(22):7905–7914
- Blazeck J et al (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol Bioeng* 109(11):2884–2895
- Blazeck J et al (2014a) Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production. *Appl Microbiol Biotechnol* 98(19):8155–8164
- Blazeck J et al (2014b) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 5:3131
- Bolivar F et al (1976) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2(2):95–113
- Brophy JA, Voigt CA (2014) Principles of genetic circuit design. *Nat Methods* 11(5):508–520
- Cambrey G et al (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res* 41(9):5139–5148
- Camps M (2009) Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Pat DNA Gene Sequences* 4(1):58–73
- Celik E, Calik P (2012) Production of recombinant proteins by yeast cells. *Biotechnol Adv* 30(5):1108–1118
- Chang AC, Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* 134(3):1141–1156
- Chao YP, Wen CS, Wang JY (2004) A facile and efficient method to achieve LacZ overproduction by the expression vector carrying the thermoregulated promoter and plasmid copy number. *Biotechnol Prog* 20(2):420–425
- Chen R (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv* 30(5):1102–1107
- Chen D, Arkin AP (2012) Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Mol Syst Biol* 8:620
- Chen X et al (2013) Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnol Adv* 31(8):1200–1223
- Cherest H, Kerjan P, Surdin-Kerjan Y (1987) The *Saccharomyces cerevisiae* MET3 gene: nucleotide sequence and relationship of the 5' non-coding region to that of MET25. *Molecular & general genetics: MGG* 210(2):307–313
- Chou CP (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl Microbiol Biotechnol* 76(3):521–532
- Cigan AM, Pabich EK, Donahue TF (1988) Mutational analysis of the HIS4 translational initiator region in *Saccharomyces cerevisiae*. *Mol Cell Biol* 8(7):2964–2975
- Clarke L, Carbon J (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* 287(5782):504–509
- Crook NC, Schmitz AC, Alper HS (2014) Optimization of a yeast RNA interference system for controlling gene expression and enabling rapid metabolic engineering. *ACS Synth Biol* 3(5):307–313
- Curran KA et al (2013a) Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab Eng* 15:55–66
- Curran KA et al (2013b) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng* 19:88–97
- Curran KA et al (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun* 5:4002
- Curry KA, Tomich CS (1988) Effect of ribosome binding site on gene expression in *Escherichia coli*. *DNA* 7(3):173–179
- Da Silva NA, Srikrishnan S (2012) Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 12(2):197–214

- Dalton AC, Barton WA (2014) Over-expression of secreted proteins from mammalian cell lines. *Protein Sci* 23(5):517–525
- Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97(12):6640–6645
- Dehli T, Solem C, Jensen PR (2012) Tunable promoters in synthetic and systems biology. *Subcell Biochem* 64:181–201
- del Solar G, Espinosa M (2000) Plasmid copy number control: an ever-growing story. *Mol Microbiol* 37(3):492–500
- del Solar G et al (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 62(2):434–464
- DiCarlo JE et al (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41(7):4336–4343
- Dunlop MJ, Keasling JD, Mukhopadhyay A (2010) A model for improving microbial biofuel production using a synthetic feedback loop. *Syst Synth Biol* 4(2):95–104
- Ellefson JW et al (2014) Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol* 32(1):97–101
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767):335–338
- Feklistov A (2013) RNA polymerase: in search of promoters. *Ann N Y Acad Sci* 1293:25–32
- Fung E et al (2005) A synthetic gene-metabolic oscillator. *Nature* 435(7038):118–122
- Futcher B, Carbon J (1986) Toxic effects of excess cloned centromeres. *Mol Cell Biol* 6(6):2213–2222
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403(6767):339–342
- Genbauffe FS, Chisholm GE, Cooper TG (1984) Tau, sigma, and delta. A family of repeated elements in yeast. *J Biol Chem* 259(16):10518–10525
- Giaever G, Nislow C (2014) The yeast deletion collection: a decade of functional genomics. *Genetics* 197(2):451–465
- Giaever G et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418(6896):387–391
- Goh S, Good L (2008) Plasmid selection in *Escherichia coli* using an endogenous essential gene marker. *BMC Biotechnol* 8:61
- Gorgoni B et al (2014) Controlling translation elongation efficiency: tRNA regulation of ribosome flux on the mRNA. *Biochem Soc Trans* 42(1):160–165
- Graumann K, Premstaller A (2006) Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnol J* 1(2):164–186
- Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* 22(7):346–353
- Guzman LM et al (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177(14):4121–4130
- Hägg P et al (2004) A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*. *J Biotechnol* 111(1):17–30
- Hamilton R, Watanabe CK, de Boer HA (1987) Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs. *Nucleic Acids Res* 15(8):3581–3593
- Harrison ME, Dunlop MJ (2012) Synthetic feedback loop model for increasing microbial biofuel production using a biosensor. *Front Microbiol* 3:360
- Holladay JE et al (2007) Top value-added chemicals from biomass: results of screening for potential candidates from biorefinery lignin. Pacific Northwest National Laboratory
- Holtz WJ, Keasling JD (2010) Engineering static and dynamic control of synthetic pathways. *Cell* 140(1):19–23
- Ito Y et al (2013) Characterization of five terminator regions that increase the protein yield of a transgene in *Saccharomyces cerevisiae*. *J Biotechnol* 168(4):486–492

- Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11(2):113–127
- Jiang W et al (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31(3):233–239
- Johnston M (1987) A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiol Rev* 51(4):458–476
- Jones KL, Kim SW, Keasling JD (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* 2(4):328–338
- Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol* 6(5):494–500
- Karim AS, Curran KA, Alper HS (2013) Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res* 13(1):107–116
- Kobori S et al (2013) A controllable gene expression system in liposomes that includes a positive feedback loop. *Mol BioSyst* 9(6):1282–1285
- Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299(1–2):1–34
- Kroll J et al (2009) Establishment of a novel anabolism-based addiction system with an artificially introduced mevalonate pathway: Complete stabilization of plasmids as universal application in white biotechnology. *Metab Eng* 11(3):168–177
- Kroll J et al (2010) Plasmid addiction systems: perspectives and applications in biotechnology. *Microb Biotechnol* 3(6):634–657
- Kroll J, Klinter S, Steinbuechel A (2011) A novel plasmid addiction system for large-scale production of cyanophycin in *Escherichia coli* using mineral salts medium. *Appl Microbiol Biotechnol* 89(3):593–604
- Labbe S, Thiele DJ (1999) Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol* 306:145–153
- Lee C et al (2006) Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J Biotechnol* 123(3):273–280
- Levin-Karp A et al (2013) Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. *ACS Synth. Biol* 2(6):327–336
- Lewis M (2005) The lac repressor. *C R Biol* 328(6):521–548
- Lewis M (2011) A tale of two repressors. *J Mol Biol* 409(1):14–27
- Li Z et al (2013) Direct and efficient xylitol production from xylan by *Saccharomyces cerevisiae* through transcriptional level and fermentation processing optimizations. *Bioresour Technol* 149:413–419
- Liang S et al (1999) Activities of constitutive promoters in *Escherichia coli*. *J Mol Biol* 292(1):19–37
- Lim HN, Lee Y, Hussein R (2011) Fundamental relationship between operon organization and gene expression. *Proc Natl Acad Sci USA* 108(26):10626–10631
- Liu L et al (2013a) How to achieve high-level expression of microbial enzymes: strategies and perspectives. *Bioengineered* 4(4):212–223
- Liu L, Redden H, Alper HS (2013b) Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*. *Curr Opin Biotechnol* 24(6):1023–1030
- Lu WC, Ellington AD (2014) Design and selection of a synthetic operon. *ACS Synth Biol* 3(6):410–415
- Madyagal M et al (2011) Gene replacement techniques for *Escherichia coli* genome modification. *Folia Microbiol (Praha)* 56(3):253–263
- Makoff AJ, Oxer MD (1991) High level heterologous expression in *E. coli* using mutant forms of the lac promoter. *Nucleic Acids Res* 19(9):2417–2421
- Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 60(3):512–538
- Malys N (2012) Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. *Mol Biol Rep* 39(1):33–39

- Martinez JL et al (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr Opin Biotechnol* 23(6):965–971
- Matsumoto Y et al (2011) Bacterial cells carrying synthetic dual-function operon survived starvation. *J Biomed Biotechnol* 2011:489265
- Mead DJ, Gardner DC, Oliver SG (1986) The yeast 2 micron plasmid: strategies for the survival of a selfish DNA. *Mol Gen Genet* 205(3):417–421
- Mieschendahl M, Petri T, Hanggi U (1986) A Novel prophage independent Trp regulated lambda-PI expression system. *Bio-Technology* 4(9):802–809
- Minton NP (1984) Improved plasmid vectors for the isolation of translational lac gene fusions. *Gene* 31(1–3):269–273
- Morris MK et al (2010) Logic-based models for the analysis of cell signaling networks. *Biochemistry* 49(15):3216–3224
- Muller PP, Trachsel H (1990) Translation and regulation of translation in the yeast *Saccharomyces cerevisiae*. *Eur J Biochem* 191(2):257–261
- Mumberg D, Muller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156(1):119–122
- Nevoigt E et al (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 72(8):5266–5273
- Nevoigt E et al (2007) Engineering promoter regulation. *Biotechnol Bioeng* 96(3):550–558
- Nielsen J (2013) Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* 4(4):207–211
- Nielsen AA et al (2016) Genetic circuit design automation. *Science* 352(6281):aac7341
- Nordstrom K (2006) Plasmid R1–replication and its control. *Plasmid* 55(1):1–26
- Okuda S et al (2007) Characterization of relationships between transcriptional units and operon structures in *Bacillus subtilis* and *Escherichia coli*. *BMC Genom* 8:48
- Partow S et al (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* 27(11):955–964
- Peubez I et al (2010) Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microb Cell Fact* 9:65
- Price VL et al (1990) Expression of heterologous proteins in *Saccharomyces cerevisiae* using the ADH2 promoter. *Methods Enzymol* 185:308–318
- Rajkumar AS, Maerkl SJ (2012) Rapid synthesis of defined eukaryotic promoter libraries. *ACS Synth Biol* 1(10):483–490
- Ramalingam KI et al (2009) Forward engineering of synthetic bio-logical AND gates. *Biochem Eng J* 47(1–3):38–47
- Redden H, Alper HS (2015) The development and characterization of synthetic minimal yeast promoters. *Nat Commun* 6:7810
- Redden H, Morse N, Alper HS (2014) The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Res*
- Ro DK et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440(7086):7940–7943
- Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5:172
- Salis HM (2011) The ribosome binding site calculator. *Methods Enzymol* 498:19–42
- Sanchez S, Demain A (2012) Special issue on the production of recombinant proteins. *Biotechnol Adv* 30(5):1100–1101
- Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32(4):347–355
- Sangsoda S, Cherest H, Surdin-Kerjan Y (1984) The expression of the MET25 gene of *Saccharomyces cerevisiae* is regulated transcriptionally. *Molecular & general genetics: MGG* 200(3):407–414
- Sauer B (1987) Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 7(6):2087–2096

- Shah P et al (2013) Rate-limiting steps in yeast protein translation. *Cell* 153(7):1589–1601
- Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. *Nature* 254(5495):34–38
- Silva-Rocha R, de Lorenzo V (2014) Chromosomal integration of transcriptional fusions. *Methods Mol Biol* 1149:479–489
- Silverstone AE, Arditti RR, Magasanik B (1970) Catabolite-insensitive revertants of lac promoter mutants. *Proc Natl Acad Sci U S A* 66(3):773–779
- Singh V (2014) Recent advancements in synthetic biology: current status and challenges. *Gene* 535(1):1–11
- Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 115(2):113–128
- Stoker NG, Fairweather NF, Spratt BG (1982) Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene* 18(3):335–341
- Stricker J et al (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* 456(7221):516–519
- Suess B et al (2012) Aptamer-regulated expression of essential genes in yeast. *Methods Mol Biol* 824:381–391
- Sun J et al (2012) Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 109(8):2082–2092
- Taxis C, Knop M (2006) System of centromeric, episomal, and integrative vectors based on drug resistance markers for *Saccharomyces cerevisiae*. *Biotechniques* 40(1):73–78
- Temme K, Zhao D, Voigt CA (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci U S A* 109(18):7085–7090
- Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 72(2):211–222
- Voss I, Steinbuchel A (2006) Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16. *Metab Eng* 8(1):66–78
- Walsh G (2010) Biopharmaceutical benchmarks 2010. *Nat Biotechnol* 28(9):917–924
- Walsh G (2014) Biopharmaceutical benchmarks 2014. *Nat Biotechnol* 32(10):992–1000
- Wang B, Buck M (2012) Customizing cell signaling using engineered genetic logic circuits. *Trends Microbiol* 20(8):376–384
- Wang RF, Kushner SR (1991) Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* 100:195–199
- Wang SH, Shih YH, Lin LY (1998) Yeast consensus initiator sequence markedly enhances the synthesis of metallothionein III in *Saccharomyces cerevisiae*. *Biotechnol Lett* 20(1):9–13
- Wang Z et al (2012) Expression and production of recombinant cis-epoxysuccinate hydrolase in *Escherichia coli* under the control of temperature-dependent promoter. *J Biotechnol* 162(2–3):232–236
- Welch M et al (2009) You're one in a googol: optimizing genes for protein expression. *J R Soc Interface* 6(Suppl 4):S467–S476
- Werpy T et al (2004) Top value added chemicals from biomass: results of screening for potential candidates from sugars and synthesis gas. [U.S. Department of Energy [Office of] Energy Efficiency and Renewable Energy
- Westfall PJ et al (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinin acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A* 109(3):E111–E118
- Winzler EA et al (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429):5901–5906
- Wu X et al (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 32(7):670–676

- Yamanishi M, Katahira S, Matsuyama T (2011) TPS1 terminator increases mRNA and protein yield in a *Saccharomyces cerevisiae* expression system. *Biosci Biotechnol Biochem* 75 (11):2234–2236
- Yamanishi M et al (2013) A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a “terminatome” toolbox. *ACS Synth Biol* 2(6):337–347
- Yao YF et al (2013) Metabolic engineering of *Escherichia coli* for production of salvianic acid A via an artificial biosynthetic pathway. *Metab Eng* 19:79–87
- Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci USA* 99(26):16587–16591
- Yoon H, Donahue TF (1992) Control of translation initiation in *Saccharomyces cerevisiae*. *Mol Microbiol* 6(11):1413–1419
- Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* 30(4):354–359
- Zhu J (2012) Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv* 30(5):1158–1170
- Zielenkiewicz U, Ceglowski P (2001) Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. *Acta Biochim Pol* 48(4):1003–1023