Anton Glieder · Christian P. Kubicek
Diethard Mattanovich · Birgit Wiltschi
Michael Sauer *Editors*

Synthetic Biology

Springer

Synthetic Biology

Anton Glieder • Christian P. Kubicek •
Diethard Mattanovich • Birgit Wiltschi •
Michael Sauer
Editors

Synthetic Biology

Springer

Editors
Anton Glieder
ACIB GmbH
Technische Universitaet Graz
Graz, Austria

Christian P. Kubicek
Research Division Biotechnology
and Microbiology
Vienna University of Technology
Vienna, Austria

Diethard Mattanovich
Department of Biotechnology
University of Natural Resources
and Life
Vienna, Austria

Birgit Wiltschi
Austrian Centre of Industrial Biotechnology
Graz, Austria

Michael Sauer
Department of Biotechnology
University of Natural Resources
and Life
Vienna, Austria

ISBN 978-3-319-22707-8 ISBN 978-3-319-22708-5 (eBook)
DOI 10.1007/978-3-319-22708-5

Library of Congress Control Number: 2015954947

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Synthetic biology is often cited as one of the largest and fastest growing but less defined trends in life science technologies. Nevertheless, driven by open technology platforms, technical standards, and success stories of applied synthetic biology, this young scientific area became more than a grant-friendly hype in the past 10–15 years.

Scientists have been manipulating genes for decades: insertion, deletion, and modification of genes and their expression have become a routine function in thousands of labs. Yet by the beginning of the twenty-first century, our ability to modify the DNA and the genetic code through molecular biology had endowed scientists to use cells as hardware, and the genetic code as the software to design microorganisms for new purposes that stretched beyond the goals that could be reached by so far used recombinant techniques. This includes new strategies for engineering the transcriptional apparatus, creating novel DNA and RNA elements, expansion of the genetic code, as well as pathway engineering and cellular remodelling towards no producer strains, and the chemical synthesis of novel biocompatible polymers. Today, scientists from a growing number of disciplines such as biology, engineering, chemistry, and bioinformatics interact as a self-defined global community in cross-disciplinary approaches applying the principles of engineering to facilitate and accelerate the design, manufacture, and/or modification of genetic materials in living organisms.

Recent advances in technologies, the availability of cheap DNA building blocks, and concerted educational events paved the way to plan efforts in silico, to understand life via building, and to engineer biology based on thousands of easily accessible well-defined parts and methods. The implementation of first industrial production processes such as the semisynthetic production of artemisinin after intense biological, chemical, and process engineering demonstrated that synthetic biology is more than useful for research but also to the benefit of human health.

This book was written by international experts in the attempt to provide a contemporary summary of the achievements in these areas as reached today, both for the purpose of updating the beginners and stimulating the development of ideas for those already working in this field.

Graz, Austria — Anton Glieder
Vienna, Austria — Christian P. Kubicek
July 2015

Contents

Programming Biology: Expanding the Toolset for the Engineering of Transcription

1

Bob Van Hove, Aaron M. Love, Parayil Kumaran Ajikumar, and Marjan De Mey

Contents

B. Van Hove • M. De Mey (✉)
Centre for Industrial Biotechnology and Biocatalysis, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
e-mail: Marjan.DeMey@UGent.be

A.M. Love • P.K. Ajikumar
Manus Biosynthesis, 1030 Massachusetts Avenue, Suite 300, Cambridge, MA 02138, USA

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_1

Abstract

Transcription is a complex and dynamic process representing the first step in gene expression that can be readily controlled through current tools in molecular biology. Elucidating and subsequently controlling transcriptional processes in various prokaryotic and eukaryotic organisms have been a key element in translational research, yielding a variety of new opportunities for scientists and engineers. This chapter aims to give an overview of how the fields of molecular and synthetic biology have contributed both historically and presently to the state of the art in transcriptional engineering. The described tools and techniques, as well as the emerging genetic circuit engineering discipline, open the door to new advances in the fields of medical and industrial biotechnology.

1.1 Introduction

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. (Watson and Crick 1953)

With this concluding remark to their groundbreaking 1953 paper, Watson and Crick laid the groundwork for what is now known as "the central dogma of molecular biology." In essence, the rule states that the molecular flow of genetic information begins with DNA, which is followed by the intermediate RNA, and finally ends with protein (Crick 1970). These processes were termed transcription and translation. Figure 1.1 shows a schematic representation of the major components involved in the process of transcription. As the field of molecular biology began unfolding, researchers elucidated various mechanisms by which gene expression is regulated and subsequently developed tools capable of manipulating these processes. Early pioneers in biotechnology recognized the opportunities for genetically engineering microorganisms and evolved the field of metabolic engineering to broaden the scope of biotechnological production of chemicals and fuels (Bailey 1991). Recently, as biology entered the post-genomic era, molecular tools and techniques had gotten so advanced that entire *new-to-nature* genetic networks could be created, enabling the development of the field of synthetic biology (Stephanopoulos 2012).

Today, scientists and engineers have a wide range of natural and synthetic tools at their disposal, which include not only techniques for regulating transcription, but also methods that target the translational and posttranslational stages of gene expression. Manipulating gene expression posttranscriptionally holds great promise as well (Chappell et al. 2013), but is outside of the scope of this chapter. We present here a valuable toolkit that can be utilized to engineer the transcription of DNA into RNA, effectively programming life itself. After giving a brief overview of

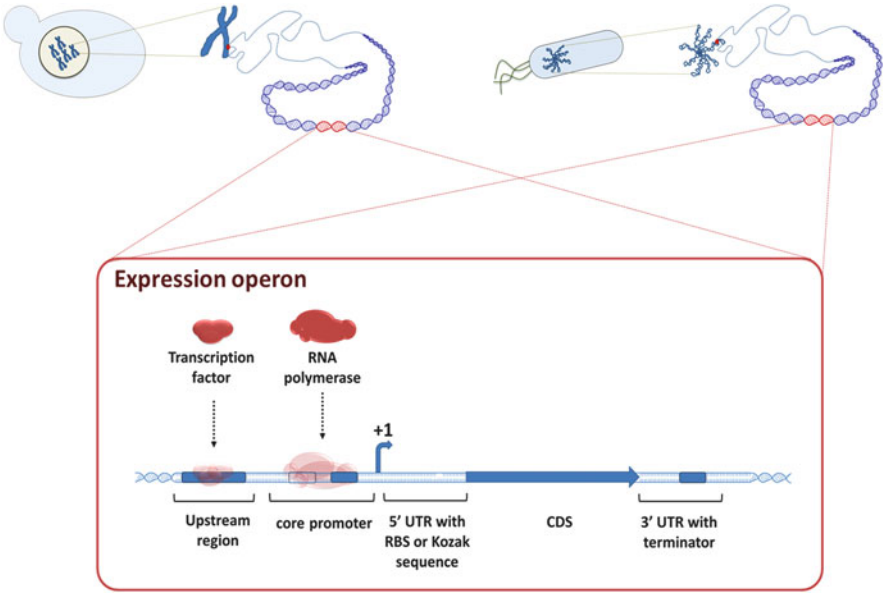

Fig. 1.1 Schematic representation of gene expression and the various components involved in the process of transcription. The central dogma of molecular biology states that DNA is transcribed to messenger RNA (mRNA), which is in turn translated to protein. Transcription is initiated by binding of the RNA polymerase (RNAP) to specific elements in the core promoter and/or upstream region. In bacteria this process can be facilitated by "UP elements" and a set of consensus hexamers at the −35 and −10 positions upstream to the transcription start site (denoted by "+1"). Recognition is primarily dictated by these consensus sequences through the action of an RNAP associated sigma factor (σ). In eukaryotes the process is more complicated, requiring at least seven different transcription factors (TFs) for the binding of RNAP II to the promoter, and regulatory elements can be several kilobases away from the transcriptional start site. Eukaryotic RNAP II-dependent promoters are not as conserved as prokaryotic promoters, but can contain a TATA element and a B recognition element (BRE). Transcriptional termination is mediated by the sequence downstream of the coding DNA sequence (CDS) called terminator. Throughout prokaryotic genomes, two classes of transcription terminators, Rho dependent and Rho independent, have been identified. During Rho-independent termination, a terminating hairpin formed on the nascent mRNA interacts with the NusA protein to stimulate release of the transcript from the RNA polymerase complex. In Rho-dependent termination, the Rho protein binds at an upstream site, translocates down the mRNA, and interacts with the RNAP complex to stimulate release of the transcript. Termination during eukaryotic transcription of mRNAs is governed by terminator signals that are recognized by protein factors associated with the RNAP II, which trigger the termination process. During the process of translation, mRNA is interpreted by a ribosome to produce a specific amino acid chain, i.e., protein. The ribosome initially binds to a Shine–Dalgarno sequence in prokaryotes and a Kozak sequence in eukaryotes located in the 5′ untranslated region (5′ UTR)

reengineered natural systems, we discuss synthetic systems and the *state-of-the-art* techniques used to construct them. Next we illustrate how to apply these techniques for the construction of complex genetic circuits, ending the chapter with applications in medicine and industry.

1.2 Reengineering Natural Systems for New Applications

1.2.1 The Beginnings

Biological organisms naturally must exert control over their transcriptome using a variety of regulatory mechanisms, several of which have been well characterized, but a host that have yet to be entirely understood. Continued discovery of natural mechanisms of transcriptional control will provide the raw material for rationally engineering natural regulatory parts, as well as designing new ones for precise control over synthetic expression systems. Current strides being made in research using genetic regulation owe their success to the early work of several groups, who were able to elucidate the transcriptional properties and regulatory aspects of transcriptional systems including the *lac* operon and viral promoters.

Since Jacob and Monod initially investigated the *lac* operon in 1961, it has been the focal point of much research concerning transcriptional regulation and has continued to provide a model basis for research today (Jacob and Monod 1961). The well-characterized *lac* operon contains discrete types of elements that are present in most bacterial promoters, including a core promoter with consensus sequences (i.e. −35 box and −10 box) and operator sequences to which regulatory proteins can bind (Oehler and Amouyal 1994). Promoters including the *lac*, *tet*, and *ara* promoters have been used for protein expression in their native form, as well as in engineered contexts. Lutz and Bujard (1997) demonstrated that elements from the aforementioned sequences can be combined to form novel tightly repressible promoters having several thousandfold better regulation than their native elements. The *lac* operon has also been the basis for predictive algorithms able to accurately correlate theoretical binding properties of transcriptional regulators to the observed repressor state, paving the way for computational approaches to inspire new synthetic promoter designs (Vilar and Saiz 2013). The ability to modularize natural operators and predict their output has allowed for the generation of promoters with novel activators or repressors and unique functionalities useful for artificial transcription systems. An alternative to using native host transcription machinery is to introduce additional RNA polymerases such as those encoded by bacteriophages and other viruses.

Viral promoters were first utilized for recombinant protein expression in the 1980s (Studier and Moffatt 1986), using a promoter and RNA polymerase from bacteriophage T7 for gene expression in *E. coli*. This work paid off tremendously, as the T7 promoter–polymerase pair is still highly regarded as a robust expression system by providing users with orthogonal control over a gene of interest. In other words, the lack of T7 promoter recognition by host sigma (σ) factors and RNA polymerase (RNAP) prevents leaky expression of genes under its control that may have toxic products or other undesirable consequences. In order to express a gene from the T7 promoter, the T7 polymerase must be integrated into the host chromosome, often in the form of the DE3 prophage under control of the *lac* promoter, permitting induction by the nonnative molecule isopropyl β-D-1-thiogalactopyranoside (IPTG) (Tabor and Richardson 1985). In addition to IPTG induction,

repression of T7 polymerase by T7 lysozyme has been demonstrated, which can be co-expressed to further reduce leaky expression (Moffatt and Studier 1987). The T7 system has been exploited even further to engineer simple genetic circuits with very low basal expression and high responsiveness to inducers (Temme et al. 2012).

Viral polymerases are also highly effective expression tools in eukaryotic hosts. Some recombinant protein expression requires highly specific environments for proper folding and/or complex posttranslational modifications such as disulfide bonds and glycosylation, which can often be more readily accomplished using eukaryotic mammalian cells and plants (Dalton and Barton 2014). In mammalian cells for instance, the Simian virus 40 and cytomegalovirus promoters have been used extensively for constitutive gene expression, typically for recombinant proteins with therapeutic applications (Condreay et al. 1999). Inducible expression can also be accomplished in higher eukaryotes through promoter–regulator systems that respond to the antibiotic tetracycline or the insect hormone ecdysone, for example (Furtht et al. 1994; No et al. 1996). This strategy, which functions both in cell culture and transgenic animals, involves expressing a ligand sensitive transcription factor (TF) and cloning the heterologous gene downstream of a promoter specifically controlled by that TF. Similarly in plants, expression of heterologous genes has been demonstrated using viral promoters as well as tissue-specific promoters (Edwards and Coruzzi 1990; Fütterer et al. 1990).

Utilizing naturally derived genetic parts to drive transcription of heterologous genes is certainly suitable for expressing large quantities of a desired protein or studying gene function, but engineering microbes to carry out complex functions requires a far more diverse set of tools. Accordingly, scientists and engineers alike continuously strive for higher expression levels and tighter control. After thorough investigations into natural systems, many of the actual components and parameters that influence transcription have been elucidated. While comprehending the basic components of transcription is very useful when natural expression systems are implemented, it furthermore enables reengineering of natural systems through combinatorial strategies.

1.2.2 Engineering Controlled Transcription: Mining for Parts

The use of endogenous regulatory systems for engineered transcription can be a very tedious process, as there are often unwanted influences from the natural cell systems. Primarily, cross talk with the cell's own regulatory mechanisms and metabolism can decrease productivity. Secondly, a transcription factor (TF)-operator couple cannot be used to regulate different genes independently (i.e., orthogonally). Independent regulation of several genes simultaneously is of special importance in the context of combining regulated modules into larger systems (Purnick and Weiss 2009). Fortunately, high-throughput sequencing technologies have brought forth an abundance of genomic databases from which new regulatory parts and systems can be mined (Fayyad et al. 1996; Stormo and Tan 2002; Pruitt et al. 2007; Silva-Rocha and de Lorenzo 2008).

Genome mining, the process of searching chromosomal DNA sequences for genetic parts or genes with a desired function, has been used to create libraries of orthogonal σ factors, repressors, and terminators (Rhodius et al. 2013; Chen et al. 2013a; Stanton et al. 2014a). Orthogonal σ factors can enable the host's RNAP to specifically recognize a set of corresponding promoters while not affecting expression of endogenous genes. The expression of such a σ factor may serve as a single control point to govern transcription of multiple heterologous genes. Incorporating inducible expression of a corresponding anti-σ factor can allow threshold-gated switch-like behavior from an engineered transcriptional system (Rhodius et al. 2013).

A typical TF mining workflow consists of first using literature or databases to assemble a library of homologous TFs with similar functions to one that is known (Bateman et al. 2004). Next, operator sites can be determined using in silico or in vitro techniques (Liu et al. 2001; Stanton et al. 2014a). Lastly, all TFs and operators must be screened in vivo for functionality and orthogonality. These libraries can be expanded tremendously by creating hybrids that combine different DNA-binding and effector domains obtained from various mined TFs (Stanton et al. 2014a). Furthermore, the vast library of parts can be expanded by selectively creating mutations in DNA-binding regions (Desai et al. 2009; Temme et al. 2012). A common way for prokaryotes as well as eukaryotes to create efficient new promoters as parts for protein expression with different strength is based on hybrid promoters, described in more detail in the chapter about new DNA and RNA parts.

1.2.3 Tandem Gene Duplication

Classical methods of expressing genes in microorganisms typically rely on high-copy number plasmids to drive ample transcription. While this is often sufficient for small-scale gene expression, it can be problematic due to genetic instability imparted by the metabolic burden associated with hosting multi-copy plasmids and expressing insoluble or toxic proteins. One can never underestimate the rapid genetic drift that often occurs in engineered microorganisms and the propensity for dividing populations of cells to bias for individual genetic variants capable of circumventing expression of heterologous genes. It has been shown that after only 40 generations, a bacterial culture can lose a desired phenotype due to propagation of mutated plasmid DNA, a phenomenon known as allele segregation (Tyo et al. 2009). Integrating genes directly into the chromosome can help solve the problem of allele segregation, but often a single copy does not provide a scientist with sufficient transcription of a gene.

Chemically inducible chromosomal evolution (CIChE—see Fig. 1.2), developed by Tyo et al. (2009), allows for tandem duplication of a chromosomally integrated gene. A synthetic cassette, which contains the gene of interest as well as an antibiotic resistance gene, is integrated into the chromosome, flanked on either side by long homologous regions of DNA. During DNA replication, the endogenous *recA* gene facilitates random homologous recombination between the two

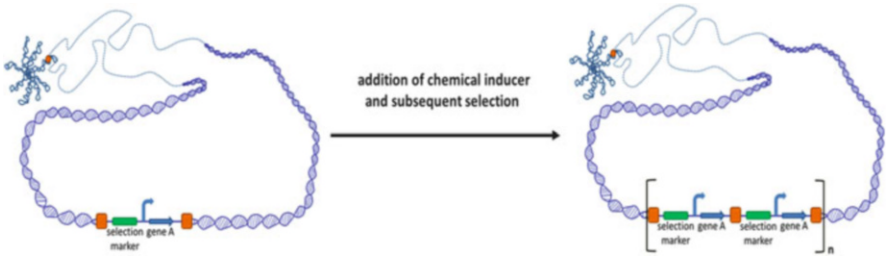

Fig. 1.2 Chemically inducible chromosomal evolution (CIChE). The CIChE DNA cassette contains the gene(s) of interest (*blue*—geneA) and a selectable marker (*green rectangle*), flanked by 1-kb homologous regions (*orange rectangle*). This CIChE cassette is delivered to the chromosome by standard methods. The chromosome is evolved to high gene copy number by addition of a chemical inducer and subsequent selection. As selection pressure increases, i.e., higher concentration of chemical inducer, only cells with many CIChE cassette duplications survive. Iterative tandem CIChE cassette duplication is accomplished by *recA*-mediated DNA crossover between the leading homologous region of one DNA strand and the trailing homologous region in another. The *recA* gene is deleted after the procedure, creating a genetically stable population (Tyo et al. 2009)

daughter DNA strands at homologous sequences. When a recombination event occurs, it results in a deletion in one cell and duplication in another. Cells that undergo duplications of the antibiotic resistance gene along with the gene of interest are selected for by increasing the concentration of the antibiotic, and over several subculturing steps a high-copy number population may be obtained. At the end of the procedure, knocking out *recA* results in a stably integrated high-copy number strain.

This technique has demonstrated its potential by generating stable strains proficient at producing lycopene (Tyo et al. 2009; Chen et al. 2013b), polyhydroxybutyrate (PHB) (Tyo et al. 2009), and shikimic acid (Cui et al. 2014) and has been modified to incorporate use of other selective agents such as triclosan (Chen et al. 2013b; Cui et al. 2014). In theory, any positive selection marker can function in this system as long as the selective compound can be titrated into solution. Alternatively, promoters duplicated in tandem have also been shown to drive stronger gene expression. In one example, up to five tandem copies of the core *tac* promoter were shown to significantly increase production of PHB to 23.7 % of total cell weight (Li et al. 2012b). These strategies are an important step forward towards stably driving heterologous gene expression to high levels.

1.2.4 Decoy Operators Modulate Transcription Factors

While it is convenient to imagine a promoter as being on or off, the reality is that transcription initiation is a stochastic process that depends on the relative abundance of associated TFs. Expression of TFs and the genes they control is temporal

and dynamic, and the relative activity of a TF depends on both its affinity towards a target DNA operator and its intracellular abundance. Due to these inherent properties, it is possible to achieve accelerations and delays in signal transduction using different types of TFs and corresponding operators. When using multiple copies of a regulated promoter, either on plasmids or tandem gene copies, unexpected TF dose–response behavior tends to occur due to an increased relative abundance of operator sequences to (TF) molecules (Brewster et al. 2014). The TF titration effect, which occurs when promoters compete for a limited amount of available TF, complicates predictive modeling and the programming of transcription (Rydenfelt et al. 2014). This effect has also been termed "retroactivity" in the context of genetic circuits, where the connecting of modules via TFs causes a delay in signal propagation analogous to impedance in electronic circuits.

One way of minimizing retroactivity is by overexpressing a TF to make sure that it is always present in excess, which is readily accomplished using inducible expression systems such as those mentioned in Sect. 1.2.1. If one includes a copy of the TF gene on the plasmid itself, every extra copy of the operator site corresponds to an extra copy of its binding TF (Amann et al. 1988; Guzman et al. 1995). While retroactivity appears to convolute TF signal transduction, it is possible to harness the titration effect itself for engineered regulation of transcription. Operators intentionally used to control relative abundances of their TFs are often termed decoys. Decoy operators serve to impede a TF from binding a target operator, while accelerating its dissociation. By using either activators or repressors alongside decoy operators, one can achieve a full spectrum of temporally varied signal transduction (see Fig. 1.3a) (Jayanthi et al. 2013).

Anand et al. (2011) propose "operator buffers," consisting of repeats of passive operator sites, to increase promoter reliability by buffering changes in promoter number. In eukaryotes, similar designs could reduce noise by protecting bound TFs from degradation (Burger et al. 2010). Decoy operators not only stabilize transcription, but also lead to qualitative changes in behavior (see Fig. 1.3b) (Lee and Maheshri 2012). High-affinity decoys convert a graded dose–response to a sharp sigmoidal-like response, while low-affinity decoys shift and broaden the transition, constituting another control knob for the metabolic engineer (Bintu et al. 2005a).

1.2.5 Choose the Gene Location Wisely

Transcription of a chromosomally integrated construct is influenced not only by its promoter and copy number, but also by its location on the chromosome. The chromosomal location can have a significant impact on the transcription of a defined promoter/gene construct that is integrated after having been characterized in another context, such as expression on a plasmid. Spatial patterns of gene expression have been demonstrated in *E. coli* and yeast, where high levels of correlation beyond the operon level are often seen (Képès 2004; Guelzim et al. 2002).

It is thus essential for the genetic engineer to consider optimal chromosomal locations when chromosomally integrating synthetic genes and operons, which

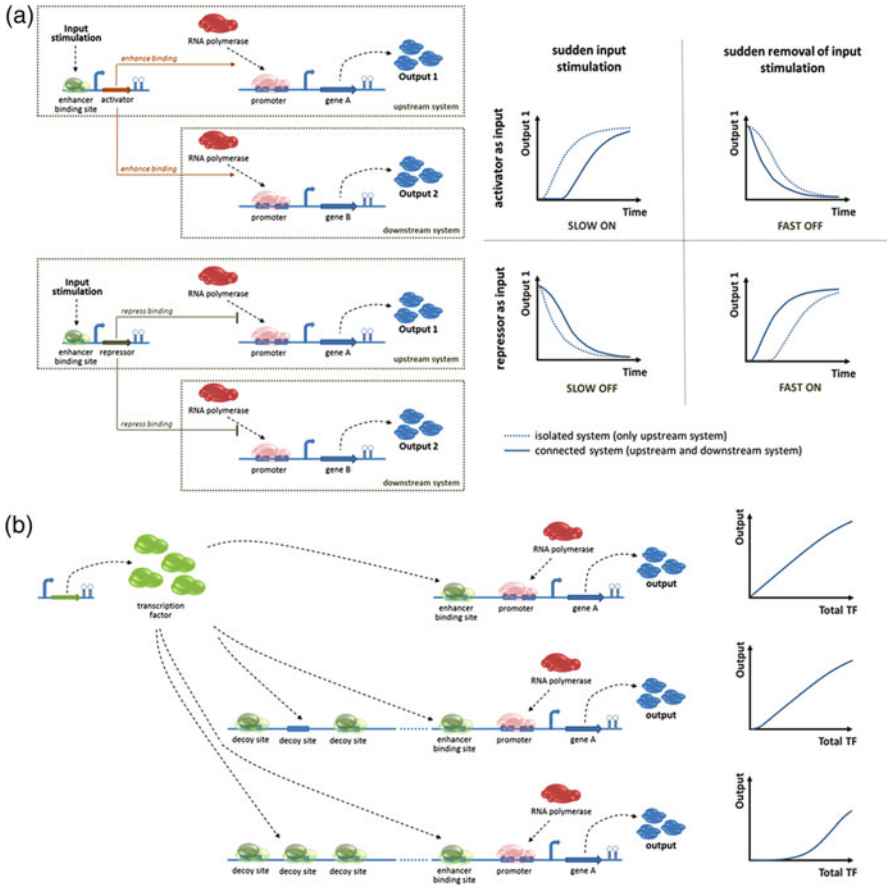

Fig. 1.3 The transcription factor titration effect. (**a**) Retroactivity is the unavoidable back action from a downstream system to an upstream system. The downstream system consumes some of the TFs in order to be expressed. Hence, the TF cannot fully take part in the network of interactions that constitutes the upstream system, resulting in a change of the upstream system behavior. The effect of retroactivity on the response to sudden input stimulation (delay) or to sudden removal of input stimulation (speedup) is shown on the *right*, for both an isolated system and a connected system (adapted from Jayanthi et al. 2013). (**b**) Operator buffer: repetitive stretches of DNA that contain TF binding sites can act as decoys that sequester TFs. These decoy sites can have important indirect effects on transcriptional regulation by altering the dose–response between a TF and its target promoter (depicted on the *right*). *Top construct*: no decoy sites; *middle*: intermediate affinity operators; *bottom*: high affinity decoys (adapted from Lee and Maheshri 2012)

often must be done empirically. As a general strategy, an integration locus is typically centered between two open reading frames (ORFs) that are convergent (Bryant et al. 2014). Design strategies such as incorporating an insulator region upstream of an integrated construct can help prevent many of the unpredictable local variations in gene expression. An effective insulator region often consists of a

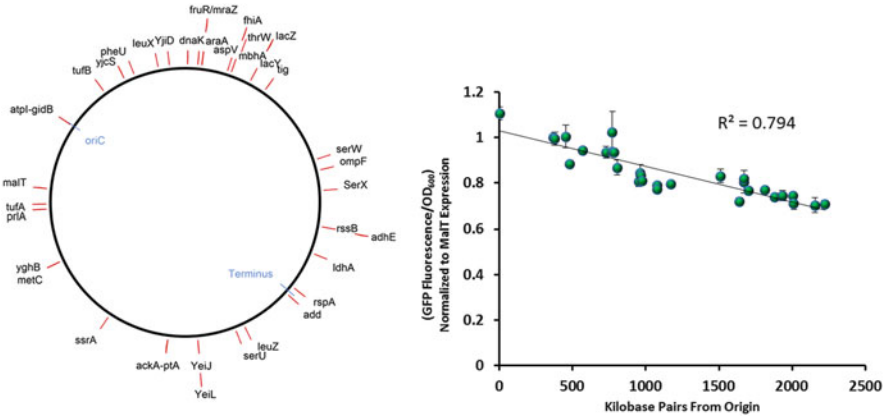

Fig. 1.4 Effect of chromosomal integration site on expression. Spatial distribution of the different tested chromosomal loci (*left*) and their corresponding gene expression as a function of their distance from the origin (*right*)

5′ terminator to prevent adjacent transcription read-through, along with an inert upstream and downstream sequence surrounding the core promoter region (Davis et al. 2011).

On a global level, gene expression in bacteria decreases with distance from the origin of replication (see Fig. 1.4—data collected by Manus Biosynthesis). This phenomenon is a result of an effectively larger copy number for genes closer to the start of DNA replication, which is exaggerated in rapidly dividing populations (Block et al. 2012). Despite this trend, there exist outlying regions where gene transcription is driven by other factors. Expression can vary up to 300-fold with outliers having severalfold higher expression than their closest neighboring genes (Bryant et al. 2014). Transcriptomics in *E. coli* have demonstrated that large genomic regions comprising up to 100 genes correlate in relative expression, which is related to local states of chromatin supercoiling (Jeong et al. 2004). This type of asymmetric expression is important to understand when considering integration of synthetic constructs, as it may have significant impacts on local expression of artificial or native surrounding genes. In addition to chromatin remodeling, local variations in concentrations of TFs can also have an impact on the transcription of genes. Kuhlman and Cox (2012) found the local concentration of the LacI repressor is greater near the inhibitor's locus, and a regulated gene was more strongly inhibited with greater proximity to the repressor gene, similar to the titration effects discussed in Sect. 1.2.4. This information is important to contemplate when designing synthetic regulatory networks as it may offer a finer degree of control over expression.

The nature of transcriptional activation and repression is even more complex in eukaryotic cells. Cis and trans enhancer elements alongside epigenetic remodeling play more complex roles in the dynamic eukaryotic chromosome (West and Fraser 2005; Fraser 2006). In addition, transcription levels can vary significantly between

different chromosomes and regions therein. In yeast, an up to almost ninefold difference was detected between 20 different sites conferring high and low expression of a *lacZ* reporter gene (Flagfeldt et al. 2009). Obtaining such dynamic ranges of gene expression simply based on location provides the genetic engineer with an additional dimension to operate in by modulating gene expression levels while retaining promoter strength and culture conditions.

1.3 Engineering Transcription: Above and Beyond Nature

The preceding sections have given an introduction to some of the various techniques one may use to exploit native genetic elements for rationally engineered systems. While an abundance of natural parts are available for manipulation, they have all evolved in host organisms to provide specific functions, which often have overlapping or conflicting interests with the genetic engineer. The ability to fully circumvent the effects of host background interference in a given expression environment ultimately requires orthogonality through synthetic engineering of custom genetic parts. At the transcriptional level, there is essentially no limit to which parts may be engineered towards rationally targeted functions. DNA stretches ranging from upstream elements and promoters to operators and terminators are frequently modified to generate new functions and optimize existing systems. Furthermore, rationally engineered TFs are becoming routinely fabricated to provide specific operations in a site-dependent manner. This rapidly expanding toolkit enables synthetic biologists and genetic engineers to accomplish what natural systems never required, thus expanding the range of possibilities that life has to offer.

1.3.1 Engineered Promoter Binding

Controlling cellular behavior relies on developing novel means to regulate the transcriptional machinery responsible for the first step in gene expression. This requires a firm understanding of the fundamental architecture comprising bacterial and eukaryotic core promoters, which enables the rational manipulation of existing regulator elements, as well as the synthetic development of new TFs and corresponding recognition sites. A core promoter is typically defined as the minimum contiguous stretch of DNA required to drive transcription initiation (Butler and Kadonaga 2002). Given the essential nature of promoters in this process, they are an attractive target for manipulation due to their ability to affect large consequences downstream.

There are significant differences between bacterial and eukaryotic promoter architecture and thus the mechanisms by which they operate. The bacterial RNAP, consisting of the five subunits $\beta\beta'\alpha_2\omega$, recruits promoter specific σ factors to drive transcription of genes throughout the cell (Browning and Busby 2004). Different σ factors are ultimately responsible for promoter recognition, which is

dictated by the −10 and −35 consensus hexamers upstream of the start site. Initial binding can also be facilitated by UP elements ~20 bp upstream of the −35 consensus sequence (Browning and Busby 2004). Transcription initiation occurs de novo with synthesis of short initiating nucleotides and proceeds after formation of an open complex with the core polymerase and σ factor ejection (Basu et al. 2014).

Eukaryotic transcription primarily differs from bacterial transcription by involving several RNAPs for expression of different classes of RNAs. Of the three main polymerases, RNAP II is responsible for protein synthesis and thus has been widely characterized and is most directly relevant for controlling expression of functional proteins and enzymes (Hahn 2004). RNAP II relies on recruitment of TFs to the core promoter, which is typically comprised of the TATA element (TATA-protein binding element), TFIIB-recognition element, initiator element, and downstream promoter element (Butler and Kadonaga 2002). In conjunction, these elements drive transcription of a downstream gene and in turn provide the foundation for engineering new promoters.

The high degree of control required for successful genetic and metabolic engineering of cells calls for a set of quality tools capable of modulating gene expression over a wide range in a reproducible manner. Early attempts to quantitatively adjust gene transcription included titrating different amounts of inducers such as IPTG with the *lac* operon, but such efforts have proven difficult to reproducibly provide consistent expression of downstream genes. Alternatively, by engineering promoters to have different transcription strengths, one can begin to accurately control transcription and even modularize gene expression of several different enzymes in a pathway at appropriate levels.

Several approaches to modulate transcription initiation rates by promoter engineering have been developed. The bacterial core promoter in particular has been subject to a significant amount of engineering by several groups, as its architecture is well understood. Varying the promoter DNA sequence can be accomplished for example with error-prone PCR (Alper et al. 2005). This technique introduces mutations into the entire promoter sequence, yet the resulting libraries are often outperformed in terms of diversity by libraries created using targeted randomization.

Starting with a consensus promoter of high strength is often ideal, as the engineering process is typically more prone to reducing promoter strength than increasing it. In addition, one can use an exogenous promoter template if a more orthogonal system with high expression is desired (Tyo et al. 2011). This approach has also been successful with mammalian expression systems such as the SV40 viral promoter, where researches have successfully randomized nonessential regions that do not participate directly in TF binding, resulting in a collection of promoters capable of driving high expression over a tenfold relative range (Tornøe et al. 2002). Furthermore, yeast promoter activity can be fine-tuned by specifically manipulating nucleosome disfavoring poly(dA:dT) tracts (Raveh-Sadka et al. 2012).

Characterizing a set of new promoters is easily accomplished by using a reporter such as GFP or luciferase, which can be screened visually or in high-throughput systems such as fluorescence-activated cell sorting (FACS). This allows screening very large diversities, an advantage that can often be necessary when engineering promoters to have activity in new organisms (Yim et al. 2013). Fluorescent reporters reliably correlate differences in transcription strength with a strong measurable signal, but ultimately the level of mRNA transcript itself should be measured using qRT-PCR, for instance, in order to accurately determine promoter strength (Kelly et al. 2009). Nonetheless, reporter-based selection techniques are so powerful for promoter engineering that prokaryotic promoters have been generated from completely random DNA fragments and error-prone PCR. By using a promoter library to drive transcription of an antibiotic resistance gene, one can also enrich the library for strong promoters by using the maximum antibiotic concentration that cells are able to grow in (Alper et al. 2005).

1.3.2 Attenuation: Regulation Through Termination

While non-intuitive, the termination of transcription can act as yet another important regulatory control point. In prokaryotes, termination is triggered by sequences that cause the RNAP to release the template and nascent RNA by means of hairpin formation, or the recruitment of a Rho factor protein that races towards the RNAP (Platt 1986). Libraries of both natural and synthetic terminator sequences of varying strength have been reported and are easily incorporated downstream of a target gene (Chen et al. 2013a) and can also be employed in multiple consecutive copies (Mairhofer et al. 2014).

Liu et al. (2011) used cell lines engineered with an expanded genetic code to harness the phenomenon known from the *trp* attenuator. By engineering the coupled transcription–translation of ORFs with peptide leader sequences containing unnatural codons corresponding to orthogonal tRNAs, they were able to create transcriptional switches, as translation of the leader peptide would only proceed through the orthogonal codons if their corresponding tRNAs were also being expressed.

Ribosome stalling is not the only known attenuator toggle mechanism (Fig. 1.5). Upon ligand binding, upstream RNA aptamers may change in conformation and propagate a response towards an attenuator stem loop affecting its state (Chappell et al. 2013), as can temperature-sensitive conformational changes (Kortmann and Narberhaus 2012). The growing collection of well-characterized aptamers makes for a wide array of small molecule sensors (Lee et al. 2004), and the SELEX[1] technique enables facile in vitro creation of novel aptamers that bind with both high affinity and specificity to virtually any ligand (Ellington and Szostak 1990).

[1] Systematic Evolution of Ligands by EXponential enrichment

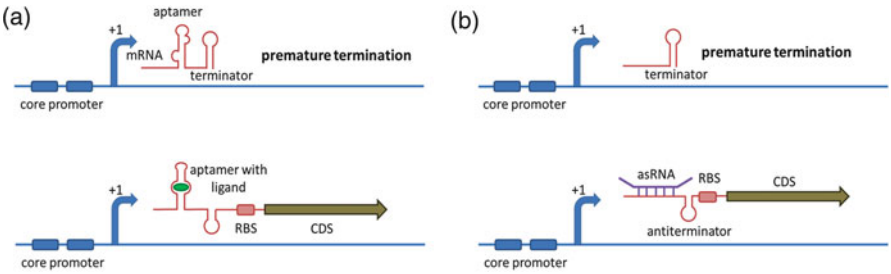

Fig. 1.5 Transcription attenuation. (**a**) Cis attenuation causes changes in the conformation of mRNA based on the binding status of a ligand, resulting in the conditional formation of a termination signal. (**b**) Trans attenuation has similar results, but is the result of a second, non-coding, RNA binding to the mRNA

Wachsmuth et al. (2013) demonstrated this principle in the creation of a synthetic theophylline-sensitive attenuator. Qi et al. (2012) took a different approach to theophylline regulated attenuation by taking advantage of the fact that attenuators can be toggled in trans by an antisense RNA. This property was first discovered in the regulation of plasmid pT181 and has since been exploited for both positive and negative regulation of synthetic constructs (Brantl and Wagner 2002; Dawid et al. 2009). Screening a library of aptamer-pT181-ncRNA fusions also resulted in a synthetic theophylline-responsive transcriptional regulator consisting of nothing but RNA (Qi et al. 2012).

One may find that the available RNA regulatory sequences acting on the initiation of translation outnumber those of the transcriptional type (Burge et al. 2013). However, strategies do exist to make use of translational regulatory elements for the engineering of transcription. One approach is fusing the sensor domains of translational regulators to a library of transcription attenuators and then selecting for attenuators that achieve a desired response in the presence of a given environmental signal (Takahashi and Lucks 2013). In addition, it has been demonstrated that RNA riboregulators responsible for terminating transcription in a Rho-dependent fashion can allow translational riboswitches to halt transcription through the use of an adapter (Liu et al. 2012a; Hollands et al. 2012). This adapter encodes a short leader peptide under control of an upstream translational riboregulator. When translation of the peptide is inhibited due to the upstream riboregulator, Rho factor can attach itself to a site on the nascent RNA that is otherwise occupied by ribosomes and terminate transcription by racing towards the RNAP (Liu et al. 2012b). Several tools exist to aid the engineer in the in silico design of novel RNA molecules (Hofacker 2003; Zuker 2003; Xayaphoummine et al. 2005). The overall balance between the diversity of sequence space and a relatively limited conformational complexity makes RNA an intriguing substrate for the creation of orthogonal transcriptional regulatory systems (Chappell et al. 2013).

1.3.3 Transcription Machinery Engineering

1.3.3.1 Hacking the Polymerase

Cells must naturally balance their production of transcriptional machinery based on environmental cues for growth and maintenance, which often have overlapping and/or conflicting functions when engineering heterologous or even innate biochemistries within an organism. Given that a prokaryotic cell on average holds 2000 molecules of RNAP, which are always subject to fluctuations based on growth phases and physical culture conditions, it is desirable to engineer orthogonal transcription machinery capable of operating independently of the cell's many other physiological needs (Segall-Shapiro et al. 2014). The implementation of functionally relevant regulatory networks requires both tight control and the ability to regulate several different genes independently without cross talk. An underlying issue with controlling biology is that the more complex a synthetic regulatory network becomes, the more difficult it becomes to create a distinct function (Temme et al. 2012). Several groups have sought to expand the current set of tools needed to create novel genetic control systems by introducing orthogonal transcription machinery, which has been most readily accomplished by using viral polymerases and their corresponding promoters to drive transcription of target genes.

The T7 phage RNAP has been used in several cases as a template for engineering orthogonal transcription, as it is a robust polymerase that is orthogonal to the host's enzymes and has been extensively characterized in both prokaryotic and eukaryotic systems (Meyer et al. 2014). Several groups have worked to expand the T7 polymerase–promoter machinery to include novel pairs that can function independent of each other. In one such case, a panel of new orthogonal T7 polymerase promoter pairs was generated through compartmentalized partnered replication. This process involved generation of a mutant library of T7 RNAPs that could drive expression of the Taq DNA polymerase under control of novel T7 promoters inside *E. coli* cells. Next, emulsion PCR of the mutant T7 RNAP genes was performed using the synthesized Taq polymerase, thus linking functionality of a mutant T7 polymerase to the subsequent amplification of the mutant gene (Meyer et al. 2014). Using this method, the authors were able to identify six novel T7 polymerase–promoter pairs through sequential rounds of mutagenesis and selection, which were all capable of specific expression from their cognate promoters. In another example, starting from a T7 RNAP previously selected for reduced burden and toxicity in *E. coli* cells, four novel and orthogonal T7 polymerase–promoter pairs were generated by swapping the promoter-recognition domain of the polymerase with those of other phage polymerases (Temme et al. 2012). The same group went on to fragment T7 RNAP into a β-core and α and σ subunits. Modulating expression of the β-core component effectively acted as a signal amplitude controller capable of tuning up or down input signals imparted by the activation by the α subunit, while output specificity was determined by the σ subunit (Segall-Shapiro et al. 2014).

Other attractive targets for engineering novel synthetic transcription machinery include bacterial σ factors, as they are the primary component in both recognizing a

core promoter and recruiting the RNAP. As an added layer of complexity, anti-σ and anti-anti-σ factors exist to add increased capabilities for cellular responses to changing environmental conditions among other stimuli (Rhodius et al. 2013). As Rhodius et al. (2013) demonstrate, the use of the alternative σ-factor subclass called extracytoplasmic function (ECF) σ-factors allows simplicity of engineering due to their reduced binding domain structure and strong evolutionary conservation. They employed a bioinformatics approach to mine for phylogenetically related σ-factors, which gave rise to 86 ECF σ-factors, 20 of which were highly orthogonal, and anti-σ partners that were used to create effective genetic switches. The above examples represent only a subset of methods to achieve orthogonal biological processes. They are nonetheless important steps forward, as generation of new sets of orthogonal polymerases and other TFs offers synthetic biologists and genetic engineers the tools required to incorporate both distinct and functional regulation inside of living cells.

1.3.3.2 Global Transcription Machinery Engineering

While orthogonal RNAPs are very useful for metabolic engineering, industrial applications often require a complicated genetic engineering approach involving the manipulation of several genes in various metabolic pathways. Typical strategies involve utilizing large-scale *omics* and computational systems biology techniques, combined with targeted protein engineering and synthetic biology manipulations to make specific changes to individual genes (Tyo et al. 2007). These approaches can often limit the maximum desired effect due to the lack of simultaneous changes in the expression of target genes, which is typically limited by construction techniques and screening requirements (Alper and Stephanopoulos 2007). An alternative to engineering specific genes and pathways is to implement combinatorial mutagenesis approaches and/or mutate proteins involved in regulating transcription at the global level. A technique known as global transcription machinery engineering (gTME) seeks to generate phenotypic diversity by mutating key proteins in the transcription process, such as σ factors and RNAP domains (Alper et al. 2006). By manipulating such key components of transcription, one can affect the expression of hundreds of genes simultaneously through mutation of a single protein (see Fig. 1.6).

gTME was first demonstrated by engineering prokaryotic σ factors, the key regulatory proteins involved in targeting the bacterial RNAP towards different promoters. This type of work has been successful in generating novel variants that are capable of tolerating unusual growth conditions and producing more of a desired product. Using error-prone PCR on the *E. coli rpoD* gene encoding the well-characterized σ^{70} factor, variants were selected that were capable of growing under normally detrimental conditions in ethanol, SDS, or both combined (Alper and Stephanopoulos 2007). Utilizing a similar approach, the authors were able to select for a metabolically productive phenotype using the red colored compound lycopene as a target product and demonstrated that a single round of gTME was more effective than several rounds of gene knockout by traditional metabolic engineering methods. Another essential piece of the bacterial RNAP machinery, *rpoA*, which encodes the α subunit often involved in TF recognition, has been targeted by gTME

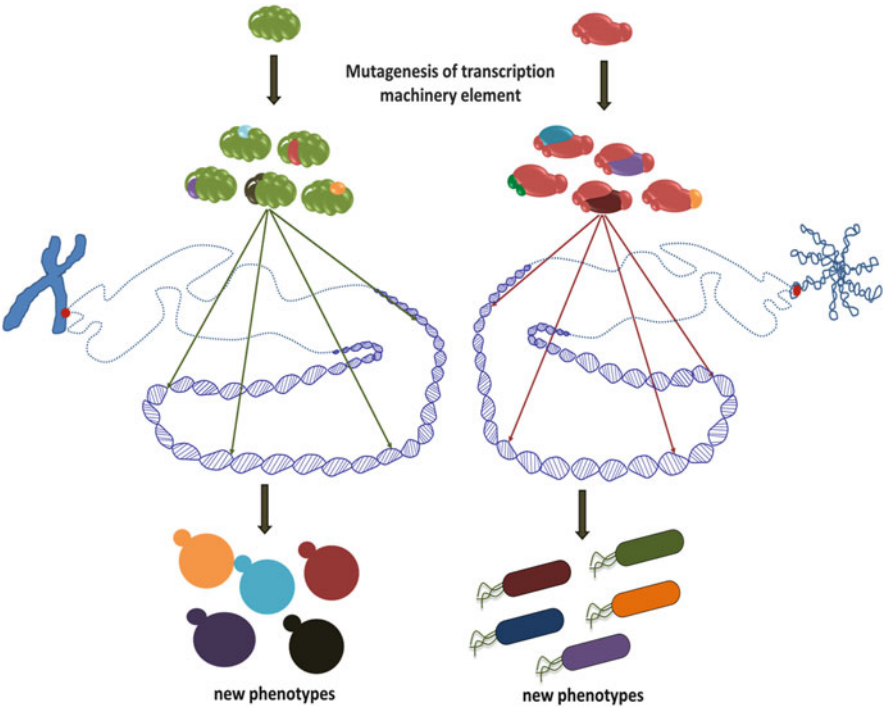

Fig. 1.6 Global transcription machinery engineering. Mutagenesis of a component of the transcription machinery (often in charge of DNA recognition and binding) results in a complete alteration of the global transcriptome (Alper and Stephanopoulos 2007)

giving rise to *E. coli* variants capable of increased tolerance to butanol and hyaluronic acid accumulation.

gTME has also been applied to eukaryotic cells by the same sort of techniques. Given that the eukaryotic RNAPII machinery involves many more TFs, there are even more potential transcriptional regulatory proteins available for targeting by gTME. In one case the yeast *SPT15* gene encoding the TATA-binding protein (TBP) and the TBP-associated protein TAF25 were subjected to random mutagenesis and screened in the presence of high ethanol and glucose concentrations. The study found variants capable of high tolerance for both compounds and observed hundreds of upregulated genes as a result of the mutant TF expression (Alper et al. 2006). Similarly, another group demonstrated that the same SPT15 TBP gene could be diversified to select variants capable of improving the yield of ethanol from *S. cerevisiae* grown on a mixed xylose and glucose sugar substrate (Liu et al. 2010).

The use of gTME to improve upon a rationally designed strain is well exemplified by Santos et al. (2012) through their engineering of *E. coli* for improved L-tyrosine production. Their research began with several gene knockouts and overexpressions to boost flux through the aromatic amino acid pathway, followed by creating random libraries of the RpoA and RpoD RNAP subunits.

Each library was subjected to a high-throughput screen based on tyrosinase enzymatic conversion of L-tyrosine to the dark pigment melanin. This resulted in a maximum increase of 113-fold L-tyrosine production over the rationally derived strain background. This study proved that gTME-induced phenotype variation correlates well with increased mutation rate in a modified unit of transcription machinery, thus allowing a degree of control to the engineer (Santos and Stephanopoulos 2008). While identifying gTME-based mutations is relatively simple, it is more tedious to characterize the change in desired phenotype and corresponding transcriptional profile, which can be accomplished using different *omics* techniques. General metrics such as population growth and pH tolerance divergence have been established in order to determine whether enough phenotypic diversity has been introduced into a library to make it worth a time-consuming screening effort (Klein-Marcuschamer and Stephanopoulos 2010). In summary, while randomized and combinatorial approaches can identify superior strains, they do not replace the need for rational manipulation of target genes and expression thereof and generally can only be effectively applied to strains that are already capable of producing a target compound (Yadav et al. 2012).

1.3.4 Artificial Transcription Factors

A more rational approach to transcriptional engineering has been used to create novel prokaryotic biosensors by exchanging the ligand-binding domain of the *E. coli* LacI TF with domains that detect a different ligand (Meinhardt et al. 2012) and by rewiring classical two-component systems using heterologous sensor kinases (Levskaya et al. 2005; Wang et al. 2013). These designs take advantage of the fact that TFs, especially those found in eukaryotes, tend to be composed of distinct DNA-binding and regulatory domains (Ansari and Mapp 2002). This modular structure has enabled researchers to build chimeric TFs out of various different DNA-binding and regulatory domains. Early examples include a potent eukaryotic transcriptional activator built from the DNA-binding domain of the GAL4 yeast TF and the activating domain of the herpes simplex virus protein VP16 (Sadowski et al. 1988). The human Krüppel-associated box (KRAB), on the other hand, leads to repression when fused to the GAL4 DNA-binding domain (Margolin et al. 1994). When designing hybrid TFs, it is even possible to combine elements from eukaryotes and prokaryotes, as exemplified by the Tet-ON/OFF system (Stanton et al. 2014b). The Tet-OFF module comprises a TetR-VP16 hybrid that strongly activates transcription unless tetracycline or one of its derivatives is present, as these prevent the TF from binding to the DNA (Gossen and Bujard 1992). This tetracycline responsiveness is reversed in the Tet-ON system due to point mutations in the TetR domain that make the synthetic TF require tetracycline for binding to its operator sequence (Gossen et al. 1995). Another class of interesting synthetic sensors can be derived from light-inducible transcriptional effectors (LITEs) that are expressed as separate proteins and bind to their DNA-binding domain only in the presence of light, enabling intensity and spatially controlled transcription (Konermann et al. 2013).

Research into synthetic eukaryotic regulatory domains has yielded activating and repressing peptides, as well as RNA molecules that activate transcription when bound to a TF (Ansari and Mapp 2002). Of special interest are regulatory domains that affect transcription by changing the structure of the chromatin, effectively editing the epigenome (Voigt and Reinberg 2013). For instance, the catalytic domain of the ten-eleven translocation 1 (TET1) protein enhances transcription by reversing methylation at CpG sites close to where the hybrid TF is bound (Maeder et al. 2013b). Contrastingly, lysine-specific demethylase 1 (LSD1) targets histones and represses transcription through methylation and indirectly by deacetylation (Mendenhall et al. 2013). While custom TFs made from natural parts are useful, the full potential of hybrid TFs was unlocked only recently with the development of custom DNA-binding domains. The key enabling technologies are zinc finger proteins (ZFPs), transcription activator-like effectors (TALEs), and clustered regularly interspaced short palindromic repeat-associated proteins (CRISPR/Cas), which will all be discussed in the next three sections. These enable the engineer to effect transcriptional regulation on any sequence at will by designing synthetic TFs in silico, assisted by software packages such as GenoCAD (Purcell et al. 2014) or web tools listed in Table 1.1.

Table 1.1 Software tools that aid in the design of custom DNA-binding domains that show minimal off-target effects

Name	URL	Zn finger	TALE	CRISPR	Ref.
CRISPR design tool	http://crispr.mit.edu			x	Hsu et al. (2013)
CRISPRer	http://bit.ly/CRISPRer			x	Grau et al. (2012)
E-CRISPR	http://www.e-crisp.org			x	Heigwer et al. (2014)
E-TALEN	http://www.e-talen.org		x		Heigwer et al. (2013)
flyCRISPR Target Finder	http://tools.flycrispr.molbio.wisc.edu			x	Gratz et al. (2014)
idTALE	http://idtale.kaust.edu.sa		x		Li et al. (2012a)
Mojo Hand	http://www.talendesign.org		x		Neff et al. (2013)
TAL Effector Nucleotide Targeter	https://tale-nt.cac.cornell.edu		x		Doyle et al. (2012)
TALENoffer	http://bit.ly/TALENoffer		x		Grau et al. (2013)
ZifDB	https://zifdb.msi.umn.edu	x			Fu and Voytas (2013)
ZiFiT Targeter	http://zifit.partners.org/ZiFiT	x	x	x	Sander et al. (2010)

These programs are mostly focused on nuclease targeting in the context of genome engineering, but are also more generally applicable for use with activator or repressor fusions

1.3.4.1 Zinc Finger Proteins

As their name suggests, ZFPs are a unique class of DNA-binding proteins that are able to form site-specific interactions with DNA through zinc-dependent tertiary motifs. First identified in 1982, zinc fingers were initially found through studying TFs required for the expression of 5S RNA genes in oocytes from *Xenopus laevis* (Klug 2010). Initial research revealed these TFs to have conserved 30-bp repeating amino acid motifs, which were found to form loop structures that coordinated zinc ions through direct interactions with two cysteines and two histidine residues, giving rise to the designation Cys_2His_2 (Klug 2010). Zinc finger transcription factors have since been found to be widely abundant regulatory proteins in eukaryotic organisms comprising up to 3 % of the human genome and have offered yet another chassis for engineering gene expression.

The Cys_2His_2 zinc finger motif has been repeatedly used for the construction of novel synthetic TFs due to its modular design. Each finger interacts with a specific three-nucleotide site on the sense strand and one nucleotide on the antisense strand, allowing multiple repeating finger subunits to contribute to increased binding affinity and specificity (Negi et al. 2008). Importantly, zinc finger recognition can occur with single-stranded DNA indicating they are able to bind non-palindromic sequences, thus offering increased design flexibility (Negi et al. 2008). In practice, stringing together three recognition finger motifs in tandem is sufficient for site-specific recognition of only nine corresponding DNA base pairs.

Expression of ZFP TFs can be easily tuned using different types of promoters to achieve the desired magnitude of regulatory effect (Pabo et al. 2001). In a recent study, artificial Cys_2His_2 zinc fingers were used to create 15 transcriptional activators with 2 to 463-fold induction and 15 repressors with 1.3 to 16-fold repression by conjugating leucine zipper or KRAB domains, respectively (Lohmueller et al. 2012). This study achieved control on a variety of simple functions using synthetic zinc fingers in various configurations in mammalian cells. Another innovative use of the Cys_2His_2 motif utilizes light-sensitive proteins from *Arabidopsis thaliana* to create a light-sensitive transcription system. Upon illumination, a ZFP-localized protein heterodimerizes with another protein conjugated to a transcriptional activator, which drives expression of a gene downstream of the ZFP binding sequence (Polstein and Gersbach 2012).

In light of the well-established structural composition of ZFPs, several groups have sought to define the amino acid residue specificity towards DNA base pairs in a predictable manner. Initial experiments with phage display have shown proof of principle in developing novel zinc finger variants by randomizing the α-helical DNA-binding motifs to create diverse libraries, followed by isolation after binding specific DNA ligands (Choo and Isalan 2000a). While somewhat successful, library generation and phage display are limited by screening capacity, as well as binding interference when incorporating preselected DNA-binding domains (Choo and Isalan 2000b). Other efforts have had success in creating limited sized libraries with common in vivo two-hybrid reporter systems, which correlate target DNA binding with transcription of a reporter gene (Hurt et al. 2003). Attempts to generalize a DNA-binding code based on amino acid sequence have had partial

success using the model ZIF268 protein, as different binding conformations and a variety of uncharacterized side chain interactions can convolute predictive models (Wolfe et al. 2000). Some groups have reported successful DNA-binding domain swapping to create novel specific recognition sequences or have engineered extra repeating DNA-binding motifs capable of recognizing up to 64 DNA triplets, resulting in enhanced specificity (Negi et al. 2008). Successful targeting of genomic DNA in mammalian cells requires a minimum six finger motifs for specific recognition, which can be optimized by varying linker length and composition (Papworth et al. 2006). Though tedious and inefficient, one can theoretically design site-specific ZFPs for any DNA sequence with enough randomization and selection of multiple modular repeating DNA-binding domains.

An alternative to rational design is using ZFPs combinatorially in a semi-rational manner. This principle has been demonstrated successfully by generating a large library of the Cys_2His_2 ZFP Zif268 through DNA shuffling of a diverse set of binding motifs, followed by fusion to transcriptional activator or repressor domains. Subsequent expression in *S. cerevisiae* led to the generation of diverse phenotypes including drug resistance, thermotolerance, and osmotolerance (Park et al. 2003). Using the same construction method, thermotolerant phenotypes were selected in *E. coli,* which were traced by chromosome immunoprecipitation to the downregulation of the *ubiX* gene (Park et al. 2005). Ultimately these techniques can lead to increased identification of novel ZFP–DNA interactions, thus expanding the set of characterized modular ZFP domains available for use.

There have been several attempts to develop rational software packages capable of predicting zinc finger arrays that are specific to a given DNA sequence input. One such example is OPEN (Oligomerized Pool ENgineering), which relies on preexisting pools of defined zinc finger DNA-binding domains that have been previously characterized empirically. The software is designed to rationally recombine the domains into three finger recognition arrays giving rise to a relatively small library of variants on the order of 10^5 unique combinations, which can be screened for binding affinity to the target DNA sequence using a bacterial two-hybrid reporter system (Maeder et al. 2008). When compared to a modular assembly method, OPEN ZFP sequences were capable of binding a target sequence with significantly higher affinity (Maeder et al. 2008). While such predictive software packages do not completely remove the screening requirement for novel DNA-binding ZFPs, they do successfully minimize the effort required and thus expedite the process significantly. Given the growing abundance of characterized ZFPs, other tools have been developed to identify existing ZFPs that will bind a given DNA sequence. One prominent example is ZiFiT (Zinc Finger Targeter), which uses a large pool of existing ZFPs that have been well characterized to identify a set of DNA-binding domains suitable for a target region (Sander et al. 2010).

While potentially potent modulators of gene expression, rational design and implementation of Cys_2His_2 zinc fingers requires the creation or assembly of existing domains followed by evaluation in a desired contextual format. Unfortunately, the relatively low success rate for rationally designed zinc fingers makes the

generation of a cross functional modular set of recognition domains challenging (Sera 2009). Despite the laborious construction and screening process required to generate new ZFPs, there has been much success reported in specific contexts as outlined here, and continued research to address these shortcomings will transform this versatile class of TFs to a widespread and robust tool.

1.3.4.2 A Tale of Transcription Activator-Like Effectors (TALEs): Adversaries Turned Allies

Recent research into host–pathogen interactions between pathogenic *Xanthomonas* bacterial species and plants has identified a new class of TFs that have evolved a mechanism to steer host gene expression towards hypertrophic phenotypes (Marois et al. 2002). To accomplish this, the bacterium injects transcription activator-like effector (TALE) proteins into plant cells. A nuclear localization sequence then guides the TALE into the nucleus, where the protein's DNA-binding domain specifically binds to its cognate target sequence. The C-terminal domain of the TALE can then activate transcription of downstream target genes, creating a more suitable environment for bacterial colonization (de Lange et al. 2014).

TALE DNA-binding domains consist of a set of tandem repeats, each encoding a single hairpin structure of approximately 19 amino acids, which collectively form a superhelix tracking a DNA sense strand. In contrast to zinc fingers, every hairpin structure contacts exactly one nucleobase, the identity of which is determined by two amino acid residues at the tip of the hairpin (Moscou and Bogdanove 2009; Boch et al. 2009). Decrypting this code has enabled researchers to target any sequence through a set of approximately 16–24 tandem repeats. It was also quickly discovered that a nuclease domain could be fused to a truncated TALE, allowing them to be used for genome editing techniques (Miller et al. 2011).

Similar to fused nuclease constructs, a transcriptional engineer can employ custom TALE domains to activate transcription in plants (Morbitzer et al. 2010), as well as prokaryotic and mammalian cells using elements that interact with RNAPs, such as VP16/64 transcriptional activators (Zhang et al. 2011; Geissler et al. 2011; Tsuji et al. 2013). Activation can be further amplified by targeting multiple upstream sites of the same gene simultaneously (Perez-Pinera et al. 2013b; Maeder et al. 2013c). Using a similar strategy, TALE repressors have been created using the SRDX domain in plants (Mahfouz et al. 2012) and SID or KRAB domains in mammalian cells (Cong et al. 2012; Garg et al. 2012) and by simply binding to the core promoter in bacteria and yeast (Blount et al. 2012; Politz et al. 2013). Furthermore, ligand-dependent TALEs have been created by inserting one or more ligand receptors in between the DNA-binding and regulatory domains. Activity of these TFs requires a conformational change within the receptor region that is triggered by binding of the ligand (Mercer et al. 2014).

To overcome any context-dependent binding issues, in silico tools such as those listed in Table 1.1 aid engineers in the selection of a target sequence and design of TALE DNA-binding domains (Liu et al. 2014). Some sequence restrictions have been lessened through protein engineering (Tsuji et al. 2013), and ambiguous recognition can actually be exploited to target multiple loci with one TALE