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

1 Programming Biology: Expanding the Toolset for the Engineering of Transcription 1
Bob Van Hove, Aaron M. Love, Parayil Kumaran Ajikumar, and Marjan De Mey

2 Novel DNA and RNA Elements 65
Julia Pitzer, Bob Van Hove, Aaron M. Love, Parayil Kumaran Ajikumar, Marjan De Mey, and Anton Glieder

3 Key Methods for Synthetic Biology: Genome Engineering and DNA Assembly 101
Astrid Weninger, Manuela Killinger, and Thomas Vogl

4 Protein Building Blocks and the Expansion of the Genetic Code 143
Birgit Wiltschi

5 Synthetic Biology for Cellular Remodelling to Elicit Industrially Relevant Microbial Phenotypes 211
Paola Branduardi

6 Microbial Platform Cells for Synthetic Biology 229
Dong-Woo Lee and Sang Jun Lee

7 Synthetic Biology Assisting Metabolic Pathway Engineering 255
Hans Marx, Stefan Pflügl, Diethard Mattanovich, and Michael Sauer

8 Molecular Modeling and Its Applications in Protein Engineering . . . 281
Emel Timucin and O. Ugur Sezerman

9 Synthetic Biopolymers 307
Christian P. Kubicek

10 Xenobiotic Life 337
Dario Cecchi and Sheref S. Mansy

Index 359

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

1.1 Introduction 2
1.2 Reengineering Natural Systems for New Applications 4
1.2.1 The Beginnings 4
1.2.2 Engineering Controlled Transcription: Mining for Parts 5
1.2.3 Tandem Gene Duplication 6
1.2.4 Decoy Operators Modulate Transcription Factors 7
1.2.5 Choose the Gene Location Wisely 8
1.3 Engineering Transcription: Above and Beyond Nature 11
1.3.1 Engineered Promoter Binding 11
1.3.2 Attenuation: Regulation Through Termination 13
1.3.3 Transcription Machinery Engineering 15
1.3.4 Artificial Transcription Factors 18
1.4 Complex Behavior Through Genetic Circuits 24
1.4.1 Biosensors Provide Circuit Inputs 26
1.4.2 Boole Meets Biology: Genetic Logic Gates 27
1.4.3 Towards Building a Biochemical Computer 31
1.4.4 Design Principles 36
1.4.5 Caveats and Perspectives 42
1.5 Transcription Engineering for New Advances in the Fields of Medicine and Industrial Biotechnology 43
1.5.1 Transcriptional Engineering in Medicine 43
1.5.2 Industrial Applications: Synthetic Biology Meets Metabolic Engineering 47
1.6 Outlook 49
References 49

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

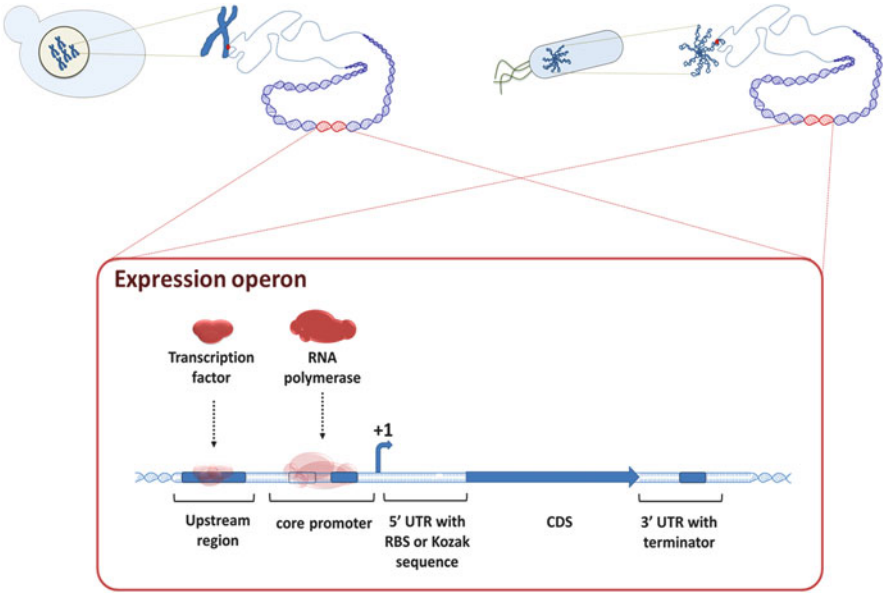


Fig. 1.1 Schematic representation of gene expression and the various components involved in the process of transcription. The central dogma of molecular biology states that DNA is transcribed to messenger RNA (mRNA), which is in turn translated to protein. Transcription is initiated by binding of the 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

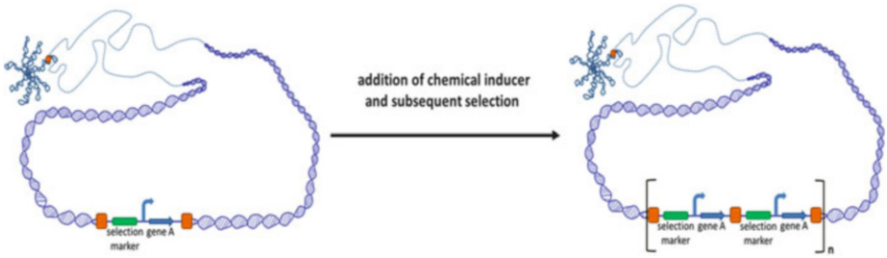


Fig. 1.2 Chemically inducible chromosomal evolution (CIChE). The CIChE DNA cassette contains the gene(s) of interest (*blue*—geneA) and a selectable marker (*green rectangle*), flanked by 1-kb homologous regions (*orange rectangle*). This CIChE cassette is delivered to the chromosome by standard methods. 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

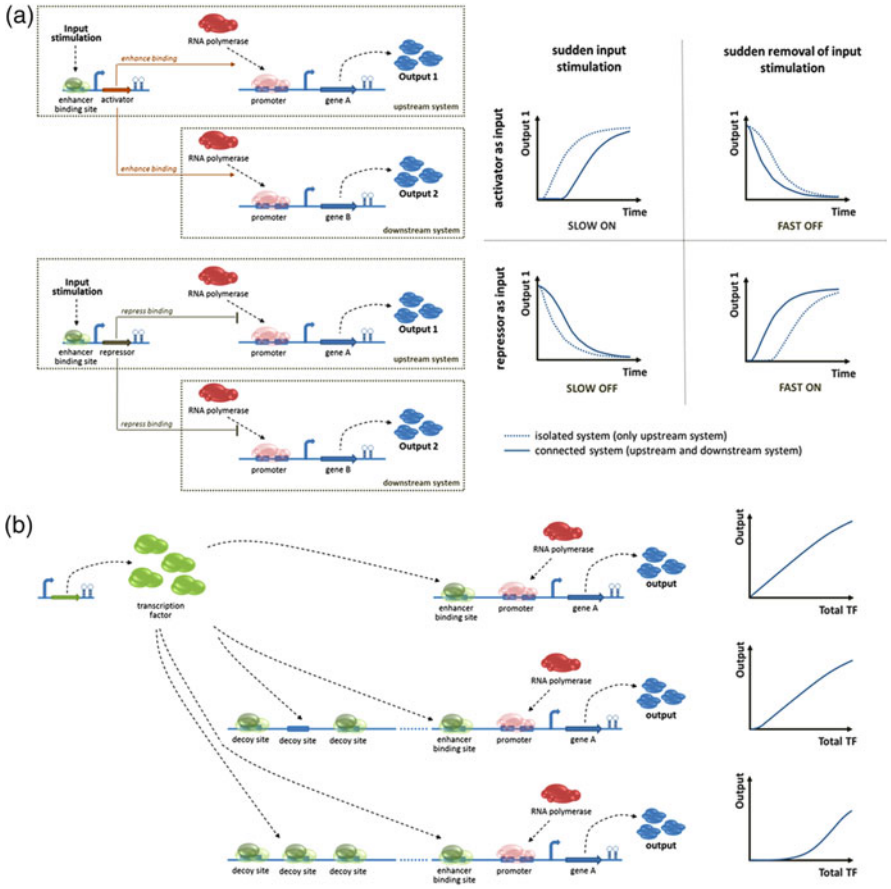


Fig. 1.3 The transcription factor titration effect. (**a**) Retroactivity is the unavoidable back action from a downstream system to an upstream system. The downstream system consumes some of the TFs in order to be expressed. Hence, the TF cannot fully take part in the network of interactions that constitutes 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

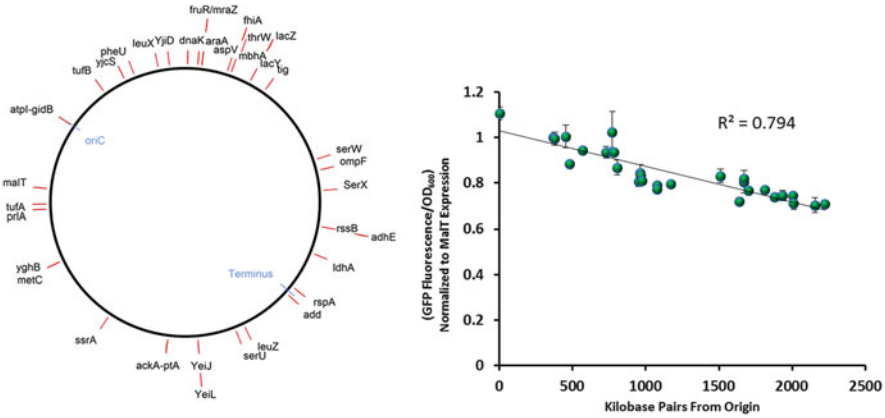


Fig. 1.4 Effect of chromosomal integration site on expression. Spatial distribution of the different tested chromosomal loci (*left*) and their corresponding gene expression as a function of their distance from the origin (*right*)

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

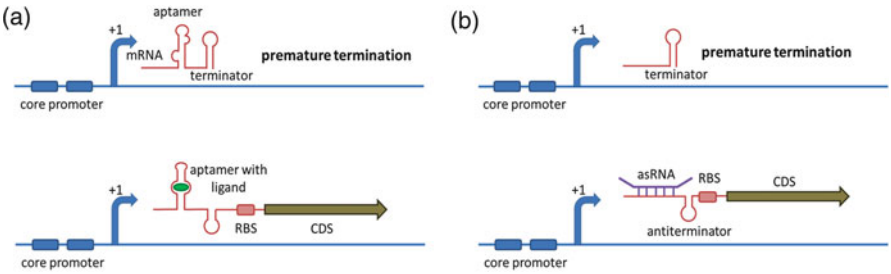


Fig. 1.5 Transcription attenuation. (**a**) Cis attenuation causes changes in the conformation of mRNA based on the binding status of a ligand, resulting in the conditional formation of a termination signal. (**b**) Trans attenuation has similar results, but is the result of a second, non-coding, RNA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

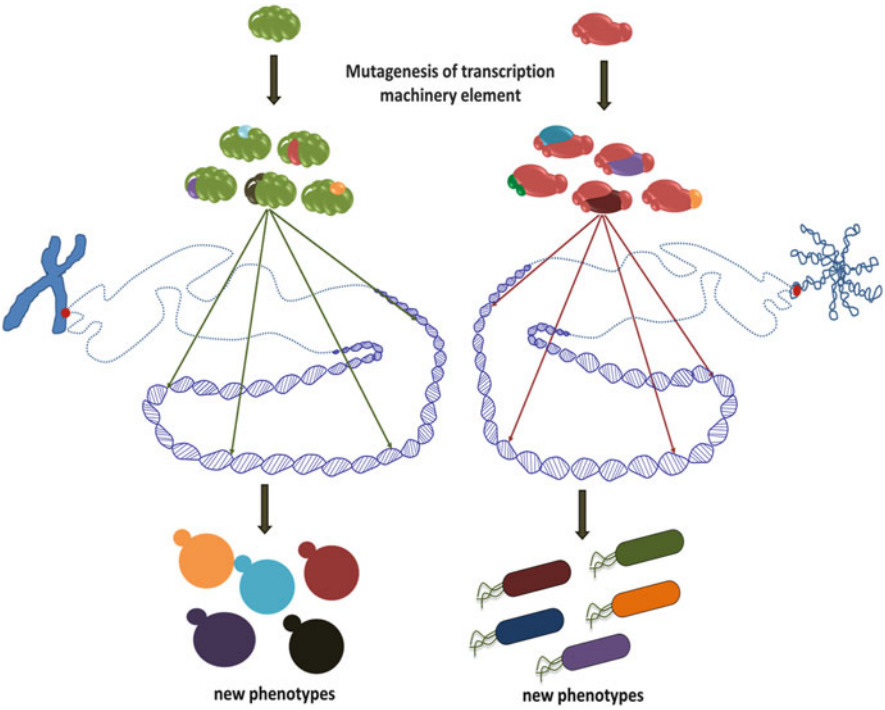


Fig. 1.6 Global transcription machinery engineering. Mutagenesis of a component of the transcription machinery (often in charge of DNA recognition and binding) results in a complete alteration of the global transcriptome (Alper and Stephanopoulos 2007)

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

(Aouida et al. 2014). Molecular cloning of TALE domains can be challenging due to their tandem repeated sequences, but techniques such as iterative Restriction Enzyme And Ligation (REAL) cloning (Sander et al. 2010), Golden Gate assembly (Weber et al. 2011; Cermak et al. 2011), Iterative Capped Assembly (ICA) (Briggs et al. 2012), Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) cloning (Reyon et al. 2012), and Ligation-Independent Cloning (LIC) (Schmid-Burgk et al. 2013) have been optimized for use in the creation of designer TALE domains. A set of orthogonal designer TALE repressors and activators gives transcriptional engineers another resource to create large-scale synthetic gene networks that operate independently from the host's regulatory systems (Blount et al. 2012; Garg et al. 2012).

1.3.4.3 RNA-Guided DNA Binding with CRISPR Technology

As viruses make up the majority of aquatic biomass on earth (Bergh et al. 1989), microbial organisms have evolved interesting mechanisms to combat foreign invaders. Aside from innate defenses such as restriction enzymes that digest alien DNA, bacteria and archaea have also evolved a recently discovered adaptive immune system in the form of genomic clustered regularly interspaced short palindromic repeats (CRISPR). Through the function of CRISPR-associated (*cas*) genes, an infected cell can integrate parts of a viral genome into its CRISPR loci as protospacers, which form a chronicle of previously encountered viruses. When surviving daughter cells are challenged by the same virus, the invading genetic material is recognized through Watson–Crick base pairing of short CRISPR transcripts (crRNA), which target Cas degradation machinery to the foreign DNA by various mechanisms (Barrangou et al. 2007; Sorek et al. 2013). In *Streptococcus pyogenes*, for example, foreign DNA is cleaved by the CRISPR-associated protein Cas9, guided by fragments of crRNA after processing by RNase III and a trans-acting crRNA (tracrRNA) (Deltcheva et al. 2011). Owing to its simplicity and robust characterization, the CRISPR/Cas9 system has been extensively used for practical applications.

Jinek et al. (2012) demonstrated the engineering potential of CRISPR systems by showing that Cas9 can be programmed to target any DNA sequence through the expression of a custom guide RNA (gRNA), requiring only a CC dinucleotide one base adjacent to the target, i.e., a protospacer adjacent motif (PAM). The gRNA was engineered as a crRNA:tracrRNA hybrid, which does not require processing by additional Cas proteins. This allowed researchers to begin employing CRISPR/Cas9 systems for RNA-guided human genome engineering (Cong et al. 2013; Mali et al. 2013b). Bacterial CRISPR/Cas9 RNA-guided targeting is functional in organisms across all domains of life, as long as proper nuclear localization sequences are included. Applications are not only limited to genome editing, as a Cas9 mutant that is incapable of cutting DNA (dCas9, a.k.a. Cas9m) still binds to the target sequence, where it acts as a steric inhibitor of RNAPs. (Qi et al. 2013; Bikard et al. 2013).

CRISPR/dCas9 gene knockdown, also known as CRISPR interference (CRISPRi), can be up to 99.9 % efficient when the gRNA is correctly designed.

The 20-bp recognition sequence of a gRNA should target sequences adjacent to a PAM on either the template strand of the target core promoter region and associated TF binding sites or on the non-template strand of the 5′ region of the transcribed sequence. Care must be taken to ensure gRNAs fold correctly and do not cross-react with off-target sites on the host genome. Increased repression can be obtained, especially in eukaryotes, by simultaneously targeting multiple non-overlapping sites in the same target gene. Contrastingly, a diminished effect can be obtained by incorporating base-pairing mismatches in the gRNA (Qi et al. 2013). It is worth noting that by expressing a gRNA with mismatches, wild-type Cas9 is unable to cleave targeted DNA, and an efficient knockdown is obtained instead (Bikard et al. 2013). A detailed workflow for the design and cloning of effective gRNA constructs has been given by Larson et al. (Larson et al. 2013).

Similarly to other artificial TFs, the dCas9 protein has been further functionalized for enhanced repression or activation properties. This potential has been well exemplified in eukaryotes by fusing regulatory components such as the VP64 activation domain to the dCas9 C-terminus (Maeder et al. 2013a; Perez-Pinera et al. 2013a; Gilbert et al. 2013; Cheng et al. 2013; Farzadfard et al. 2013), or the RNA binding MS2 bacteriophage coat protein (Mali et al. 2013a). In the latter case, the RNA-binding protein's recognition site is added to the 3′ end of the gRNA. Activation domains tethered to the gRNA are less effective compared to dCas9 fusions, but they do enable researchers to use the dCas9 protein for both activation and repression in the same cell by changing or omitting binding sites on the gRNA. CRISPR/dCas9 transcriptional activation in bacteria is less common, but can be performed by fusing the omega subunit of the bacterial RNAP to dCas9 (Bikard et al. 2013). CRISPR/dCas9 repression in eukaryotes was improved by expressing the KRAB and Mxi1 repressor domains as protein fusions (Gilbert et al. 2013; Farzadfard et al. 2013). Although similar fusion proteins have been described for both ZFPs and TALE domains, it is much easier to target Cas9 fusions to new sequences, i.e., exchanging a 20-bp stretch in the gRNA. Herein lies the main advantage CRISPR systems have over previous technology and is the reason the technology has been received so well by the scientific community, and has quickly accelerated in use (Copeland et al. 2014). An overview of the significant attributes of ZFPs, TALEs, and CRISPR technologies is outlined in Table 1.2.

1.4 Complex Behavior Through Genetic Circuits

While the ability to command the expression of a single gene is impressive, even more exciting is the possibility of constructing networks of interconnected genes with complex regulation. Synthetic biologists focus on creating new molecular tools capable of altering gene expression and study the ways in which these components can be assembled into networks that respond to changing cellular environments thereby adjusting gene expression accordingly. Metabolic engineers look at a cell not as a collection of macromolecules, but as a highly efficient chemical factory that transforms raw materials into high-value products. Enzymes

Table 1.2 Comparison of zinc fingers, TALE, and CRISPR

Molecular design					
TF	*DNA recognition*	*Module size*	*DNA bases recognized/ module*	*Modules required*	*Sequence biases*
ZFPs (Zif28)	Repeating Cys_2His_2 motif	30–40 amino acids ~40 kDa	9–12	3–4	GNN triplet
TALE	Repeating TALE motif	19 amino acids ~105Kda	1	16–24	5′ thiamine target
CRISPR/ Cas9	Guide RNA	Cas9 ~160 kDa +1 Guide RNA	20	1	PAM
Properties					
TF	*Design flexibility*	*Engineering time*	*Modularity*	*Specificity*	*Historical validation*
Zinc finger Zif28	High	High	Low	Medium	High
TALE	High	Medium	Medium	High	Medium
CRISPR/ dCas9	Medium	Low	High	Medium	Low
Reported issues					
ZFPs (Zif28)	Irreproducible and not all triplets have aa fingers validated, bias for GNN triplets				
TALE	Difficulty cloning tandem arrays				
CRISPR/ Cas9	Each module restricted to 20 bp and inability to use activating/repressing dCas9 conjugates in parallel				

Each of the synthetic TFs discussed in Sects. 1.3.4.1–1.3.4.3 is examined in terms of physical characteristics, as well as their relative benefits, drawbacks, and unique facets

can be thought of as machines that perform unit operations, which transfer metabolites to one another in a network comprising a pathway. In a nutshell, metabolic engineering concerns the flux of metabolites through the system to generate a product, while synthetic biology concerns the flow of information. This abstraction justifies a new model for biological engineering, specifically a departure from the view of cells as factories, in favor of understanding them as circuits of interconnected components that constantly shuffle data back and forth.

The idea that genetic regulatory networks are comparable to electronic circuits is not new and was in fact proposed by Sugita as early as 1963 (Sugita 1963). Seeking to expand upon Jacob and Monod's (1961) seminal work on inducible and repressible expression, and fueled by the cybernetic movement that took flight in the 1940s (Apter 1966), he designed and built digital circuit models of hypothetical genetic networks. These genetic designs included the bistable switches, oscillators, and logic gates that decades later would form the foundation of genetic circuit engineering (Hasty et al. 2002b). This concept was revisited a number of times in the ensuing years (Kauffman 1969; Thomas 1973, 1991; McAdams and Shapiro 1995;

Weiss et al. 2002), but only in the last decade with the advancements of genomics and synthetic biology culminated the translation of theory into practice (McAdams and Arkin 2000; Purnick and Weiss 2009).

1.4.1 Biosensors Provide Circuit Inputs

Fundamentally, a genetic circuit can receive input signals either from other circuits or from the detection of biochemical and physical changes. The binding or release of a TF activates a promoter, after which, through the processes of transcription and translation, the signal propagates into the designed circuit (de Las Heras et al. 2010). Genetic circuit signals generally take the form of macromolecules such as protein and RNA, but can also be small molecules like N-acyl homoserine lactone for intercellular communication (Salis et al. 2009).

Engineers have an abundance of TFs at their disposal for use as highly specific and sensitive biosensors (Michener et al. 2012). The DNA-binding properties of a sensor TF change in the presence of a chemical ligand such as IPTG as discussed before or in response to physical environmental conditions, such as the temperature-sensitive cI857 repressor (Remaut et al. 1981) and light-sensitive EL222 LOV domain (Nash et al. 2012). Extracellular stimuli can be detected using two-component systems consisting of a membrane-bound sensor kinase and a cytoplasmic response regulator (Salis et al. 2009). Stimulation of the sensor kinase triggers a phosphorylation cascade resulting in activation of the response regulator TF. Sensor kinases can also be semi-rationally engineered to either interact with a different response regulator or to recognize another ligand (Looger et al. 2003; Salis et al. 2009).

Three-hybrid systems, best known from yeast, provide another example of rationally engineered novel chemical sensors. A TF is split into its DNA-binding domain and transcription-activating domain. Next, each domain is fused to a separate protein that binds strongly to a ligand, i.e., the compound that the sensor responds to. Consequently, a functional TF complex capable of both binding to the DNA and activating transcription is assembled only when the ligand is present and bound by both components (Baker et al. 2002). An example of a two-hybrid sensor system is a split TF in which a “prey” and “bait” domain bind to each other in the presence of light (Shimizu-Sato et al. 2002). Light sensing has been demonstrated by several other groups and exhibits the exciting potential of combining biosensors from nature with human ingenuity (Levskaya et al. 2005; Tabor et al. 2012; Ohlendorf et al. 2012; Schmidl et al. 2014).

More complex conditions such as culture density or damage to the chromosome can be detected by interfacing an engineered circuit to a cell’s natural quorum sensing or SOS response networks (Kobayashi et al. 2004). Synthetic DNA-binding domains such as those discussed before greatly improve the applicability of biological sensors by allowing scientists to target virtually any desired promoter. In addition, tools including the translation–transcription adapter construct mentioned in Sect. 1.3.2 (Liu et al. 2012a) makes a plethora of RNA-based sensors [not

discussed here but reviewed in Serganov and Nudler (2013)] usable for the regulation of transcription. The abundance and diversity of reported biosensors are fundamental tools enabling new innovative genetic circuit design.

1.4.2 Boole Meets Biology: Genetic Logic Gates

Transcriptional genetic circuits contain input modules that convert stimuli into signals that are fed into the circuit in the form of active TF molecules. The relationship between the concentration of a TF and the strength of the stimulus can be quantified by a transfer function (Brophy and Voigt 2014). Even though TF concentration is a continuously dynamic variable, it can be represented more simply as being "low" (nM range) or "high" (μM range). This type of binary classification provides an analogous situation to that of electronic circuits (Buchler et al. 2003). In a digital circuit built from interconnected transistors, for instance, a voltage of less than 0.8 V is considered "low," while 2.0 V and up is "high." Constraining variables to just two values, i.e., logical "false" (0) and "true" (1), permits the use of Boolean algebra, which is ideally suited for the analysis and design of circuits (Boole 1854; Shannon 1938).

Logic gates are the real-world implementation of Boolean functions and form the basis of any electronic or genetic circuit. They yield a single output signal after performing algebraic operations on one or more inputs. The most straightforward logic gates are the buffer and inverter. The buffer outputs a logical 1 if and only if it receives a 1, while the inverter, which is more commonly known as a NOT gate, outputs 1 if and only if the input signal is 0. In their most basic form, logic gates correspond to promoters that require one or more input TFs for activation or repression (de Las Heras et al. 2010). In essence, these types of gates can be actuated by any of the transcriptional signals described earlier in this chapter. In contrast to electrical circuits, there is no easy way to physically insulate different signal carriers inside of a cell, so orthogonality is vital to prevent cross talk. Fortunately, as outlined in the previous sections, a wide range of mined, synthetic, and hybrid TFs exists, as well as orthogonal σ-factors and even entire RNAPs (Temme et al. 2012; Rhodius et al. 2013; Stanton et al. 2014a).

Figure 1.7 displays the logic symbols corresponding to the buffer and NOT gates, as well as log–log plots of their respective transfer functions. The x-axis indicates the concentration of the input TF, and the y-axis shows the relative fold change of the promoter output signal. The fold change is calculated as the ratio of the probability that an active RNAP is bound to the promoter in the presence of an input TF, compared to the situation without a TF (Bintu et al. 2005a, b).

These curves demonstrate a number of key parameters that describe a gate's performance: the dynamic range, threshold, and sensitivity. Dynamic range is a measure of the difference in output intensity between the OFF and ON states. A broad dynamic range is required for efficient stimulation of downstream logic gates in a circuit. The threshold is defined as the input value where the output reaches half of its maximal value. This point must fall within the dynamic range of upstream

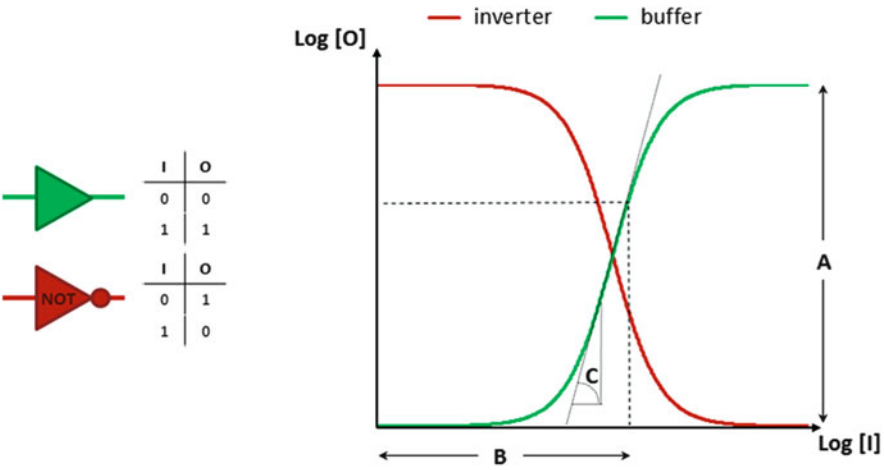


Fig. 1.7 The buffer and NOT gate. Logic symbol and truth table of the buffer and NOT gate (*left*). Transfer curves of the buffer and NOT gate. A buffer gate displays a high output if and only if its input is also high; a NOT gate is the inverse. Gates are characterized by their (A) dynamic range, (B) threshold input at which the output reaches the half of its maximum, and (C) sensitivity or degree of cooperativity, displayed as the slope on the log–log plot (right). I = input; O = output (Brophy and Voigt 2014)

logic gates connected to an input (Brophy and Voigt 2014). Finally, the sensitivity, also called the degree of cooperativity, describes the shape of the response curve.

In the most simplistic case, a transfer function follows a hyperbolic relationship similar to Michaelis–Menten kinetics. In reality, it is often the case that a small percentage increase in an input signal is amplified into a larger percentage increase of the output (Zhang et al. 2013). This phenomenon, termed the ultra-sensitive response, is brought by cooperative binding of the TF to the operator. Cooperativity often occurs when a TF's affinity towards its target operator is higher once a TF homo-multimer has formed or when multiple tandem operators recruit TFs to a promoter region as discussed in Sect. 1.2.4. Bintu et al. (2005a, b) provide an excellent biochemical and mathematical explanation of these and other mechanisms.

While single input gates serve as a model to introduce key concepts of genetic circuitry, implementing complex algorithms requires the use of gates with multiple inputs. Since every added input doubles the total number of possible input states, the set of possible logic operations is expanded as well. This concept is demonstrated by the AND logic gate, which is outlined with symbols and truth tables in Fig. 1.8. This gate displays a high output signal if and only if both inputs are high. Genetic circuit engineers are not required to create every type of gate biochemically, because the logical operation of one gate can readily be emulated by a combination of gates. For instance, a NAND gate can be created by inverting the output of an AND gate using a NOT gate. Moreover, a combination of either

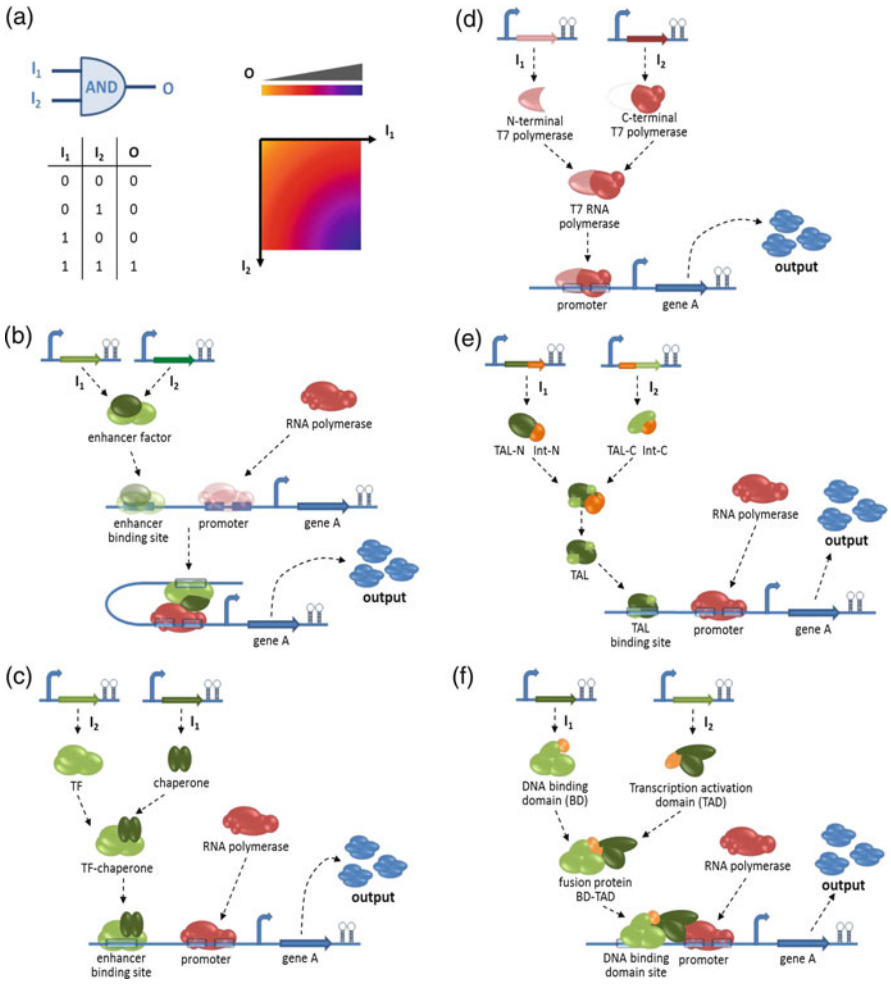


Fig. 1.8 The AND gate. (**a**) Logic symbol, truth table, and transfer function of an AND gate. The two inputs are designated I_1 and I_2. The value of output O is displayed as color intensity (simulated data). Schematic mechanism of an AND gate implementation using a (**b**) heterodimeric TF, (**c**) TF–chaperone pair, (**d**) split T7 polymerase, (**e**) intein joined split TALE, and (**f**) two-hybrid system. (Schematic mechanisms adapted from Kramer et al. 2005; Guido et al. 2006; Moon et al. 2012; Shis and Bennett 2013; Lienert et al. 2013)

NAND or NOR gates can be used to recreate every possible logic gate (Ran et al. 2012). The reader is refered to the works of Sheffer (1913) for a mathematical proof of this statement.

As previously indicated, there are several physical mechanisms by which one can create gates with the same logic. Continuing with the AND example, a hybrid promoter with both LacI and TetR operators constitutes a simple AND gate that responds to IPTG and anhydrotetracycline (Cox et al. 2007). This type of

construction does not constitute a true transcriptional gate, as the inputs are small molecules rather than the products of two genes under control of different operators. In contrast, promoters of the σ-54 type have been used to create true transcriptional AND gates (Fig. 1.8b). For instance, the *Pseudomonas syringae hrpL* output promoter requires the presence of two cooperatively binding TFs named HrpR and HrpS. The first input promoter controls *hrpR* expression, while the second drives *hrpS* (Wang et al. 2011). A transcriptional AND gate can also be created using an input promoter to drive expression of a transcriptional activator that requires the presence of a specific chaperone protein, which is controlled by a second input promoter (Fig. 1.8c). InvF and SicA, derived from the *Salmonella* Pathogenicity Island 1 (SPI-1), form one such TF–chaperone pair, and several orthologs are available through genomic mining (Moon et al. 2012).

A different type of AND gate utilizes split orthogonal RNAP mutants that require the expression of each domain for function (Shis and Bennett 2013; Schaerli et al. 2014) (Fig. 1.8d). The "resource allocator" outlined in Sect. 1.3.3.1 further expands on this type of mechanism by fragmenting the T7 RNAP to make a promoter-recognition (alpha) domain interchangeable (Segall-Shapiro et al. 2014). By expressing competing alpha subunits, the output signal can be redirected to different targets or dampened in the case of a nonfunctional alpha subunit. In another example AND gate, one input promoter drives the transcription of a gene encoding the T7 RNAP, while the second promoter expresses an RNA molecule required as a cofactor for translation of the first mRNA (Anderson et al. 2007). Other example AND gates include two- or three-hybrid systems (Bronson et al. 2008), split ZFP activators (Lohmueller et al. 2012), and split TALE activators (Fig. 1.8e) (Lienert et al. 2013). The AND gates described here are easily converted to NAND gates by connecting the output promoter to a repressor TF.

In mammalian cells, researchers have taken advantage of the fact that eukaryotic promoters are more often regulated from a distance and by multiple TFs. In such cases, a promoter controlled by two activating TFs becomes an OR gate when each TF is controlled by a separate input promoter (Fig. 1.8f) (Kramer et al. 2005). Hybrid activators created using synthetic ZFP–DNA-binding domains described in Sect. 1.3.4.1 have been used to create fully orthogonal logic gates of this type (Lohmueller et al. 2012). NOR gates can be created similarly by substituting both transcriptional activators with repressor TFs (Kramer et al. 2005; Lienert et al. 2013). This type of gate is more amenable to bacteria, because prokaryotic promoters tend to be more easily repressed than activated (Ran et al. 2012). Placing two input promoters in tandem upstream of a repressor TF gene also exhibits NOR behavior (Tamsir et al. 2011). While this section outlines an overview of the available transcriptional logic gates, many more have been described in literature (Buchler et al. 2003; Goñi-Moreno and Amos 2012; Nielsen et al. 2013; Brophy and Voigt 2014).

1.4.3 Towards Building a Biochemical Computer

Construction of genetic logic gates only provides the tools necessary for subsequent combination of several parts into meaningful genetic circuits. In the year 2000, two milestone circuits were released: the bistable genetic toggle switch by Gardner et al. (2000) and the repressilator by Elowitz and Leibler (2000). The former demonstrated that synthetic gene networks can express switch-like behavior and maintain the state of the switch across generations. The latter demonstrated that a cell could be engineered to reliably switch states in a predictable temporal manner. Together these basic circuits form a foundation upon which genetic engineers and synthetic biologists can theoretically build self-replicating computers (Salis et al. 2009; Moe-Behrens 2013).

1.4.3.1 Volatile and Nonvolatile Memory

Maintaining a steady transcriptional state allows for the construction of genetic circuits where the output not only depends on current inputs, but also on a variable saved in memory. A well-studied example of such a circuit occurring naturally is the mechanism by which the lambda prophage determines whether to remain in a chromosomally integrated lysogenic state or convert into an active lytic phase. In essence, a stable lysogenic state is maintained by the CI protein, which represses most phage genes. When a bacterial host's SOS response is triggered, CI is degraded causing the activation of a number of genes that allow the prophage to excise itself and enter a lytic state, which is then maintained by the Cro protein (Johnson et al. 1981; McAdams and Shapiro 1995). This behavior is achieved through interaction between two TFs that repress each other's synthesis and is the basis upon which Gardner et al. (2000) designed the synthetic toggle switch shown in Fig. 1.9a. The switch is composed of two promoters that each drive synthesis of a repressor inhibiting the other promoter. The state is set by inactivating one of the repressors through heat shock or by adding a chemical inducer. Mathematical models demonstrate that bistability depends upon the transfer functions of the inverters making up the system. To this end, each repressor should display cooperative binding (ultrasensitivity), and the promoter strength (dynamic range) should be balanced.

A variation that uses activators instead of repressors has been applied to yeast systems (Ajo-Franklin et al. 2007), and a push-on–push-off circuit, or T-latch, has been created by combining the memory module with a NOR gate so the same sensory input can be used to toggle both states (Lou et al. 2010). The original switch design has since been expanded to include systems that can be set or reset by input promoters rather than by applying heat or chemicals (Hillenbrand et al. 2013).

Similar to a computer's random access memory (RAM), toggle switches that rely on TFs are inherently volatile. Data cannot be stored indefinitely due to repressor degradation and turnover. To address this concern, systems have been created that allow a cell to use its own DNA as a more permanent storage medium (Inniss and Silver 2013). By connecting an output promoter of a circuit to regions accessible to site-specific recombinases such as Cre and FLP, specific regions can

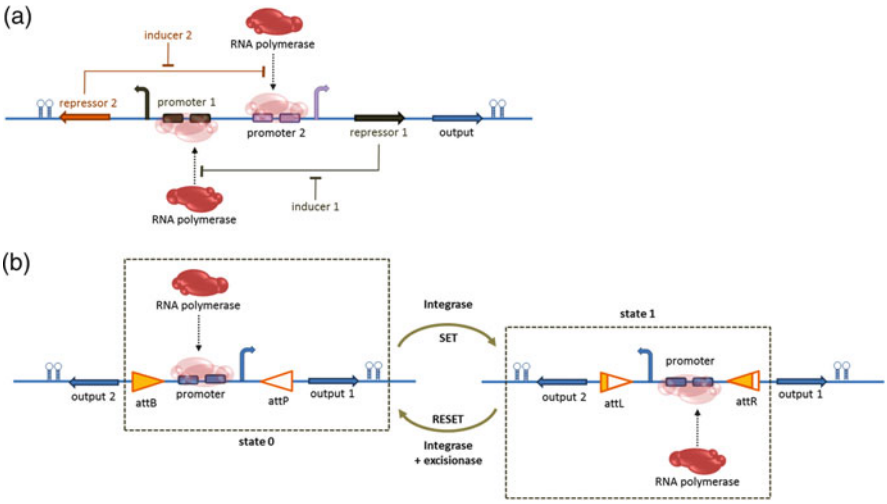


Fig. 1.9 Towards building a biochemical computer—MEMORY: (**a**) Genetic construct of a bistable genetic toggle switch: Repressor *n* inhibits transcription from promoter *n* unless inducer *n* is present ($n = \{1,2\}$) (adapted from Gardner et al. 2000). (**b**) Architecture, mechanisms, and operation of a recombinase addressable data (RAD) module. The DNA inversion RAD module is driven by two generic transcription input signals, set and reset. A set signal drives expression of an integrase that inverts a DNA element serving as a genetic data register. Flipping the register converts flanking attB and attP sites to attL and attR sites, respectively. A reset signal drives expression of the integrase as well as an excisionase and restores both the register orientation and the original flanking attB and attP sites. The register itself encodes a constitutive promoter that initiates strand-specific transcription. Following successful set or reset operations, mutually exclusive transcription output "1" or "0" is activated. For the RAD module pictured here, a "0" or "1" register state produces "output 1" or "output 2," respectively (adapted from Bonnet et al. 2012)

be programmably excised and reintegrated in a genome. More complex storage systems make use of orthogonal invertases (Ham and Lee 2006; Ham et al. 2008). Improving upon these designs, Bonnet and colleagues developed the rewritable Recombinase Addressable Data (RAD) SET/RESET latch that saves one bit of data by reversibly inverting an output promoter (Bonnet et al. 2012). In this example, a RAD module is controlled by two inputs: the SET input promoter controls the expression of an integrase that inverts a region flanked by recognition sites, and the RESET promoter drives both the integrase and an excisionase that reverses the direction of the integrase (Fig. 1.9b).

Increasing the storage capacity of a genetic memory bank requires the introduction of an orthogonal integrase for every additional register. Memory arrays that can hold up to 1375 bytes of information in 2 kb of DNA have been created through genome mining (Yang et al. 2014). Entirely new registers have been successfully created using ZFP and TALE hybrids, again demonstrating the maturity of both

technologies (Mercer et al. 2012; Gaj et al. 2014). The inverted fragment does not have to be a promoter, as any orientation-sensitive regulatory element will suffice. Moreover, by utilizing combinations of promoters and unidirectional terminators, Boolean logic has been implemented in genetic circuits that both compute and remember (Siuti et al. 2013, 2014). Both volatile and nonvolatile memory have actually been used to develop counters that record the number of times a stimulus is applied (Friedland et al. 2009; Subsoontorn and Endy 2012).

1.4.3.2 Clock Generators: Biological Metronomes

Many digital circuits incorporate a clock generator that produces an output continuously oscillating between high and low states. Oscillatory behavior is closely related to bistability and is seen in nature as well, as exemplified by the circadian oscillator of Cyanobacteria (Ishiura 1998). Elowitz and Leibler (2000) built a synthetic oscillator by daisy-chaining three repressible promoters into a cyclic negative feedback loop and linking one of the repressors to a reporter gene (Fig. 1.10a). In this case, the time delay between inhibition of a repressor's synthesis and its proteolytic degradation forms the basis of the oscillatory behavior. In vivo, the repressilator exhibits oscillations with a period of 2–3 h. Genetic circuits like this one do exhibit inherent noise, owing to the stochastic nature of chemical reactions taking place on the nano-molar scale (Elowitz and Leibler 2000).

With this in mind, researchers have developed genetic "relaxation oscillators," which consist of an activator and a repressor (Fig. 1.10b). The activator (A) enhances the expression of itself and a repressor (R). R counteracts A either by disabling A by binding to it or preventing its synthesis. In some cases R directly represses its own synthesis as well (Barkai and Leibler 2000; Hasty et al. 2002a; Atkinson et al. 2003). A circuit of this type takes advantage of the principle of hysteresis rather than time delay. At intermediate concentrations of A and R, the system remains in steady state, but a small increase of A over R causes run-away expression of A due to the positive feedback loop. R increases progressively as well until the system reaches a second steady state with high concentrations of A and R. From this state, a small increase of R over A disables the positive feedback loop, causing the system to quickly fall back to the first steady state (Savageau 2002; Lomnitz and Savageau 2014). A similar circuit has been implemented in mammalian cells as well (Tigges et al. 2010), and it has been indicated by some models that the positive feedback loop is dispensable, as long as the circuit is finely tuned (Stricker et al. 2008; Mather et al. 2009). Of special interest to metabolic engineers is a third type of transcriptional oscillator termed the "metabolator." In contrast to the other oscillators, it uses the relative concentrations of different metabolites to affect its state, rather than the concentrations of TFs (Fung et al. 2005). Such systems are useful for controlling relative amounts of enzymes needed to catalyze steps in a metabolic pathway, where the levels of different intermediates must be maintained within certain acceptable ranges.

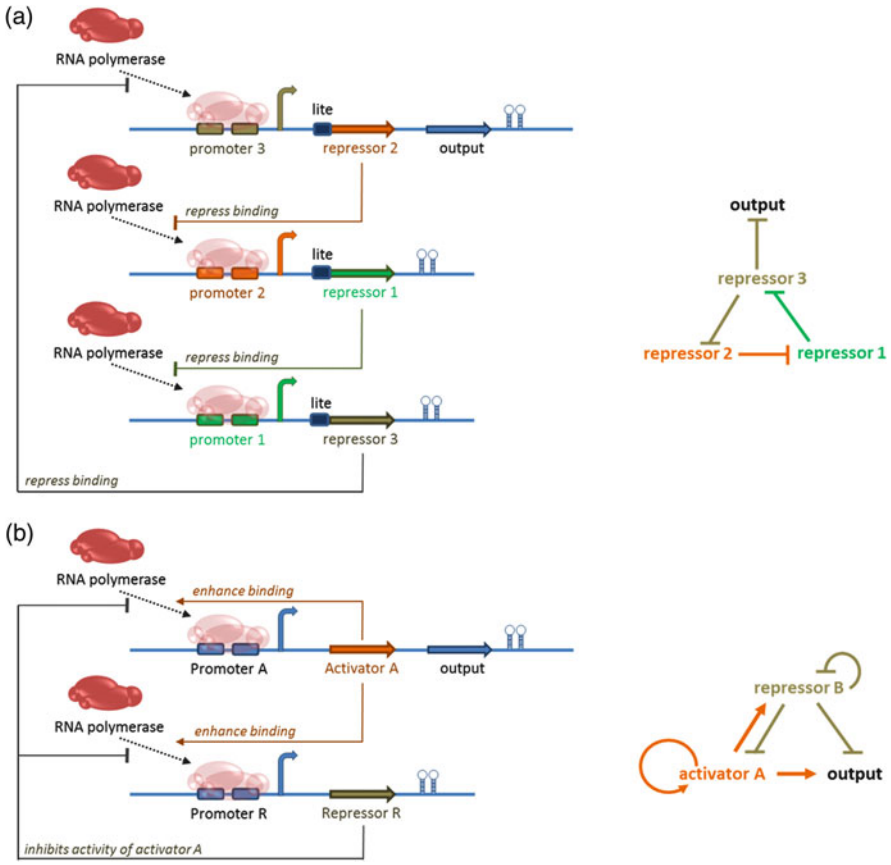


Fig. 1.10 Towards building a biochemical computer—CLOCKS. (**a**) Time-delay oscillator: the repressilator is a cyclic negative feedback loop composed of three repressor genes and their corresponding promoters, as shown schematically on the *right*. The stability of the three repressors is reduced by the presence of destruction tags (denoted "lite") (adapted from Barkai and Leibler 2000; Elowitz and Leibler 2000). (**b**) Relaxation oscillator: a positive element, activator A, increases its own expression and that of a negative element, repressor R. Strong binding of R to A inhibits A activity and thus represses the expression of both elements by binding to the promoters P_A and P_R (adapted from Barkai and Leibler 2000)

1.4.3.3 Network Interfaces Enable Multicellular Computing

Different cells carrying oscillators such as those discussed in the previous section run out of phase in relation to each other; however, they can be synchronized across an entire population through the use of quorum sensing. Quorum sensing is used in nature to detect cell density, as is exemplified by *Alivibrio fischeri*, which uses it to activate bioluminescence. In essence, each cell enzymatically produces a diffusible

small inducer molecule (e.g., an acyl homoserine lactone or AHL), which is communicated to all other cells in the medium. The light-generating system, along with AHL synthesis itself, is repressed unless AHL is present at a high enough concentration. Integral to these systems is the fact that a single cell can never reach activating AHL levels on its own (Waters and Bassler 2005). At high cell densities, AHL can be used to synchronize the state of an oscillator across an entire population. Building upon simulated models (McMillen et al. 2002; Garcia-Ojalvo et al. 2004), a number of multicellular clocks that operate on this principle have been built. The oscillations are brought about by the expression of an AHL degradation gene in response to high AHL concentrations. At high enough levels of AHL, the entire population begins degrading it until the levels drop below a certain threshold, after which AHL synthesis simultaneously starts again in all cells (Danino et al. 2010; Mondragón-Palomino et al. 2011).

Another application of quorum sensing is the creation of multicellular genetic circuits. Robust circuits have been created by culturing a number of engineered strains together, each carrying a subunit of the circuit, which use diffusing signal molecules or excreted metabolites to act as "chemical wires" between cells or colonies (Tamsir et al. 2011; Ji et al. 2013; Silva-Rocha and de Lorenzo 2014). On solid media, a spatial concentration gradient can be formed that radiates from colonies emitting an inducer. This enables the generation of programmed spatial patterns of gene expression capable of being visualized by using fluorescent reporters (Basu et al. 2005). The concentration of an inducer also fluctuates over time, which has been exploited to create pulse-generating networks (Basu et al. 2004).

The optical edge detection circuit shown in Fig. 1.11 is another example of complex circuitry that makes use of intercellular communication within a population of genetically identical cells (Tabor et al. 2009). A bacterial lawn of cells carrying this circuit visualizes the boundaries between dark and light regions of a projected image. This complex behavior results from a relatively low number of components: a transcriptional light sensor, a quorum sensing system, a pigment output system, and the logic gates connecting these parts. The photoreceptor is in fact a darkness (i.e., NOT light) sensor, which consists of a hybrid two-component system where the extracellular domain of the EnvZ sensor kinase was replaced by a cyanobacterial photoreceptor (Levskaya et al. 2005). In cells not directly hit by light, the response regulator is phosphorylated, activating a promoter that controls the synthesis of the lambda repressor (CI) and an inducer-synthesizing enzyme (LuxI). The promoter governing pigment production is repressed by CI and activated by the LuxR–AHL complex and is thus active only in the presence of cells secreting AHL. Such conditions are met only at the edges where dark regions meet light regions.

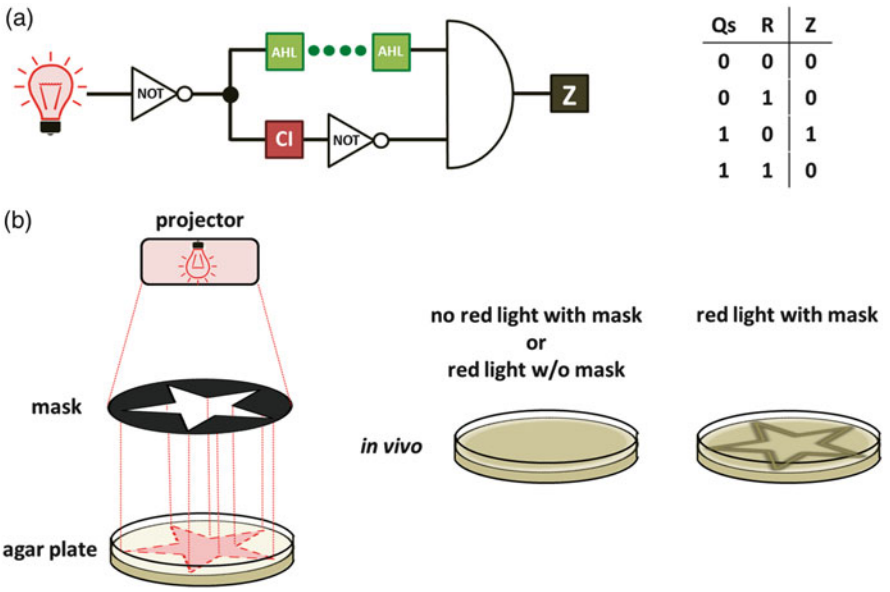


Fig. 1.11 Towards building a biochemical computer—NETWORK. (**a**) Edge Detector Circuit and truth table. *Red light* represses the expression of two genes: a biosynthetic gene for a membrane diffusible quorum sensing activator (AHL) and a dominant transcriptional repressor (CI). The output of the circuit (*Z*; beta-galactosidase) is produced only in the presence of AHL and the absence of CI. This can only occur at the light/dark boundary. (**b**) Light is projected through a mask onto a large community (lawn) of bacteria grown on an agar plate. To find the edges, bacteria in the dark produce a communication signal that diffuses across the dark/light boundary. Bacteria in the dark cannot respond to the communication signal. Only bacteria that are exposed to light and receive the signal become positive for the expression of a visible reporter gene. The sum of this activity over the entire two-dimensional population results in visualization of the edges of the input image (adapted from Tabor et al. 2009)

1.4.4 Design Principles

Building a functional genetic circuit is not as simple as connecting different logic gates. Connected components must be tuned with respect to their dynamic range and thresholds. Fortunately, several design principles have been formulated to assist genetic engineers with this task.

Recreating the behavior of natural circuits is a useful exercise when building synthetic ones, as it uncovers empirically determined properties that are not obvious in theory (Wall et al. 2004). For example, during the creation of a synthetic version of the lambda phage state-switching circuit, researchers discovered that simply substituting the viral TFs CI and Cro with TetR and LacI is not enough to obtain the desired behavior. The reason for this is that TetR, unlike CI, does not exhibit cooperative DNA binding or positive autoregulation. Success was eventually attained by combinatorially substituting each of the operator sites in the circuit and promoter sequences and RBSs with mutant alleles (Atsumi and Little 2006).

A similar strategy has been used to create a set of functional logic gates, demonstrating that combinatorial approaches can successfully yield a large diversity of behaviors originating from a limited set of genetic parts (Guet et al. 2002).

Screening combinatorial libraries for functional circuits is often unpredictable and becomes infeasible as the number of variants increases. On the other hand, directed evolution can mimic the process of natural selection by iteratively introducing random mutations while applying a selective pressure. This technique has been proven to rapidly evolve a functional circuit from two improperly matched logic gates based on CI- and LacI-mediated repression (Yokobayashi et al. 2002). In this example, error-prone PCR of the *cI* gene introduced random mutations, after which circuit performance was evaluated by measuring expression of a fluorescent reporter protein, which enables the use of fluorescence-activated cell sorting (FACS). After successive rounds of mutation and selection, functional circuits emerged due to mutations that reduced translation initiation efficiency or disrupted cooperative DNA binding of CI. High-throughput selection can also be performed by coupling circuit performance to cell survival (Collins et al. 2006; Cui et al. 2014), or the ability to create infectious phages (Esvelt et al. 2011).

Despite various advancements in high-throughput screening technology, the sequence space of a genetic circuit and its components is simply too vast to solely rely on random mutagenesis for complex circuits. As the field is maturing, it is becoming increasingly possible to optimize circuits with a more rational approach. To this end, the engineer can perform operations on three distinct levels: (1) fine-tune the transfer functions of discrete components, (2) connect additional signal processing parts in series, and (3) expand a network with parallel loops.

1.4.4.1 Turning the Control Knobs of Discrete Components

A fundamental understanding of transfer functions, introduced in Sect. 1.4.2, is central to the rational optimization of genetic circuit components. To recall, a transfer function describes the nonlinear relationship between input and output signals of a circuit component. This is frequently determined empirically by applying a chemical inducer input and subsequently measuring a fluorescent reporter protein connected to the output. The sigmoidal curve that often results suggests a model described by the Hill equation:

$$\theta(x) = \frac{x^n}{K^n + x^n} \tag{1.1}$$

This biochemical equation was first proposed in a study of the hemoglobin protein, where it describes the fraction of binding sites on the protein occupied by its ligand oxygen, as a function of the concentration of the free ligand x. In the context of transcriptional regulation, it also serves as an approximation of the probability that an operator site is occupied by a TF (Ang et al. 2013). In this case, the K parameter is equal to the TF concentration at which half the sites are occupied, while the n parameter indicates the degree of cooperative binding ($n > 1$ indicates positive cooperativity). In the most basic case of transcriptional activation

with one operator site, gene expression (i.e., the rate of change in protein concentration y) increases linearly as a function of the fraction of operator sites bound by their TF:

$$\frac{dy}{dt} = k' + k\left(\frac{x^n}{K^n + x^n}\right) \tag{1.2}$$

The basal rate of transcription is given by the term k', while the maximum increase in response to an input signal is represented by k. In case of repression through binding of a TF to one operator site, transcription over the leakage level instead increases as a function of the free operator site fraction:

$$\frac{dy}{dt} = k' + k\left(1 - \frac{x^n}{K^n + x^n}\right) = k' + k\left(\frac{K^n}{K^n + x^n}\right) \tag{1.3}$$

These equations are population averaged and are empirical approximations for the simplest cases of transcriptional regulation. More complex models of RNAP binding kinetics are described elsewhere (Bintu et al. 2005b). A scaled version of these equations can also be used to describe the steady-state concentration of a reporter protein, as long as protein degradation is assumed to be linear (Ang et al. 2013). This is useful as it is often difficult to measure rate of change.

By studying these models, it becomes possible to rationally fine-tune the performance of a circuit component in a number of ways (Arpino et al. 2013; Ang et al. 2013; Brophy and Voigt 2014). Figure 1.12 shows the effect of various operations on the transfer function of a transcriptional NOT gate (simulated data). Modifying k' leads to a vertical shift (Fig. 1.12a), where expression levels increase while the dynamic range stays the same. One way to accomplish this would be to constitutively express another copy of the gene connected to the output. This is distinct from a vertical scaling operation (Fig. 1.12b), which corresponds to multiplying equation (1.2) in its entirety, which would simultaneously alter the dynamic range and threshold. The easiest way to accomplish vertical scaling is through the gene dosage effect (i.e., expressing the circuit on a multi-copy plasmid). Cloning the circuit into a highly expressed region on the genome can also have the same effect.

Influencing the affinity of a TF for its operator site can set the threshold of a gate by affecting K, resulting in horizontal scaling (Fig. 1.12c). Importantly, the sequence of an operator site and its relative position can affect promoter leakage (i.e., residual expression at maximal repression). Tuning only the low-end level of the curve would require modifying k and k' at the same time but in the opposite direction. Adding additional operator sites to a promoter decreases leakage, but also causes horizontal scaling since more TF molecules are needed to saturate the additional copies.

Setting the threshold through horizontal scaling is not often feasible in reality, but recall that the vertical shift in Fig. 1.12a also has an effect on the switching

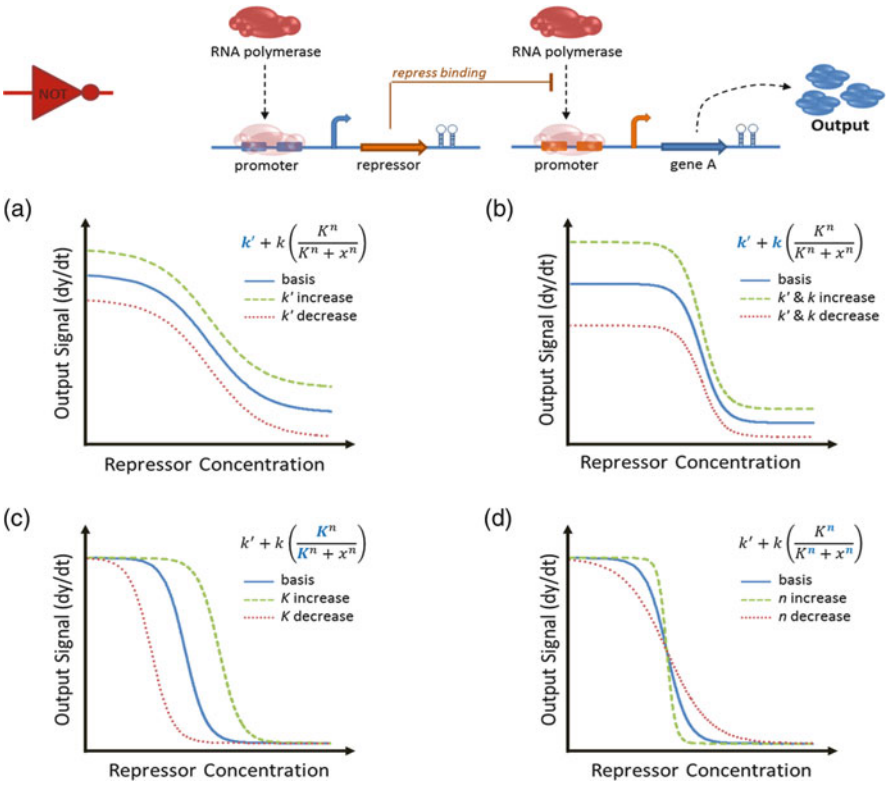


Fig. 1.12 Rational optimization of genetic circuit components. The effect of various operations on the transfer function of a transcriptional NOT gate by changing the parameter k' (**a**), k' & k equally (**b**), K (**c**), and n (**d**)

threshold. The steepness of the transfer curve as it crosses the threshold is determined by the n parameter. Cooperative binding of a TF to DNA results in a more switch-like behavior (Fig. 1.12d). Other strategies to achieve switch-like behavior include DNA looping and sequestration strategies that capture a TF through protein–protein interactions or decoy operators. Thanks to promoter library technology, tuning genetic circuits has become more commonplace. Nevertheless, many of the synthetic biology-related controls involve posttranscriptional strategies outside the scope of this chapter, such as the engineering of ribosomal binding sites (Salis 2011), transcript stability (Carrier and Keasling 1997), and protein degradation rates (McGinness et al. 2006).

1.4.4.2 Serial Signal Processing Parts

Due to biochemical limitations, it is not always feasible to attain the required signal properties by rationally fine-tuning a circuit component directly. In such cases, a dedicated signal processing part should be used to modulate the system to required

specifications. For example, a TF NOT gate can act as a simple inverting amplifier when a repressor has a high affinity for its operator, and the output promoter is strong (Karig and Weiss 2005). Alternatively, a "transcriptor" rewritable buffer gate can be used to amplify a signal by expressing a recombination directionality factor (RDF), which causes integrase-mediated inversion of a terminator or strong promoter sequence. In the absence of an input signal, the sequence is flipped back to its original state (Bonnet et al. 2013).

If the gain factor requires circumstantial adjusting, an externally tunable amplifier can be used. A buffer based on hrpR/hrpS ultrasensitive activation (i.e., the second AND gate in Fig. 1.8, modified so one promoter drives both hrpR and hrpS) amplifies the signal at a gain that is set by the expression of a third protein called hrpV, which inhibits hrpS. The gain factor is then inversely correlated to concentration of hrpV (Wang et al. 2014). When a single output signal is to be connected to multiple downstream modules, a load driver device can be placed in series to mitigate the effects of retroactivity. This device consists of a short phosphorylation cascade with a fast response time and acts as a buffer against high loads placed on the output (Mishra et al. 2014).

A band-pass filter is a different kind of signal processing device that allows signals to pass only if their intensity falls within a defined range. It is generally constructed from a high and low threshold detector, as well as a negating combiner (Basu et al. 2003). In this instance, the input promoter simultaneously drives the expression of two TFs with different individual switching characteristics. When the input signal crosses the lower threshold, one of these TFs causes activation of the output (i.e., lifting of constitutive repression); however, if the signal crosses the upper threshold, the second TF reaches a high enough concentration to propagate a negative signal towards the output, turning it back off. If the user also induces expression of an enzyme that degrades one of the signal-carrying compounds, the threshold can be tuned externally (Sohka et al. 2009).

1.4.4.3 Parallel Network Loops

Natural genetic circuits rarely consist of parts in series, but instead have often evolved as genetic networks with parallel connections and feedback loops. Despite an apparent lack of organization, regulatory patterns are seen more often than would be expected in a random network. These network motifs each carry out a specific information-processing function that can also be of use for the design and optimization of synthetic circuits (Alon 2007a, b).

The first of these motifs is the negative autoregulation (NAR—see Fig. 1.13a) loop, in which a TF represses its own transcription. An NAR displays two important functions: speeding up response times and reducing cell-to-cell variability. The increased response time results from the ability to use a much stronger promoter to drive TF transcription. After a rapid initial rise, the TF concentration is locked into a steady state near its repressive threshold, never reaching levels that would otherwise be toxic to the cell. The associated reduction in cell-to-cell variability results from a narrower protein level distribution clustered around the mean.

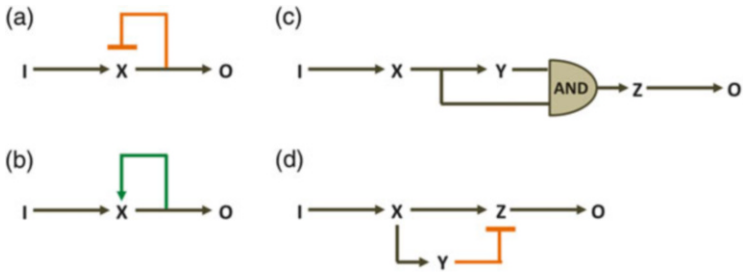


Fig. 1.13 Network motifs: (**a**) negative autoregulation, (**b**) positive autoregulation, (**c**) coherent type I feed-forward loop, and (**d**) incoherent type I feed-forward loop (adapted from Alon 2007a, b)

Positive autoregulation (PAR—see Fig. 1.13b), caused by a TF enhancing its own synthesis, has the exact opposite effect. Longer associated response times are explained as follows: at early stages, a weak basal expression leads to a slow increase in TF concentration, which suddenly increases dramatically once the concentration reaches the activation threshold. In this case, cell-to-cell variability is increased as not every cell reaches the threshold at the same time. This may even cause the formation of a bimodal distribution of protein concentration levels. This behavior can be useful when implemented as a volatile memory element or as a mixed phenotype that can better respond to environmental changes.

Feed-forward loops (FFLs—see Fig. 1.13c, d) form a second family of network motifs, and two members will be described here. A coherent type I FFL can be constructed as follows: both X AND Y transcriptional activators are required to activate an output Z, but the input only drives X, which in turn enhances transcription of Y. This type of FFL functions as a persistence detector and a sign-sensitive delay element. When the input signal is switched to a high state, expression of X commences, followed by the synthesis of Y. Z is expressed only when Y reaches a threshold (i.e., input persistence) required to activate expression of Z. It is called a sign-sensitive delay, because this delay only occurs when the input is switched from low to high (a.k.a. the "rising edge" in the electronics industry) and not when the input is turned off again (the "falling edge"), since Z is inactivated as soon as X is degraded. Replacing the AND gate with an OR gate makes the delay occur only on the falling edge. In this case, on the rising edge Z is activated immediately following expression of X, but on the falling edge Z inactivation requires that both X and Y are degraded, with residual X still capable of activating expression of Y.

In an incoherent type I FFL, the input signal again drives X, which in turn activates output Z. In addition, X also activates expression of Y, which then represses Z. This type of FFL is used to create pulse generators and accelerate response times. The pulse dynamics are a consequence of the delay between the activation of Y and Z and the accumulation of Y above the threshold required to completely repress Z. If a circuit is tuned in such a way that Y never completely

represses Z, a steady-state concentration of Z is reached after activation. This concentration limit enables the use of very strong activation of Z by X, generating a fast response time. Contrastingly, if repression of Z is made very strong, the circuit functions as a fold-change detector. The amplitude and duration of the pulse generated by this circuit depend only on the fold change of the input and not on its absolute value. In other words, shifting the concentration of a chemical inducer from 1 μM to 5 μM has the same effect as shifting from 2 μM to 10 μM, whereas a transition from 1 μM to 10 μM would produce a stronger pulse (Goentoro et al. 2009).

1.4.5 Caveats and Perspectives

As stated before in this chapter, designing a functional genetic circuit is not trivial. Many of the failure modes are a consequence of assembling many components into one genetic construct (Brophy and Voigt 2014). If the dynamic ranges of the parts are not tuned properly to one another, reduced performance or even total loss of logic function might result. Furthermore, previously characterized parts can malfunction when genetic context is not adequately considered, as promoter and RBS strengths are often influenced by surrounding DNA sequences (Yao et al. 2013). Despite these concerns, there do exist strategies to remedy different forms of sequence dependence (Davis et al. 2011; Mutalik et al. 2013).

Unexpected behavior can also occur due to a lack of orthogonality, which leads to unwanted cross talk between parts. The phenomenon of retroactivity, described earlier in Sect. 1.2.4, is a common problem when host effects are not fully considered (Jayanthi et al. 2013). In other cases, synthetic genetic parts can be toxic by titrating host factors or depleting cellular resources. Furthermore, any negative effect a circuit has on a host's growth rate imparts a selective pressure for malfunctioning mutants. For this reason, it is especially important to invest in the creation of genetically stable hosts. Keeping expression levels within lower ranges offers a means to achieve genetic stability, though it often comes at the expense of a strain's productivity (Sleight and Sauro 2013).

The performance of a genetic circuit in an industrial bioreactor compared to small-scale tests indicates that scaling factors (e.g., medium composition, mixing, aeration) can have a significant impact on synthetic regulatory networks (Moser et al. 2012). One must also consider the stochastic nature of biological systems, which are characterized by intrinsic and extrinsic noise (Elowitz et al. 2002). While these concerns are appropriate, genetic circuitry that filters out noise is available (Nelson et al. 2013), and noise can actually be useful in some contexts as well (Eldar and Elowitz 2010). Even if future results indicate that the abstraction of gene regulatory networks as digital circuits is not a good model, analogue computing can still find meaningful use (Daniel et al. 2013).

Learning the common problems associated with genetic circuits has only increased our understanding of biology. Coupling library-based approaches with in silico modeling has reduced the requirement for post hoc fine-tuning (Ellis

et al. 2009). To this end, computer-aided design (CAD) software has been developed to further establish synthetic biology as an engineering discipline (Ghosh et al. 2011; Rodrigo and Jaramillo 2013). Researchers now possess an end-to-end workflow for the engineering of biological networks (Beal et al. 2012), and our knowledge has become so advanced that there is even a vision towards building a genetic compiler (Clancy and Voigt 2010), which would actually make programming biology a reality. The potential of genetic circuits truly is exciting, and the current state of the art has already begun to advance the fields of medicine and industrial biotechnology, which will be outlined in the subsequent sections.

1.5 Transcription Engineering for New Advances in the Fields of Medicine and Industrial Biotechnology

1.5.1 Transcriptional Engineering in Medicine

Only decades ago, treating patients on the nanoscale remained the work of science fiction authors. Today, biomedical engineers are turning such fantasies into realities by engineering living cells to act as microscopic tools that can diagnose and target various diseases. In this section, current progress in the fields of infectious disease, diagnostics, diabetes, and oncology is outlined to exemplify translational applications of genetic circuit design.

1.5.1.1 Combatting Infectious Diseases: Fighting Fire with Fire

Pathogenic bacteria are increasingly acquiring resistance to many antibiotics, necessitating the development of new biotechnological strategies, such as the medical use of bacteriophages. Rather than simply using phages to infect and kill pathogenic bacteria, more complex strategies can be employed that try to reduce the risk of drug resistance caused by natural selection. In one instance, researchers engineered a bacteriophage vector to express a mutant version of the LexA TF in its host, generating a form of in vivo gTME (Lu and Collins 2009). LexA is an endogenous repressor of a tightly regulated network of genes comprising the SOS response, which can be triggered by DNA damage often caused by bactericidal antibiotics. The mutation in the TF confers resistance to proteolytic degradation, which increases repression of the SOS response making a pathogen more susceptible to antibiotics.

In another example, *E. coli* was engineered to prevent cholera infection (Duan and March 2010). *Vibrio cholerae* makes use of a natural quorum sensing genetic circuit to determine when to express virulence genes (e.g., low to intermediate cell densities). The engineered *E. coli* hijacks this network by secreting the circuit's inducer, causing the pathogen to overestimate its own density and repress its virulence genes. Furthermore, multiple groups have implemented QS circuits in pathogen targeting bacteria that selectively swim towards *Pseudomonas aeruginosa*, by detecting the inducer it secretes (Gupta et al. 2013; Hwang et al. 2014). Once arrived at the target site, high inducer concentrations trigger

the expression of a biofilm degrading nuclease and a toxic peptide that kills the target bacteria.

Many diseases are spread by vector organisms, which often makes targeting such hosts equally as effective as directly targeting the pathogen itself. In one example, researchers sought to combat dengue fever by implementing a genetic circuit into its mosquito host *Aedes aegypti*. This mosquito expresses a toxic protein in adult wing tissue if it is female and is not exposed to tetracycline (Fu et al. 2010). The toxin prevents females from mating, unless tetracycline is added to their feed (e.g., in lab or factory conditions). In theory, releasing these mosquitoes into the wild would lead to a marked drop in total population, as the engineered females are effectively sterile and males can actively propagate the genetic circuit into the population by mating with both wild-type and engineered variants (Wise de Valdez et al. 2011).

1.5.1.2 Microscopic Doctors that Diagnose and Medicate

The medical diagnostics industry already makes frequent use of biological parts, for example, in antibody-based assays. These passive components are limited in their application, which remains mostly in vitro. Transcription circuit technology allows the development of in vivo active and dynamic systems. For example, genetic memory elements have been implemented in an engineered strain of *E. coli* that keeps a record of environmental conditions as it passes through the intestine (Kotula et al. 2014). The state of the memory element is determined by analyzing the expression of a reporter protein in the fecal matter. In the future, this could lead to an affordable screen for digestive tract disorders. Rather than simply driving a reporter protein, biological sensors can also be used in gene therapy as a way to dose in vivo synthesized therapeutics in response to external cues (Weber and Fussenegger 2012). Cues of a chemical nature include amino acids, vitamins, phytochemicals, and synthetic hormone-like molecules (Aubel and Fussenegger 2010). The most well researched of the artificial physical stimuli is light, and the field of optogenetics is devoted entirely to the use of light to control biological processes in vivo (Toettcher et al. 2011).

Utilizing optogenetics, type II diabetes has been treated in mice by implanting cultured mouse cells carrying a blue light-inducible genetic circuit controlling the expression of glucagon-like peptide 1 (Ye et al. 2011). Due to the limited penetration depth of visible light, invasive fiber optic implants are often required to deliver the stimulus to the target tissue. A different group solved this problem by using radio waves to selectively heat metal nanoparticles coated with an antibody (Stanley et al. 2012). The antibody coating guides the nanoparticles to engineered cells that display a corresponding epitope tag and carry a synthetic genetic circuit. These cells release insulin in response to heat generated by the irradiated nanoparticles. Magnetite cationic liposomes perform a similar function when subjected to an oscillating magnetic field (Yamaguchi et al. 2014). While innovative, both of the aforementioned circuits require an external stimulus for function which limits the applicability as glucose homeostasis controllers.

Closed-loop glucose controller circuits are able to function independently of outside hardware by responding to blood glucose concentrations directly, which more closely resembles natural systems (Cheung et al. 2000; Won et al. 2009). These circuits are referred to as "prosthetic networks," because they replace an existing malfunctioning system (Weber and Fussenegger 2012). Hyperuricemia has been treated with prosthetic networks in cell culture and in mice, using engineered cells that maintain the blood urate concentration within a fixed range by continuously monitoring its levels and responding by expressing heterologous urate oxidase as needed (Kemmer et al. 2010). Even when an existing metabolic sensor is unavailable, the symptoms caused by a metabolite imbalance can be used as a stimulus for an engineered circuit as well. This concept has been illustrated in mice engineered to express insulin in response to diabetic ketoacidosis, which can be detected by a biological pH sensor (Ausländer et al. 2014).

1.5.1.3 Target the Tumor and Nothing But the Tumor

Today's physicians can choose from a range of therapies to treat cancer. Unfortunately, the majority of current treatments cause considerable side effects with extensive damage to healthy tissues and often allow recurrence due to incomplete tumor destruction. An ideal pharmaceutical would be able to sense its local environment, travel towards tumor sites, and specifically kill cancerous cells. For safety and ease of monitoring, such a drug should respond to external triggers and produce an externally detectable signal (Forbes 2010). One can consider all these requirements as an algorithm, with the therapy manifested as a physical implementation. Thanks to advances in the field of genetic expression engineering, biological systems now offer the means to achieve these goals. These systems include engineered bacteria and lymphocytes, as well as viruses and other gene therapy vectors (June 2007; Forbes 2010; Singh et al. 2012), which often rely on transcriptional tools to function.

There have been three primary methods demonstrated to specifically target tumors. The most widely characterized method utilizes bacterial or viral vectors to detect and/or localize towards a cancerous growth. One mechanism enabling this feat is the use of sensors that measure oxygen levels (Kizaka-Kondoh et al. 2009). Cancer is characterized by uncontrolled cell growth, which consumes more oxygen than typical somatic cells, and creates hypoxic microenvironments that are not found elsewhere in the body (Forbes 2010). The *E. coli fdhF* promoter is activated in hypoxic conditions and has been implemented in a circuit that drives expression of a *Yersinia pseudotuberculosis* invasin protein if cell densities are sufficiently high and oxygen concentrations are low (Anderson et al. 2006). *Salmonella typhimurium* naturally seeks out tumor tissues (Kasinskas and Forbes 2006), and it can even be directed to specific regions within a tumor by performing selective receptor knockouts (Kasinskas and Forbes 2007). This behavior is enabled by the FnR fumarate and nitrate reduction regulator TF. By applying promoter engineering strategies, leaky expression in non-hypoxic conditions was virtually eliminated permitting targeted delivery of toxic payloads (Ryan et al. 2009).

Using viral vectors to deliver a genetic construct into a cell's nucleus permits implementation of hybrid promoters, which can tap into a cell's entire network of endogenous TFs. Viral systems can use the endogenous hypoxia-inducible transcription factor (HIF-1) as an oxygen sensor, for example (Kizaka-Kondoh et al. 2009). The network also includes TFs related to cell-cycle control, which provide much more information about the local environment than a crude oxygen sensor (Nettelbeck et al. 2000; Dorer and Nettelbeck 2009). Integrating the various endogenous signal inputs into one meaningful response is enabled by the use of logic gates. These confer robustness to noise by minimizing false-positive cell state identification (Nissim and Bar-Ziv 2010). Xie et al. (2011) have extended this concept to a cell-type classifier circuit that senses the transcription of microRNAs and kills the cell only if it expresses a very distinct profile associated with cancer.

The second class of targeting methods involves the use of tissue-specific promoters that activate cytotoxic modules only in an organ where a physician has located a tumor. This technique has been demonstrated for the treatment of prostate cancer through transcriptionally regulated prostate-targeted gene therapy that is reviewed elsewhere (Lu 2009).

Lastly, if both autonomous strategies fail, a surgeon can give an external cue indicating the location of a tumor by injecting a chemical inducer or applying a localized physical stimulus such as radiation. Chemical inducers such as L-arabinose and salicylic acid are nontoxic with tightly regulated outputs that have been used in this context. In another stimulus using gamma-radiation to trigger a RecA/LexA system it directly penetrates tumor tissue and is not limited by diffusion (Forbes 2010). Radiation therapy can also be supplemented with gene therapy by using a hybrid promoter that responds to both hypoxia and ionizing radiation (Greco et al. 2002). This type of combination therapy offers a solution to the problem of radiation-resistant hypoxic tumor cells.

Destroying a tumor after it has been identified is another applied area for the engineering of transcription, as the same TFs that identify cancerous cells can also provide a target for cytotoxic agents. *Clostridium novyi* engineered to produce antibodies against a naturally occurring oncogenic TF has been used to combat a tumor's resistance to treatment, by sequestering the TF away from its operator sequences (Groot et al. 2007). Another strategy entails increasing the expression of a functional p53 tumor suppressor TF, since the natural gene is often mutated in cancerous cells (Lane et al. 2010). Clinical trials are under way for therapies of this type, and some products are in active use (Pol et al. 2014).

It is clear that gene therapy and engineered microorganisms will define therapeutic approaches in the coming decades. A proper understanding of the engineering of transcription is indispensable for the design and creation of these incredibly powerful but equally complex strategies. As outlined in the following section, learnings from these therapeutic designs combined with the many tools available to genetic engineers can also be translated to other useful areas of biotechnology.

1.5.2 Industrial Applications: Synthetic Biology Meets Metabolic Engineering

As industrial biotechnology matures, the field is poised to establish new sustainable sources of essential chemicals used throughout daily life. Improved methods for engineering synthetic biological systems offer the potential to efficiently produce pharmacologically active compounds, industrially important bulk chemicals, and liquid fuels for transportation (Carothers et al. 2009). The diverse set of natural molecules can be harnessed through controlled bioprocesses to yield complex chemical structures that are unattainable through traditional chemical synthesis. For instance, the molecule taxadiene, a precursor to the potent anticancer pharmaceutical taxol, is a complex cyclized molecule that can be generated in a single step from glucose in vivo (Ajikumar et al. 2010). The state of the art in the majority of applied industrial biotechnology relies on overexpression of endogenous and/or heterologous enzymes to produce a desired product in large quantity. While these approaches demonstrate the potential for biotechnology to replace traditional synthetic organic chemistry, they are not sufficient to direct biosynthetic processes towards widespread adoption.

Generation of new modular cloning and expression techniques, combined with insight garnered from *omics* technologies, has bolstered capabilities in metabolic pathway engineering (Bhan et al. 2013). Combinatorial approaches have also shown success in generating diversity in pathway expression, as seen in multiplex automated genome engineering (MAGE), which produces genomic diversity by simultaneously modifying several genetic expression levels to optimize ratios of different enzymes (Yadav et al. 2012). Ultimately obtaining high product titers and yields is of most importance for industrial fermentation processes, and controlling these variables requires fine-tuning levels of gene expression to direct flux appropriately, while balancing metabolic intermediates and cofactors through regulatory circuits (Bution et al. 2014). This effort will require novel control mechanisms that can be integrated into production strains through an interplay of synthetic biology, metabolic engineering, and systems biology analysis.

Current bioengineering in the industrial regime is less like the tightly controlled switches used in electrical engineering and more like the systems level optimizations used in mechanical engineering (Church et al. 2014). Creating genetic controls that generate well-defined levels of gene expression can offer ideal stoichiometric ratios of enzymes and metabolic intermediates, and introducing feedback controls can help prevent the accumulation of toxic levels of a compound, as well as prevent excessive metabolic burden (Carothers et al. 2009). There are several items that must be addressed before building a synthetic biochemical controller, including identification of a signaling molecule that reflects the desired metabolic state, a sensor to monitor the signal, a controller to process the sensory input, and enough knowledge of the metabolic pathway to identify the rate-limiting steps (Farmer and Liao 2000).

Widespread implementation of complex genetic circuits at the commercial scale has not yet been achieved; however, there have been considerable efforts

demonstrating relatively simple proof of concept circuits in yeast and bacterial hosts, some of which have actually been successfully implemented for use in large-scale fermentations. One such control circuit has been demonstrated through the engineering of a strain for enhanced lycopene production by optimization of the prokaryotic methyl-erythritol phosphate (MEP) pathway (Farmer and Liao 2000). In this example, both the signaling molecule acetyl phosphate and Ntr regulon sensor are utilized to sense excess glucose metabolism and activate downstream synthesis of lycopene to prevent the buildup of the toxic intermediate acetate. When tested, this system provided significantly enhanced lycopene production over the unmodified control and threefold less acetate accumulation, indicating the engineered metabolic regulation was functioning as predicted.

Other examples of this concept are illustrated through the use of dynamic sensor–regulator systems to increase biodiesel production of fatty acids. In one example, a sensor named FapR isolated from *Bacillus subtilis* was used for its ability to respond to a malonyl-CoA fatty acid intermediate. Subsequently, it was engineered to control expression of the downstream enzyme acetyl-CoA carboxylase, which becomes toxic at high levels (Liu et al. 2013). By incorporating a negative feedback loop that prevented overexpression of downstream genes and accumulation of malonyl-CoA, the circuit effectively alleviated toxicity and metabolic burden resulting in 34 % increased fatty acid titer. In a very similar report, a naturally occurring fatty acid-sensitive TF FadR was used due to its ability to derepress when bound to acyl-CoA, another key intermediate in fatty acid biosynthesis. By engineering hybrid promoters with the FadR operator that were able to turn on in the presence of the acyl-CoA intermediate, fatty acid biosynthetic modules were dynamically induced only when necessary intermediates were present, thus limiting accumulation of toxic ethanol and preventing unnecessary gene expression. These effects ultimately increased the titer to 1.5 g/l and the yield threefold to 28 % of the theoretical maximum (Zhang et al. 2012). While convenient, the implementation of such a regulatory network relies on a TF that is sensitive to a key metabolic intermediate, which is not often readily available.

One way around the reliance on characterized TFs is the use of promoters that are known to be activated during a stressful metabolic state associated with the engineering of a desired metabolic pathway. Dahl et al. (2013) have exemplified this strategy by demonstrating the use of transcriptomics to identify native *E. coli* promoters that respond to the toxicity of pathway intermediates and then using the stress response promoters to drive transcription of target genes in an engineered pathway, thus creating a dynamic feedback control mechanism. Here the engineered strain doubles the production of the target compound amorphadiene by using endogenous stress response promoters as compared to commonly used inducible promoters, providing both an increase in product titer and elimination of reliance on expensive inducer compounds.

Beyond simple sensor regulatory circuits, more complex circuit design may soon have practical functionality. Dynamically controlling gene expression using sensor–regulator feedback loops has already been well demonstrated, but synthesizing more complex mixtures of drop-in biofuels or other compounds of

interest would likely require both extensive protein and transcriptional engineering. This will rely on computational tools to aid in initial design, an approach that must be used more in the future to make significant leaps in engineering microorganisms (Way et al. 2014). Eventually, feedback mechanisms could also improve product yields by accelerating pathway component expression and reducing rise time to steady state, while oscillatory circuits could be used to program product formation that takes place in discrete steps for asymmetric syntheses or to cycle between phases of product formation and active transport out of the cell (Carothers et al. 2009). Use of increasingly complex synthetic regulation will undoubtedly continue to harness novel pathways for microbial biosynthesis.

1.6 Outlook

While still in their infancy, medical and industrial biotechnology are set to grow exponentially as new tools in transcriptional engineering become available, thanks to important work in synthetic biology and metabolic engineering. The close collaboration between these two fields is essential for moving forward. While metabolic engineering often entails a "top-down approach," where metabolic pathways are retrofitted to direct flux towards a target compound, synthetic biology takes a "bottom-up approach" by constructing new genetic functions with biological parts that can find a more general use (Nielsen et al. 2014). Exploring the synergies between these fields of research, biotechnology is entering a new age of expansion, which will undoubtedly continue to offer humanity new cures to disease, as well as sources of affordable and reliable materials available for use in everyday life.

Research in biotechnology is only beginning to unveil its vast potential. The tremendous speed at which the field has evolved is, in no small part, thanks to the formation of a community striving for openness and dissemination of knowledge and resources. Organizations including the iGEM competition and BioBricks Foundation enrich the community not only with knowledge and tools, but with strong values and ideology as well (Smolke 2009). The establishment of the Registry of Standard Biological Parts (Endy 2005) and the creation of a Synthetic Biology Open Language (Galdzicki et al. 2014) would not have been possible without these ideals. With the numerous examples of translational research illustrated in this chapter, it is clear that programming biology has already made a strong impression on several applied fields. As scientists and engineers continue to decrypt the complex regulation observed in nature and rationally build new forms of life, one can only imagine how far the limits of biology can be pushed.

References

Ajikumar PK, Xiao W, Tyo KEJ et al (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science 330:70–74

Ajo-Franklin CM, Drubin DA, Eskin JA et al (2007) Rational design of memory in eukaryotic cells. Genes Dev 21:2271–2276. doi:10.1101/gad.1586107

Alon U (2007a) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461. doi:10.1038/nrg2102

Alon U (2007b) An introduction to systems biology: design principles of biological circuits. Math Comput Biol Ser. Chapman Hall/CRC. doi:citeulike-article-id:1314150

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678–12683. doi:10.1073/pnas.0504604102

Alper H, Moxley J, Nevoigt E et al (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565–1568. doi:10.1126/science.1131969

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9:258–267. doi:10.1016/j.ymben.2006.12.002

Amann E, Ochs B, Abel K-J (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301–315

Anand R, Rai N, Thattai M (2011) Promoter reliability in modular transcriptional networks. Methods Enzymol 497:31–49. doi:10.1016/B978-0-12-385075-1.00002-0

Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355:619–627. doi:10.1016/j.jmb.2005.10.076

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3:133. doi:10.1038/msb4100173

Ang J, Harris E, Hussey BJ et al (2013) Tuning response curves for synthetic biology. ACS Synth Biol 2:547–567. doi:10.1021/sb4000564

Ansari AZ, Mapp AK (2002) Modular design of artificial transcription factors. Curr Opin Biotechnol 6:765–772

Aouida M, Piatek MJ, Bangarusamy DK, Mahfouz MM (2014) Activities and specificities of homodimeric TALENs in *Saccharomyces cerevisiae*. Curr Genet 60:61–74. doi:10.1007/s00294-013-0412-z

Apter M (1966) Cybernetics and development, 1st edn. Pergamon Press, Oxford

Arpino JAJ, Hancock EJ, Anderson J et al (2013) Tuning the dials of synthetic biology. Microbiology 159:1236–1253. doi:10.1099/mic.0.067975-0

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. Cell 113:597–607

Atsumi S, Little JW (2006) A synthetic phage lambda regulatory circuit. Proc Natl Acad Sci U S A 103:19045–19050. doi:10.1073/pnas.0603052103

Aubel D, Fussenegger M (2010) Mammalian synthetic biology–from tools to therapies. Bioessays 32:332–345. doi:10.1002/bies.200900149

Ausländer D, Ausländer S, Charpin-El Hamri G et al (2014) A synthetic multifunctional mammalian pH sensor and CO_2 transgene-control device. Mol Cell 55:397–408. doi:10.1016/j.molcel.2014.06.007

Bailey JE (1991) Towards a science of metabolic engineering. Science 252:1668–1875

Baker K, Bleczinski C, Lin H et al (2002) Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. Proc Natl Acad Sci U S A 99:16537–16542. doi:10.1073/pnas.262420099

Barkai N, Leibler S (2000) Circadian clocks limited by noise. Nature 403:267–268. doi:10.1038/35002258

Barrangou R, Fremaux C, Deveau H et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. doi:10.1126/science.1138140

Basu RS, Warner BA, Molodtsov V et al (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. J Biol Chem 289:24549–24559. doi:10.1074/jbc.M114.584037

Basu S, Gerchman Y, Collins CH et al (2005) A synthetic multicellular system for programmed pattern formation. Nature 434:1130–1134. doi:10.1038/nature03461

Basu S, Karig D, Weiss R (2003) Engineering signal processing in cells: towards molecular concentration band detection. Nat Comput 2:463–478. doi:10.1023/B:NACO.0000006774.27778.f0

Basu S, Mehreja R, Thiberge S et al (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A 101:6355–6360. doi:10.1073/pnas.0307571101

Bateman A, Coin L, Durbin R et al (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141. doi:10.1093/nar/gkh121

Beal J, Weiss R, Densmore D et al (2012) An end-to-end work flow for engineering of biological networks from high-level speci fications. ACS Synth Biol 1:317–331

Bergh O, Børsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. Nature 340:467–468. doi:10.1038/340467a0

Bhan N, Xu P, Koffas MA (2013) Pathway and protein engineering approaches to produce novel and commodity small molecules. Curr Opin Biotechnol 24:1137–1143. doi:10.1016/j.copbio.2013.02.019

Bikard D, Jiang W, Samai P et al (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437. doi:10.1093/nar/gkt520

Bintu L, Buchler NE, Garcia HG et al (2005a) Transcriptional regulation by the numbers: applications. Curr Opin Genet Dev 15:125–135. doi:10.1016/j.gde.2005.02.006

Bintu L, Buchler NE, Garcia HG et al (2005b) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15:116–124. doi:10.1016/j.gde.2005.02.007

Block DHS, Hussein R, Liang LW, Lim HN (2012) Regulatory consequences of gene translocation in bacteria. Nucleic Acids Res 40:8979–8992. doi:10.1093/nar/gks694

Blount BA, Weenink T, Vasylechko S, Ellis T (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One 7, e33279. doi:10.1371/journal.pone.0033279

Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512. doi:10.1126/science.1178811

Bonnet J, Subsoontorn P, Endy D (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. Proc Natl Acad Sci U S A 109:8884–8889. doi:10.1073/pnas.1202344109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1202344109

Bonnet J, Yin P, Ortiz ME et al (2013) Amplifying genetic logic gates. Science 340:599–603. doi:10.1126/science.1232758

Boole G (1854) An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities. Walton and Maberly, London

Brantl S, Wagner EGH (2002) An antisense RNA-mediated transcriptional attenuation mechanism functions in *Escherichia coli*. J Bacteriol 184:2740–2747. doi:10.1128/JB.184.10.2740

Brewster RC, Weinert FM, Garcia HG et al (2014) The transcription factor titration effect dictates level of gene expression. Cell 156:1312–1323. doi:10.1016/j.cell.2014.02.022

Briggs AW, Rios X, Chari R et al (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res 40, e117. doi:10.1093/nar/gks624

Bronson JE, Mazur WW, Cornish VW (2008) Transcription factor logic using chemical complementation. Mol Biosyst 4:56–58. doi:10.1039/b713852k

Brophy JAN, Voigt CA (2014) Principles of genetic circuit design. Nat Methods 11:508–520. doi:10.1038/nmeth.2926

Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol 2:57–65. doi:10.1038/nrmicro787

Bryant JA, Sellars LE, Busby SJW, Lee DJ (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. Nucleic Acids Res 42:11383–11392. doi:10.1093/nar/gku828

Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100:5136–5141. doi:10.1073/pnas.0930314100
Burge SW, Daub J, Eberhardt R et al (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226–D232
Burger A, Walczak AM, Wolynes PG (2010) Abduction and asylum in the lives of transcription factors. Proc Natl Acad Sci U S A 107:4016–4021. doi:10.1073/pnas.0915138107
Bution ML, Molina G, Abrahão MRE, Pastore GM (2014) Genetic and metabolic engineering of microorganisms for the development of new flavor compounds from terpenic substrates. Crit Rev Biotechnol 8551:1–13. doi:10.3109/07388551.2013.855161
Butler JEF, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 16:2583–2592. doi:10.1101/gad.1026202
Carothers JM, Goler JA, Keasling JD (2009) Chemical synthesis using synthetic biology. Curr Opin Biotechnol 20:498–503. doi:10.1016/j.copbio.2009.08.001
Carrier TA, Keasling JD (1997) Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. Biotechnol Prog 13:699–708. doi:10.1021/bp970095h
Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39, e82. doi:10.1093/nar/gkr218
Chappell J, Takahashi MK, Meyer S et al (2013) The centrality of RNA for engineering gene expression. Biotechnol J 8:1379–1395. doi:10.1002/biot.201300018
Chen Y-J, Liu P, Nielsen AAK et al (2013a) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 10:659–664. doi:10.1038/nmeth.2515
Chen Y-Y, Shen H-J, Cui Y-Y et al (2013b) Chromosomal evolution of *Escherichia coli* for the efficient production of lycopene. BMC Biotechnol 13:6. doi:10.1186/1472-6750-13-6
Cheng AW, Wang H, Yang H et al (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res 23:1163–1171. doi:10.1038/cr.2013.122
Cheung AT, Dayanandan B, Lewis JT et al (2000) Glucose-dependent insulin release from genetically engineered K cells. Science 290:1959–1962. doi:10.1126/science.290.5498.1959
Choo Y, Isalan M (2000) Advances in zinc finger engineering. Curr Opin Struct Biol 10:411–416
Church GM, Elowitz MB, Smolke CD et al (2014) Realizing the potential of synthetic biology. Nat Rev Mol Cell Biol 15:289–294. doi:10.1038/nrm3767
Clancy K, Voigt CA (2010) Programming cells: towards an automated "Genetic Compiler". Curr Opin Biotechnol 21:572–581. doi:10.1016/j.copbio.2010.07.005
Collins CH, Leadbetter JR, Arnold FH (2006) Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. Nat Biotechnol 24:708–712. doi:10.1038/nbt1209
Condreay JP, Witherspoon SM, Clay WC, Kost TA (1999) Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. Proc Natl Acad Sci U S A 96:127–132
Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. doi:10.1126/science.1231143
Cong L, Zhou R, Kuo Y-C et al (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun 3:968. doi:10.1038/ncomms1962
Copeland MF, Politz MC, Pfleger BF (2014) Application of TALEs, CRISPR/Cas and sRNAs as trans-acting regulators in prokaryotes. Curr Opin Biotechnol 29C:46–54. doi:10.1016/j.copbio.2014.02.010
Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Syst Biol 3:145. doi:10.1038/msb4100187
Crick F (1970) Central dogma of molecular biology. Nature 227:561–563. doi:10.1038/227561a0

Cui Y-Y, Ling C, Zhang Y-Y et al (2014) Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. Microb Cell Fact 13:21. doi:10.1186/1475-2859-13-21

Dahl RH, Zhang F, Alonso-Gutierrez J et al (2013) Engineering dynamic pathway regulation using stress-response promoters. Nat Biotechnol 31:1039–1046. doi:10.1038/nbt.2689

Dalton AC, Barton WA (2014) Over-expression of secreted proteins from mammalian cell lines. Protein Sci 23:517–525. doi:10.1002/pro.2439

Daniel R, Rubens JR, Sarpeshkar R, Lu TK (2013) Synthetic analog computation in living cells. Nature 497:619–623. doi:10.1038/nature12148

Danino T, Mondragón-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. Nature 463:326–330. doi:10.1038/nature08753

Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 39:1131–1141. doi:10.1093/nar/gkq810

Dawid A, Cayrol B, Isambert H (2009) RNA synthetic biology inspired from bacteria: construction of transcription attenuators under antisense regulation. Phys Biol 6:025007. doi:10.1088/1478-3975/6/2/025007

De Lange O, Binder A, Lahaye T (2014) From dead leaf, to new life: TAL effectors as tools for synthetic biology. Plant J 78:753–771. doi:10.1111/tpj.12431

De Las Heras A, Carreño CA, Martínez-García E, de Lorenzo V (2010) Engineering input/output nodes in prokaryotic regulatory circuits. FEMS Microbiol Rev 34:842–865. doi:10.1111/j.1574-6976.2010.00238.x

Deltcheva E, Chylinski K, Sharma CM et al (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471:602–607. doi:10.1038/nature09886

Desai TA, Rodionov DA, Gelfand MS et al (2009) Engineering transcription factors with novel DNA-binding specificity using comparative genomics. Nucleic Acids Res 37:2493–2503. doi:10.1093/nar/gkp079

Dorer DE, Nettelbeck DM (2009) Targeting cancer by transcriptional control in cancer gene therapy and viral oncolysis. Adv Drug Deliv Rev 61:554–571. doi:10.1016/j.addr.2009.03.013

Doyle EL, Booher NJ, Standage DS et al (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res 40:117–122. doi:10.1093/nar/gks608

Duan F, March JC (2010) Engineered bacterial communication prevents *Vibrio cholerae* virulence in an infant mouse model. Proc Natl Acad Sci U S A 107:11260–11264. doi:10.1073/pnas.1001294107

Edwards J, Coruzzi G (1990) Cell-specific gene expression in plants. Annu Rev Genet 24:275–303

Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467:167–173. doi:10.1038/nature09326

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822

Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol 27:465–471. doi:10.1038/nbt.1536

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338. doi:10.1038/35002125

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297:1183–1186. doi:10.1126/science.1070919

Endy D (2005) Foundations for engineering biology. Nature 438:449–453. doi:10.1038/nature04342

Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. Nature 472:499–503. doi:10.1038/nature09929

Farmer W, Liao J (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat Biotechnol 18:533–537

Farzadfard F, Perli SD, Lu TK (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth Biol 2:604–613. doi:10.1021/sb400081r

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17:37–54
Flagfeldt DB, Siewers V, Huang L, Nielsen J (2009) Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. Yeast 26:545–551. doi:10.1002/yea
Forbes NS (2010) Engineering the perfect (bacterial) cancer therapy. Nat Rev Cancer 10:785–794. doi:10.1038/nrc2934
Fraser P (2006) Transcriptional control thrown for a loop. Curr Opin Genet Dev 16:490–495. doi:10.1016/j.gde.2006.08.002
Friedland AE, Lu TK, Wang X et al (2009) Synthetic gene networks that count. Science 324:1199–1202. doi:10.1126/science.1172005
Fu F, Voytas DF (2013) Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C_2H_2 zinc fingers and engineered zinc finger arrays. Nucleic Acids Res 41:452–455. doi:10.1093/nar/gks1167
Fu G, Lees RS, Aw D et al (2010) Female-specific flightless phenotype for mosquito control. Proc Natl Acad Sci U S A 107:4550–4554. doi:10.1073/pnas.1014662107
Fung E, Wong WW, Suen JK et al (2005) A synthetic gene-metabolic oscillator. Nature 435:118–122. doi:10.1038/nature03508
Furtht PA, Onget LST, Bgert H et al (1994) Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc Natl Acad Sci 91:9302–9306
Fütterer J, Bonneville JM, Hohn T (1990) Cauliflower mosaic virus as a gene expression vector for plants. Physiol Plant 79:154–157
Gaj T, Sirk SJ, Barbas CF (2014) Expanding the scope of site-specific recombinases for genetic and metabolic engineering. Biotechnol Bioeng 111:1–15. doi:10.1002/bit.25096
Galdzicki M, Clancy KP, Oberortner E et al (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. Nat Biotechnol 32:545–550. doi:10.1038/nbt.2891
Garcia-Ojalvo J, Elowitz MB, Strogatz SH (2004) Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. Proc Natl Acad Sci U S A 101:10955–10960. doi:10.1073/pnas.0307095101
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342. doi:10.1038/35002131
Garg A, Lohmueller JJ, Silver PA, Armel TZ (2012) Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res 40:7584–7595. doi:10.1093/nar/gks404
Geissler R, Scholze H, Hahn S et al (2011) Transcriptional activators of human genes with programmable DNA-specificity. PLoS One 6, e19509. doi:10.1371/journal.pone.0019509
Ghosh S, Matsuoka Y, Asai Y et al (2011) Software for systems biology: from tools to integrated platforms. Nat Rev Genet 12:821–832. doi:10.1038/nrg3096
Gilbert LA, Larson MH, Morsut L et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154:442–451. doi:10.1016/j.cell.2013.06.044
Goentoro L, Shoval O, Kirschner MW, Alon U (2009) The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol Cell 36:894–899. doi:10.1016/j.molcel.2009.11.018
Goñi-Moreno A, Amos M (2012) A reconfigurable NAND/NOR genetic logic gate. BMC Syst Biol 6:126. doi:10.1186/1752-0509-6-126
Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A 89:5547–5551
Gossen M, Freundlieb S, Bender G, Muller G (1995) Transcriptional activation by tetracyclines in mammalian cells. Science 268:1766–1769
Gratz SJ, Ukken FP, Rubinstein CD et al (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. Genetics 196:961–971. doi:10.1534/genetics.113.160713

Grau J, Boch J, Posch S (2013) TALENoffer: genome-wide TALEN off-target prediction. Bioinformatics 29:2931–2932. doi:10.1038/mt.2013.56.Streubel
Grau J, Keilwagen J, Gohr A (2012) Jstacs: a Java framework for statistical analysis and classification of biological sequences. J Mach Learn Res 13:1967–1971
Greco O, Marples B, Dachs GU et al (2002) Novel chimeric gene promoters responsive to hypoxia and ionizing radiation. Gene Ther 9:1403–1411. doi:10.1038/sj.gt.3301823
Groot AJ, Mengesha A, van der Wall E et al (2007) Functional antibodies produced by oncolytic clostridia. Biochem Biophys Res Commun 364:985–989. doi:10.1016/j.bbrc.2007.10.126
Guelzim N, Bottani S, Bourgine P, Képès F (2002) Topological and causal structure of the yeast transcriptional regulatory network. Nat Genet 31:60–63. doi:10.1038/ng873
Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296:1466–1470. doi:10.1126/science.1067407
Guido NJ, Wang X, Adalsteinsson D et al (2006) A bottom-up approach to gene regulation. Nature 439:856–860. doi:10.1038/nature04473
Gupta S, Bram EE, Weiss R (2013) Genetically programmable pathogen sense and destroy. ACS Synth Biol 2:715–723. doi:10.1021/sb4000417
Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J Bacteriol 177:4121–4130
Hahn S (2004) Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol 11:394–403. doi:10.1038/nsmb763
Ham T, Lee S (2006) A tightly regulated inducible expression system utilizing the fim inversion recombination switch. Biotechnol Bioeng 94:1–4. doi:10.1002/bit
Ham TS, Lee SK, Keasling JD, Arkin AP (2008) Design and construction of a double inversion recombination switch for heritable sequential genetic memory. PLoS One 3, e2815. doi:10.1371/journal.pone.0002815
Hasty J, Dolnik M, Rottschäfer V, Collins J (2002a) Synthetic gene network for entraining and amplifying cellular oscillations. Phys Rev Lett 88:148101. doi:10.1103/PhysRevLett.88.148101
Hasty J, McMillen D, Collins JJ (2002b) Engineered gene circuits. Nature 420:224–230. doi:10.1038/nature01257
Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. Nat Methods 11:122–123. doi:10.1038/nmeth.2812
Heigwer F, Kerr G, Walther N et al (2013) E-TALEN: a web tool to design TALENs for genome engineering. Nucleic Acids Res 41, e190. doi:10.1093/nar/gkt789
Hillenbrand P, Fritz G, Gerland U (2013) Biological signal processing with a genetic toggle switch. PLoS One 8, e68345. doi:10.1371/journal.pone.0068345
Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431. doi:10.1093/nar/gkg599
Hollands K, Proshkin S, Sklyarova S et al (2012) Riboswitch control of Rho-dependent transcription termination. Proc Natl Acad Sci U S A 109:5376–5381. doi:10.1073/pnas.1112211109
Hsu PD, Scott DA, Weinstein JA et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827–832. doi:10.1038/nbt.2647
Hurt JA, Thibodeau SA, Hirsh AS et al (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. Proc Natl Acad Sci U S A 100:12271–12276. doi:10.1073/pnas.2135381100
Hwang IY, Tan MH, Koh E et al (2014) Reprogramming microbes to be pathogen-seeking killers. ACS Synth Biol 3:228–237. doi:10.1021/sb400077j
Inniss MC, Silver PA (2013) Building synthetic memory. Curr Biol 23:R812–R816. doi:10.1016/j.cub.2013.06.047
Ishiura M (1998) Expression of a gene cluster *kaiABC* as a circadian feedback process in Cyanobacteria. Science 281:1519–1523. doi:10.1126/science.281.5382.1519

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–356. doi:10.1016/S0022-2836(61)80072-7

Jayanthi S, Nilgiriwala KS, Del Vecchio D (2013) Retroactivity controls the temporal dynamics of gene transcription. ACS Synth Biol 2:431–441. doi:10.1021/sb300098w

Jeong KS, Ahn J, Khodursky AB (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. Genome Biol 5:R86. doi:10.1186/gb-2004-5-11-r86

Ji W, Shi H, Zhang H et al (2013) A formalized design process for bacterial consortia that perform logic computing. PLoS One 8, e57482. doi:10.1371/journal.pone.0057482

Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. doi:10.1126/science.1225829

Johnson AD, Poteete AR, Lauer G et al (1981) Lambda Repressor and cro – components of an efficient molecular switch. Nature 294:217–223

June CH (2007) Science in medicine Adoptive T cell therapy for cancer in the clinic. doi:10.1172/JCI32446.1466

Karig D, Weiss R (2005) Signal-amplifying genetic circuit enables *in vivo* observation of weak promoter activation in the Rhl quorum sensing system. Biotechnol Bioeng 89:709–718. doi:10.1002/bit.20371

Kasinskas RW, Forbes NS (2006) Salmonella typhimurium specifically chemotax and proliferate in heterogeneous tumor tissue *in vitro*. Biotechnol Bioeng 94:710–721. doi:10.1002/bit.20883

Kasinskas RW, Forbes NS (2007) Salmonella typhimurium lacking ribose chemoreceptors localize in tumor quiescence and induce apoptosis. Cancer Res 67:3201–3209. doi:10.1158/0008-5472.CAN-06-2618

Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 22:437–467

Kelly JR, Rubin AJ, Davis JH et al (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng 3:4. doi:10.1186/1754-1611-3-4

Kemmer C, Gitzinger M, Daoud-El Baba M et al (2010) Self-sufficient control of urate homeostasis in mice by a synthetic circuit. Nat Biotechnol 28:355–360. doi:10.1038/nbt.1617

Képès F (2004) Periodic transcriptional organization of the E. coli genome. J Mol Biol 340:957–964. doi:10.1016/j.jmb.2004.05.039

Kizaka-Kondoh S, Tanaka S, Harada H, Hiraoka M (2009) The HIF-1-active microenvironment: an environmental target for cancer therapy. Adv Drug Deliv Rev 61:623–632. doi:10.1016/j.addr.2009.01.006

Klein-Marcuschamer D, Stephanopoulos G (2010) Method for designing and optimizing random-search libraries for strain improvement. Appl Environ Microbiol 76:5541–5546. doi:10.1128/AEM.00828-10

Klug A (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem 79:213–231. doi:10.1146/annurev-biochem-010909-095056

Kobayashi H, Kaern M, Araki M et al (2004) Programmable cells: interfacing natural and engineered gene networks. Proc Natl Acad Sci U S A 101:8414–8419. doi:10.1073/pnas.0402940101

Konermann S, Brigham MD, Trevino AE et al (2013) Optical control of mammalian endogenous transcription and epigenetic states. Nature 500:472–476. doi:10.1038/nature12466

Kortmann J, Narberhaus F (2012) Bacterial RNA thermometers: molecular zippers and switches. Nat Rev Microbiol 10:255–265. doi:10.1038/nrmicro2730

Kotula JW, Kerns SJ, Shaket LA et al (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. Proc Natl Acad Sci U S A 111:4838–4843. doi:10.1073/pnas.1321321111

Kramer BP, Fischer M, Fussenegger M (2005) Semi-synthetic mammalian gene regulatory networks. Metab Eng 7:241–250. doi:10.1016/j.ymben.2005.02.005

Kuhlman TE, Cox EC (2012) Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. Mol Syst Biol 8:610. doi:10.1038/msb.2012.42

Lane DP, Cheok CF, Lain S (2010) P53-based cancer therapy. Cold Spring Harb Perspect Biol 2: a001222. doi:10.1101/cshperspect.a001222

Larson MH, Gilbert LA, Wang X et al (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc 8:2180–2196. doi:10.1038/nprot.2013.132

Lee JF, Hesselberth JR, Meyers LA, Ellington AD (2004) Aptamer database. Nucleic Acids Res 32:D95–D100

Lee T-H, Maheshri N (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. Mol Syst Biol 8:576. doi:10.1038/msb.2012.7

Levskaya A, Chevalier AA, Tabor JJ et al (2005) Synthetic biology: engineering *Escherichia coli* to see light. Nature 438:441–442

Li L, Piatek MJ, Atef A et al (2012a) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. Plant Mol Biol 78:407–416. doi:10.1007/s11103-012-9875-4

Li M, Wang J, Geng Y et al (2012b) A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. Microb Cell Fact 11:19. doi:10.1186/1475-2859-11-19

Lienert F, Torella JP, Chen J-H et al (2013) Two- and three-input TALE-based AND logic computation in embryonic stem cells. Nucleic Acids Res 41:9967–9975. doi:10.1093/nar/gkt758

Liu C, Qi L, Lucks J et al (2012) An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. Nat Methods 9:1088–1094. doi:10.1038/NMETH.2184

Liu CC, Qi L, Yanofsky C, Arkin AP (2011) Regulation of transcription by unnatural amino acids. Nat Biotechnol 29:164–168. doi:10.1038/nbt.1741

Liu D, Xiao Y, Evans B, Zhang F (2013) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. ACS Synth Biol. doi:10.1021/sb400158w

Liu H, Yan M, Lai C et al (2010) gTME for improved xylose fermentation of *Saccharomyces cerevisiae*. Appl Biochem Biotechnol 160:574–582. doi:10.1007/s12010-008-8431-9

Liu J, Chen Y, Jiao R (2014) TALEN-mediated *Drosophila* genome editing: protocols and applications. Methods 69:22–31. doi:10.1016/j.ymeth.2014.04.010

Liu X, Brutlag D, Liu J (2001) Bioprospector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pacific Symp Biocomput 138:127–138

Lohmueller JJ, Armel TZ, Silver PA (2012) A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. Nucleic Acids Res 40:5180–5187. doi:10.1093/nar/gks142

Lomnitz JG, Savageau MA (2014) Strategy revealing phenotypic differences among synthetic oscillator designs. ACS Synth Biol 3:686–701. doi:10.1021/sb500236e

Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423:185–190. doi:10.1038/nature01556

Lou C, Liu X, Ni M et al (2010) Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. Mol Syst Biol 6:350. doi:10.1038/msb.2010.2

Lu TK, Collins JJ (2009) Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. Proc Natl Acad Sci U S A 106:4629–4634. doi:10.1073/pnas.0800442106

Lu Y (2009) Transcriptionally regulated, prostate-targeted gene therapy for prostate cancer. Adv Drug Deliv Rev 61:572–588. doi:10.1016/j.addr.2009.03.014

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res 25:1203–1210

Maeder M, Linder S, Cascio V, Fu Y (2013a) CRISPR RNA-guided activation of endogenous human genes. Nat Methods 10:977–979. doi:10.1038/NMETH.2598

Maeder M, Thibodeau-Beganny S, Osiak A (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31:294–301. doi:10.1016/j.molcel.2008.06.016.Rapid

Maeder ML, Angstman JF, Richardson ME et al (2013b) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol 31:1137–1142. doi:10.1038/nbt.2726

Maeder ML, Linder SJ, Reyon D et al (2013c) Robust, synergistic regulation of human gene expression using TALE activators. Nat Methods 10:243–245. doi:10.1038/nmeth.2366

Mahfouz MM, Li L, Piatek M et al (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Mol Biol 78:311–321. doi:10.1007/s11103-011-9866-x

Mairhofer J, Wittwer A, Cserjan-Puschmann M, Striedner G (2014) Preventing T7 RNA polymerase read-through transcription – a synthetic termination signal capable of improving bioprocess stability. ACS Synth Biol. doi:10.1021/sb5000115

Mali P, Aach J, Stranges PB et al (2013a) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31:833–838. doi:10.1038/nbt.2675

Mali P, Yang L, Esvelt KM et al (2013b) RNA-guided human genome engineering via Cas9. Science 339:823–826. doi:10.1126/science.1232033

Margolin JF, Friedman JR, Meyer WK et al (1994) Krüppel-associated boxes are potent transcriptional repression domains. Proc Natl Acad Sci U S A 91:4509–4513

Marois E, Van den Ackerveken G, Bonas U (2002) The *Xanthomonas* type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. Mol Plant Microb Interact 15:637–646. doi:10.1094/MPMI.2002.15.7.637

Mather W, Bennett M, Hasty J, Tsimring L (2009) Delay-induced degrade-and-fire oscillations in small genetic circuits. Phys Rev Lett 102:068105. doi:10.1103/PhysRevLett.102.068105

McAdams H, Shapiro L (1995) Circuit stimulation of genetic networks. Science 269:650–656

McAdams HH, Arkin A (2000) Gene regulation: towards a circuit engineering discipline. Curr Biol 10:R318–R320. doi:10.1016/S0960-9822(00)00440-1

McGinness KE, Baker TA, Sauer RT (2006) Engineering controllable protein degradation. Mol Cell 22:701–707. doi:10.1016/j.molcel.2006.04.027

McMillen D, Kopell N, Hasty J, Collins JJ (2002) Synchronizing genetic relaxation oscillators by intercell signaling. Proc Natl Acad Sci U S A 99:679–684. doi:10.1073/pnas.022642299

Meinhardt S, Manley MW, Becker NA et al (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. Nucleic Acids Res 40:11139–11154. doi:10.1093/nar/gks806

Mendenhall EM, Williamson KE, Reyon D et al (2013) Locus-specific editing of histone modifications at endogenous enhancers. Nat Biotechnol 31:1133–1136. doi:10.1038/nbt.2701

Mercer AC, Gaj T, Fuller RP, Barbas CF (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Res 40:11163–11172. doi:10.1093/nar/gks875

Mercer AC, Gaj T, Sirk SJ et al (2014) Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors. ACS Synth Biol 3:723–730. doi:10.1021/sb400114p

Meyer AJ, Ellefson JW, Ellington AD (2014) Directed evolution of a panel of orthogonal T7 RNA polymerase variants for *in vivo* or *in vitro* synthetic circuitry. ACS Synth Biol. doi:10.1021/sb500299c

Michener JK, Thodey K, Liang JC, Smolke CD (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. Metab Eng 14:212–222. doi:10.1016/j.ymben.2011.09.004

Miller JC, Tan S, Qiao G et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148. doi:10.1038/nbt.1755

Mishra D, Rivera PM, Lin A et al (2014) A load driver device for engineering modularity in biological networks. Nat Biotechnol 32:1268–1275

Moe-Behrens G (2013) The biological microprocessor, or how to build a computer with biological parts. Comput Struct Biotechnol 7, e201304003
Moffatt BA, Studier FW (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell 49:221–227
Mondragón-Palomino O, Danino T, Selimkhanov J et al (2011) Entrainment of a population of synthetic genetic oscillators. Science 333:1315–1319. doi:10.1126/science.1205369
Moon TS, Lou C, Tamsir A et al (2012) Genetic programs constructed from layered logic gates in single cells. Nature 491:249–253. doi:10.1038/nature11516
Morbitzer R, Römer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 107:21617–21622. doi:10.1073/pnas.1013133107
Moscou M, Bogdanove A (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501
Moser F, Broers NJ, Hartmans S et al (2012) Genetic circuit performance under conditions relevant for industrial bioreactors. ACS Synth Biol 1:555–564. doi:10.1021/sb3000832
Mutalik VK, Guimaraes JC, Cambray G et al (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10:354–360. doi:10.1038/nmeth.2404
Nash AI, Mcnulty R, Elizabeth M et al (2012) Correction for Nash et al., Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. Proc Natl Acad Sci U S A 109:5904. doi:10.1073/pnas.1204100109
Neff KL, Argue DP, Ma AC et al (2013) Mojo Hand, a TALEN design tool for genome editing applications. BMC Bioinf. doi:10.1186/1471-2105-14-1
Negi S, Imanishi M, Matsumoto M, Sugiura Y (2008) New redesigned zinc-finger proteins: design strategy and its application. Chem A Eur J 14:3236–3249. doi:10.1002/chem.200701320
Nelson E, Kurz V, Perry N (2013) Biological noise abatement: coordinating the responses of autonomous bacteria in a synthetic biofilm to a fluctuating environment using a stochastic bistable switch. ACS Synth Biol 3:286–297
Nettelbeck D, Jérôme V, Müller R (2000) Gene therapy designer promoters for tumour targeting. Trends Genet 16:174–181
Nielsen AAK, Segall-Shapiro TH, Voigt CA (2013) Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. Curr Opin Chem Biol 17:878–892. doi:10.1016/j.cbpa.2013.10.003
Nielsen J, Fussenegger M, Keasling J et al (2014) Engineering synergy in biotechnology. Nat Chem Biol 10:319–322. doi:10.1038/nchembio.1519
Nissim L, Bar-Ziv RH (2010) A tunable dual-promoter integrator for targeting of cancer cells. Mol Syst Biol 6:444. doi:10.1038/msb.2010.99
No D, Yao T, Evans R (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A 93:3346–3351
Oehler S, Amouyal M (1994) Quality and position of the three lac operators of E. coli define efficiency of repression. EMBO J 13:3348–3355
Ohlendorf R, Vidavski RR, Eldar A et al (2012) From dusk till dawn: one-plasmid systems for light-regulated gene expression. J Mol Biol 416:534–542. doi:10.1016/j.jmb.2012.01.001
Pabo C, Peisach E, Grant R (2001) Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem 70:313–340
Papworth M, Kolasinska P, Minczuk M (2006) Designer zinc-finger proteins and their applications. Gene 366:27–38. doi:10.1016/j.gene.2005.09.011
Park K-S, Jang Y-S, Lee H, Kim J-S (2005) Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells. J Bacteriol 187:5496–5499. doi:10.1128/JB.187.15.5496-5499.2005
Park K-S, Lee D, Lee H et al (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. Nat Biotechnol 21:1208–1214. doi:10.1038/nbt868

Perez-Pinera P, Kocak D, Vockley C (2013a) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10:973–976. doi:10.1038/nmeth.2600.RNA-guided
Perez-Pinera P, Ousterout DG, Brunger JM et al (2013b) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat Methods 10:239–242. doi:10.1038/nmeth.2361
Platt T (1986) Transcription termination and the regulation of gene expression. Annu Rev Biochem 55:339–372
Pol J, Bloy N, Obrist F, Eggermont A (2014) Trial Watch: oncolytic viruses for cancer therapy. Oncolmmunology 3:e28694-1–e28694-13
Politz MC, Copeland MF, Pfleger BF (2013) Artificial repressors for controlling gene expression in bacteria. Chem Commun 49:4325–4327. doi:10.1039/c2cc37107c
Polstein LR, Gersbach CA (2012) Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. J Am Chem Soc 134:16480–16483. doi:10.1021/ja3065667
Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65. doi:10.1093/nar/gkl842
Purcell O, Peccoud J, Lu TK (2014) Rule-based design of synthetic transcription factors in eukaryotes. ACS Synth Biol 3:737–744. doi:10.1021/sb400134k
Purnick PEM, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422. doi:10.1038/nrm2698
Qi L, Lucks JB, Liu CC et al (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res 40:5775–5786. doi:10.1093/nar/gks168
Qi LS, Larson MH, Gilbert L et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi:10.1016/j.cell.2013.02.022
Ran T, Douek Y, Milo L, Shapiro E (2012) A programmable NOR-based device for transcription profile analysis. Sci Rep 2:641. doi:10.1038/srep00641
Raveh-Sadka T, Levo M, Shabi U et al (2012) Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet 44:743–750. doi:10.1038/ng.2305
Remaut E, Stanssens P, Fiers W (1981) Plasmid vectors for high-efficiency expression controlled by the pL promoter of coliphage lambda. Gene 15:81–93
Reyon D, Tsai SQ, Khayter C et al (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30:460–465. doi:10.1038/nbt.2170
Rhodius VA, Segall-Shapiro TH, Sharon BD et al (2013) Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. Mol Syst Biol 9:702. doi:10.1038/msb.2013.58
Rodrigo G, Jaramillo A (2013) AutoBioCAD: full biodesign automation of genetic circuits. ACS Synth Biol 2:230–236. doi:10.1021/sb300084h
Ryan RM, Green J, Williams PJ et al (2009) Bacterial delivery of a novel cytolysin to hypoxic areas of solid tumors. Gene Ther 16:329–339. doi:10.1038/gt.2008.188
Rydenfelt M, Cox RS III, Garcia H, Phillips R (2014) Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. Phys Rev E 89:1–50
Sadowski I, Ma J, Triezenberg S, Ptashne M (1988) Gal4-VP16 is an unusually potent transcriptional activator. Nature 335:563–564
Salis H, Tamsir A, Voigt C (2009) Engineering bacterial signals and sensors. Contrib Microbiol 16:194–225. doi:10.1159/000219381
Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498:19–42. doi:10.1016/B978-0-12-385120-8.00002-4
Sander JD, Maeder ML, Reyon D et al (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res 38:W462–W468. doi:10.1093/nar/gkq319

Santos CNS, Stephanopoulos G (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. Curr Opin Chem Biol 12:168–176. doi:10.1016/j.cbpa.2008.01.017

Santos CNS, Xiao W, Stephanopoulos G (2012) Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in *Escherichia coli*. Proc Natl Acad Sci U S A 109:13538–13543. doi:10.1073/pnas.1206346109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1206346109

Savageau MA (2002) Alternative designs for a genetic switch: analysis of switching times using the piecewise power-law representation. Math Biosci 180:237–253

Schaerli Y, Gili M, Isalan M (2014) A split intein T7 RNA polymerase for transcriptional AND-logic. Nucleic Acids Res 42:12322–12328. doi:10.1093/nar/gku884

Schmid-Burgk J, Schmidt T, Kaiser V (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. Nat Biotechnol 31:76–81. doi:10.1038/nbt.2460.A

Schmidl S, Sheth R, Wu A, Tabor J (2014) Refactoring and optimization of light-switchable *Escherichia coli* two-component systems. ACS Synth Biol 3(11):820–831

Segall-Shapiro TH, Meyer AJ, Ellington AD et al (2014) A "resource allocator" for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol 10:742

Sera T (2009) Zinc-finger-based artificial transcription factors and their applications. Adv Drug Deliv Rev 61:513–526. doi:10.1016/j.addr.2009.03.012

Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152:17–24. doi:10.1016/j.cell.2012.12.024

Shannon C (1938) A symbolic analysis of relay and switching circuits. Trans Inst Electr Eng 57:713–723

Sheffer H (1913) A set of five independent postulates for Boolean algebras, with application to logical constants. Trans Am Math Soc 14:481–488

Shimizu-Sato S, Huq E, Tepperman JM, Quail PH (2002) A light-switchable gene promoter system. Nat Biotechnol 20:1041–1044. doi:10.1038/nbt734

Shis D, Bennett M (2013) Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. Proc Natl Acad Sci U S A 110:5028–5033. doi:10.1073/pnas.1220157110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1220157110

Silva-Rocha R, de Lorenzo V (2014) Engineering multicellular logic in bacteria with metabolic wires. ACS Synth Biol 3:204–209. doi:10.1021/sb400064y

Silva-Rocha R, de Lorenzo V (2008) Mining logic gates in prokaryotic transcriptional regulation networks. FEBS Lett 582:1237–1244. doi:10.1016/j.febslet.2008.01.060

Singh PK, Doley J, Kumar GR et al (2012) Oncolytic viruses and their specific targeting to tumour cells. Indian J Med Res 136:571–584

Siuti P, Yazbek J, Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. Nat Biotechnol 31:448–452. doi:10.1038/nbt.2510

Siuti P, Yazbek J, Lu TK (2014) Engineering genetic circuits that compute and remember. Nat Protoc 9:1292–12300. doi:10.1038/nprot.2014.089

Sleight SC, Sauro HM (2013) Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. ACS Synth Biol 2:519–528. doi:10.1021/sb400055h

Smolke CD (2009) Building outside of the box: iGEM and the BioBricks Foundation. Nat Biotechnol 27:1099–1102. doi:10.1038/nbt1209-1099

Sohka T, Heins RA, Phelan RM et al (2009) An externally tunable bacterial band-pass filter. Proc Natl Acad Sci U S A 106:10135–10140. doi:10.1073/pnas.0901246106

Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem 82:237–266. doi:10.1146/annurev-biochem-072911-172315

Stanley SA, Gagner JE, Damanpour S et al (2012) Radio-wave heating of iron oxide nanoparticles can regulate plasma glucose in mice. Science 336:604–608. doi:10.1126/science.1216753

Stanton BC, Nielsen AAK, Tamsir A et al (2014a) Genomic mining of prokaryotic repressors for orthogonal logic gates. Nat Chem Biol 10:99–105. doi:10.1038/nchembio.1411
Stanton BC, Siciliano V, Ghodasara A et al (2014b) Systematic transfer of prokaryotic sensors and circuits to Mammalian cells. ACS Synth Biol 3:880–891. doi:10.1021/sb5002856
Stephanopoulos G (2012) Synthetic biology and metabolic engineering. ACS Synth Biol 1:514–525. doi:10.1021/sb300094q
Stormo GD, Tan K (2002) Mining genome databases to identify and understand new gene regulatory systems. Curr Opin Microbiol 5:149–153
Stricker J, Cookson S, Bennett MR et al (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456:516–519. doi:10.1038/nature07389
Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189:113–130
Subsoontorn P, Endy D (2012) Design and analysis of genetically encoded counters. Procedia Comput Sci 11:43–54. doi:10.1016/j.procs.2012.09.006
Sugita M (1963) Functional analysis of chemical systems *in vivo* using a logical circuit equivalent. II. The idea of a molecular automation. J Theor Biol 4:179–192
Tabor JJ, Levskaya A, Voigt CA (2012) Multichromatic control of gene expression in *Escherichia coli*. J Mol Biol 405:315–324. doi:10.1016/j.jmb.2010.10.038.Multichromatic
Tabor JJ, Salis HM, Simpson ZB et al (2009) A synthetic genetic edge detection program. Cell 137:1272–1281. doi:10.1016/j.cell.2009.04.048
Tabor S, Richardson CC (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc Natl Acad Sci U S A 82:1074–1078
Takahashi MK, Lucks JB (2013) A modular strategy for engineering orthogonal chimeric RNA transcription regulators. Nucleic Acids Res 41:7577–7588. doi:10.1093/nar/gkt452
Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical "wires.". Nature 469:212–215. doi:10.1038/nature09565
Temme K, Hill R, Segall-Shapiro TH et al (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res 40:8773–8781. doi:10.1093/nar/gks597
Thomas R (1973) Boolean formalization of genetic control circuits. J Theor Biol 42:563–585
Thomas R (1991) Regulatory networks seen as asynchronous automata: a logical description. J Theor Biol 153:1–23
Tigges M, Dénervaud N, Greber D et al (2010) A synthetic low-frequency mammalian oscillator. Nucleic Acids Res 38:2702–2711. doi:10.1093/nar/gkq121
Toettcher JE, Voigt CA, Weiner OD, Lim WA (2011) The promise of optogenetics in cell biology: interrogating molecular circuits in space and time. Nat Methods 8:35–38. doi:10.1038/nmeth.f.326
Tornøe J, Kusk P, Johansen TE, Jensen PR (2002) Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297:21–32
Tsuji S, Futaki S, Imanishi M (2013) Creating a TALE protein with unbiased 5'-T binding. Biochem Biophys Res Commun 441:262–265. doi:10.1016/j.bbrc.2013.10.060
Tyo KE, Alper HS, Stephanopoulos GN (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. Trends Biotechnol 25:132–137. doi:10.1016/j.tibtech.2007.01.003
Tyo KEJ, Ajikumar PK, Stephanopoulos G (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol 27:760–765. doi:10.1038/nbt.1555
Tyo KEJ, Nevoigt E, Stephanopoulos G (2011) Directed evolution of promoters and tandem gene arrays for customizing RNA synthesis rates and regulation. Methods Enzymol 497:135–155. doi:10.1016/B978-0-12-385075-1.00006-8
Vilar JMG, Saiz L (2013) Reliable prediction of complex phenotypes from a modular design in free energy space: an extensive exploration of the *lac* operon. ACS Synth Biol 2:576–586
Voigt P, Reinberg D (2013) Epigenome editing. Nat Biotechnol 31:1097–1099. doi:10.1038/nbt.2756

Wachsmuth M, Findeiß S, Weissheimer N et al (2013) De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res 41:2541–2551. doi:10.1093/nar/gks1330

Wall ME, Hlavacek WS, Savageau MA (2004) Design of gene circuits: lessons from bacteria. Nat Rev Genet 5:34–42. doi:10.1038/nrg1244

Wang B, Barahona M, Buck M (2014) Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks. Nucleic Acids Res 42:9484–9492. doi:10.1093/nar/gku593

Wang B, Barahona M, Buck M, Schumacher J (2013) Rewiring cell signalling through chimaeric regulatory protein engineering. Biochem Soc Trans 41:1195–1200. doi:10.1042/BST20130138

Wang B, Kitney RI, Joly N, Buck M (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat Commun 2:508. doi:10.1038/ncomms1516

Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21:319–346. doi:10.1146/annurev.cellbio.21.012704.131001

Watson J, Crick F (1953) Molecular structure of nucleic acids. Nature 171:737–738

Way JC, Collins JJ, Keasling JD, Silver PA (2014) Integrating biological redesign: where synthetic biology came from and where it needs to go. Cell 157:151–161. doi:10.1016/j.cell.2014.02.039

Weber E, Gruetzner R, Werner S et al (2011) Assembly of designer TAL effectors by Golden Gate cloning. PLoS One 6, e19722. doi:10.1371/journal.pone.0019722

Weber W, Fussenegger M (2012) Emerging biomedical applications of synthetic biology. Nat Rev Genet 13:21–35. doi:10.1038/nrg3094

Weiss R, Homsy G, Knight T Jr (2002) Toward in vivo digital circuits. Evol as Comput Nat Comput Ser 275–295. doi:10.1007/978-3-642-55606-7_14

West AG, Fraser P (2005) Remote control of gene transcription. Hum Mol Genet 14:R101–R111. doi:10.1093/hmg/ddi104

Wise de Valdez MR, Nimmo D, Betz J et al (2011) Genetic elimination of dengue vector mosquitoes. Proc Natl Acad Sci U S A 108:4772–4775. doi:10.1073/pnas.1019295108

Wolfe SA, Ramm EI, Pabo CO (2000) Combining structure-based design with phage display to create new Cys 2 His 2 zinc finger dimers. Cell 8:739–750

Won JC, Rhee BD, Ko KS (2009) Glucose-responsive gene expression system for gene therapy. Adv Drug Deliv Rev 61:633–640. doi:10.1016/j.addr.2009.03.015

Xayaphoummine A, Bucher T, Isambert H (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. Nucleic Acids Res 33:W605–W610. doi:10.1093/nar/gki447

Xie Z, Wroblewska L, Prochazka L et al (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells. Science 333:1307–1311. doi:10.1126/science.1205527

Yadav VG, De Mey M, Giaw Lim C et al (2012) The future of metabolic engineering and synthetic biology: towards a systematic practice. Metab Eng 14:233–241. doi:10.1016/j.ymben.2012.02.001

Yamaguchi M, Ito A, Ono A et al (2014) Heat-inducible gene expression system by applying alternating magnetic field to magnetic nanoparticles. ACS Synth Biol 3:273–279. doi:10.1021/sb4000838

Yang L, Nielsen AAK, Fernandez-Rodriguez J et al (2014) Permanent genetic memory with >1-byte capacity. Nat Methods 1–9. doi:10.1038/nmeth.3147

Yao A, Fenton T, Owsley K (2013) Promoter element arising from the fusion of standard BioBrick parts. ACS Synth Biol 2:111–120

Ye H, Daoud-El Baba M, Peng R-W, Fussenegger M (2011) A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. Science 332:1565–1568. doi:10.1126/science.1203535

Yim SS, An SJ, Kang M et al (2013) Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. Biotechnol Bioeng 110:2959–2969. doi:10.1002/bit.24954

Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. Proc Natl Acad Sci U S A 99:16587–16591. doi:10.1073/pnas.252535999

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 30:354–359. doi:10.1038/nbt.2149

Zhang F, Cong L, Lodato S et al (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–154. doi:10.1038/nbt1775

Zhang Q, Bhattacharya S, Andersen M (2013) Ultrasensitive response motifs: basic amplifiers in molecular signalling networks. Open Biol 3:130031

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. doi:10.1093/nar/gkg595

Novel DNA and RNA Elements

2

Julia Pitzer, Bob Van Hove, Aaron M. Love,
Parayil Kumaran Ajikumar, Marjan De Mey, and Anton Glieder

Contents

2.1 Introduction 66
2.2 Synthetic Promoters 67
2.2.1 Hybrid Promoters 68
2.2.2 Common Strategies for the Engineering of Prokaryotic Promoters 70
2.2.3 Common Methods for the Engineering of Eukaryotic Promoters 73
2.3 Terminators 77
2.4 Ribozymes 78
2.5 Riboswitches 80
2.6 Small RNAs 81
2.6.1 Detection, Prediction, and Classification of Small RNAs 81
2.6.2 Small RNA Processing Pathways 82
2.6.3 Functions and Applications of Small RNAs 84
2.7 Long Noncoding RNAs 85
2.8 Aptamers and Adaptamers 86
2.9 DNA Barcodes 88
2.10 DNA Machines 89
2.11 DNA Walker 91
References 94

J. Pitzer • A. Glieder (✉)
Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria
e-mail: a.glieder@tugraz.at

B. Van Hove • M. De Mey
Centre for Industrial Biotechnology and Biocatalysis, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

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_2

Abstract

Impressive advances in the field of synthetic biology go hand in hand with the discovery, design, and use of novel DNA and RNA elements. Efficient synthesis of large oligonucleotides and double-stranded DNA parts, chip-based synthesis of DNA libraries, and a detailed understanding of fundamental biological mechanisms and increased capacities in bioinformatics enable new findings and applications.

In this chapter, reengineered and model-based designed DNA parts such as promoters, terminators, and aptamers, which can be used for controlled gene expression, are discussed. Furthermore, recent advances in the RNA world concerning small RNAs, ribozymes, and riboswitches and novel applications of synthetic nucleotide sequences in the form of DNA walkers and DNA machines are emphasized. The knowledge and experiences gained in the investigations of large numbers of natural and novel DNA and RNA parts have built the foundation for the design of new functions of polynucleotides and model-based regulation of gene and pathway expression.

2.1 Introduction

DNA and RNA elements provide common and some of the most efficient tools to regulate the expression of genes and pathways. In contrast to other factors, such as the genetic background and cell physiology of host strains, the engineering and use of DNA and RNA elements are also more simple and knowledge-based approaches are frequently applied. The use of such synthetic nucleic acid parts also facilitates approaches, which build on simple design and therefore can be made more or less free of unknown natural regulatory effects. This makes the design and generation of novel DNA and RNA parts a key step toward model-based regulation of protein expression and regulatory circuit design. The most studied DNA and RNA elements include promoters, ribosome binding sites, terminators, ribozymes, riboswitches, small RNAs, aptamers, as well as DNA machines and walker elements, and recently also DNA sequence elements are used for protein scaffolding (see Fig. 2.1).

Synthetic control of protein expression can occur at different levels. Several elements influence the transcription of genes such as promoters, synthetic transcriptional amplifiers (Blazeck et al. 2011, 2012), 5′ untranslated regions (UTRs), and multiple cloning sites (Crook et al. 2011), as well as terminators and 3′ UTRs (Chen et al. 2013; Curran et al. 2013), either by a direct influence on transcription or termination efficiency or by differences in transcript stability. Furthermore, the selection markers, the vectors, and the genetic context of individual systems effect expression.

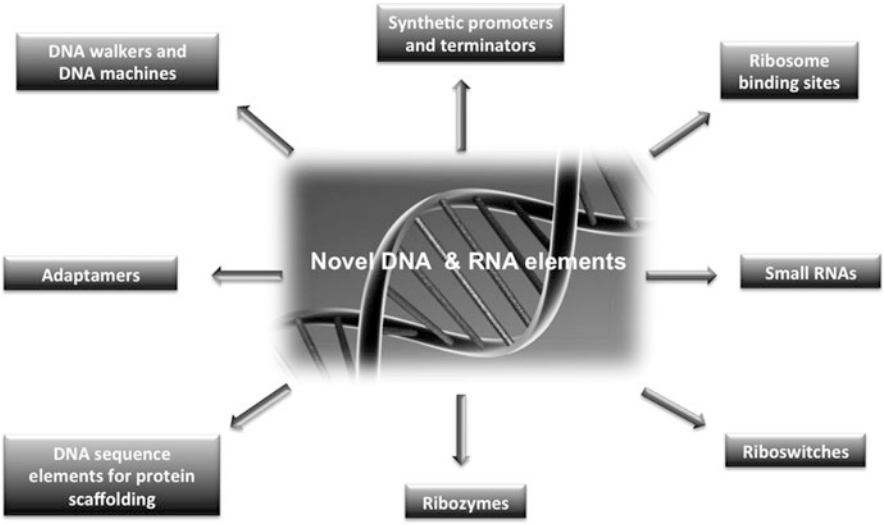


Fig. 2.1 Novel DNA and RNA elements. The most studied DNA and RNA elements comprise synthetic promoters and terminators, ribosome binding sites, small RNAs, ribozymes, riboswitches, DNA sequences for protein scaffolding, adaptamers, DNA walkers, and DNA machines. A picture of DNA molecule by Caroline Davis from https://www.flickr.com/photos/53416677@N08/4973532326

2.2 Synthetic Promoters

Engineering of promoters is a popular tool with high impact on protein expression. In order to fully understand transcriptional regulation, many different elements have to be considered, which often interact. However, in spite of such possible interactions, the modular design and construction based on the combinations of different synthetic nucleic acid parts are feasible.

Efficient and controllable promoters and knowledge about the involved transcriptional regulatory systems are essential for the optimization of protein expression (Vogl et al. 2013). There is increasing interest in using redesigned and synthetic promoters, since they broaden the natural biodiversity, facilitate the individual fine-tuning of the expression of the target gene (Ruth and Glieder 2010), and provide opportunities to overcome unknown or unexpected intrinsic regulation effects from natural promoters. In addition difficulties in respect to strain stabilities due to their sequence similarity and tendency to homologous recombination events can be avoided.

Although our knowledge about gene expression is constantly increasing, today we are able to predict transcript levels and translation initiation to some extent but not the optimal strength to express a maximum amount of biologically active recombinant protein, which differs between specific targets. Synthetic promoters

can span a wide range of expression levels and can therefore be used for many different purposes. They are especially useful for applications like the optimization of metabolic pathways. Traditional strategies focused either on gene knockout or on strong overexpression but in several cases none of these two extreme approaches lead to the desired results. Libraries of synthetic promoters provide a continuous set of different expression levels and allow fine-tuned control of gene expression (Hammer et al. 2006). However, one important aspect that has to be kept in mind concerning all different published studies about promoter engineering is the fact that the finally measured differences in the expression due to different promoter variants can have several reasons. They may be due to changed transcription, but they could also be caused by differences in mRNA stability or translation initiation (Vogl et al. 2014), since mostly reporter proteins are used to evaluate such variants rather than direct quantification of differences in transcript levels.

Promoter engineering was classified into four main strategies by Blazeck and Alper (Blazeck and Alper 2013), indicating a strong emphasis on top-down approaches rather than ab initio design of unprecedented promoter sequences so far:

1. Saturation mutagenesis of spacer regions
2. Random mutagenesis by ep-PCR
3. Hybrid promoter engineering
4. Direct modification of transcription factor binding sites (TFBSs)

Secondary structure elements probably have an important impact in successful strategies, too. More recently even the design of fully synthetic eukaryotic core promoter sequences including 5′ UTRs was demonstrated (Vogl et al. 2014), and the existence of large datasets about yeast promoter regulation (Sharon et al. 2012) with repetitive regulatory patterns paved the way toward model-based bottom-up approaches in near future. The direct modification of computationally predicted and experimentally proven TFBSs can be applied for synthetic circuit design (Blazeck and Alper 2013) as well as controlled protein expression (Hartner et al. 2008).

While for the engineering of bacterial promoters, the first of the four strategies discussed above is the most popular one for fine-tuned expression of pathways. For eukaryotes, on the other hand, ep-PCR is very common for the same purpose. Hybrid promoters provided efficient novel prokaryotic promoters already in the early days of genetic engineering and recombinant protein production and are similarly useful for strong overexpression in eukaryotic hosts.

2.2.1 Hybrid Promoters

Common methods used to express heterologous and recombinant proteins in yeast or bacteria often involve the use of hybrid promoters (see Fig. 2.2). These constructs represent some of the first means to control transcription by merging important elements such as operators and consensus sequences of multiple promoters. They are advantageous in that select characteristics relating to

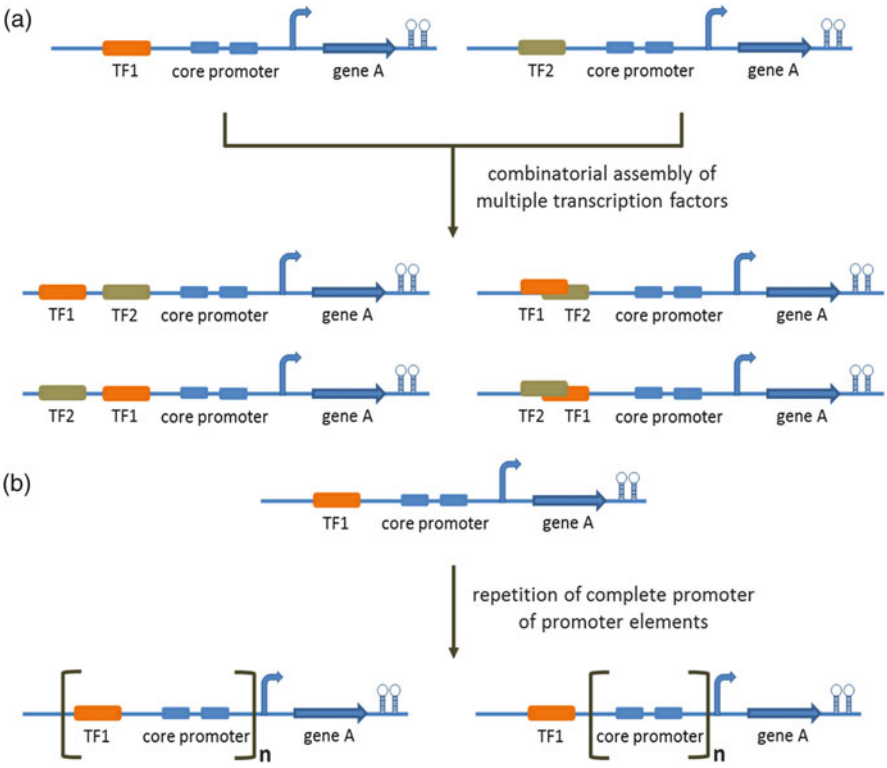


Fig. 2.2 Hybrid promoters. Construction of chimeric promoters by (**a**) merging of promoter elements and (**b**) replication of promoter (elements). n = number of tandem repeats

transcriptional regulation and/or strength can be combined in novel ways. Some of the oldest and best characterized hybrid promoters are the *tac*/*trc* promoters, which both consist of the −35 region from the *trp* promoter and the −10 region from the lacUV5 promoter (Comstock 1983). The two promoters effectively merge the strength of the *trp* promoter with the lactose- or IPTG-induced regulation of the *lac* promoter, yielding a tightly regulated strong hybrid promoter that is functional in *E. coli*.

Years of empirical data gathering have brought to light a set of general design rules for hybrid promoters. It has been shown, for example, that placing activators upstream relative to the core promoter enhances RNAP binding by recruiting activating TFs, and placing repressor operators such that they overlap with the core promoter often hinders RNAP binding (Guazzaroni and Silva-Rocha 2014). Hybrid promoters have been proven to be able to amplify genetic expression over several orders of magnitude using activating operators and/or upstream activating sequences UASs and furthermore are capable of robust repression with just one repression operator in the core promoter (Cox et al. 2007). Upstream promoter (UP) elements have been shown to play a critical role in some prokaryotic systems

by increasing expression up to 90-fold, similar to homologous enhancer elements present in higher-order eukaryotic systems (Ross et al. 1998). Once a UAS and core promoter of interest are identified, expression can be enhanced further by creating tandem repeats of the complete promoter (Li et al. 2012) or one or several UASs, which enhances TF recruitment toward the core promoter (Blazeck and Alper 2013).

More complex hybrid promoters can be generated by combining operators capable of responding to different TFs, each with their own respective ligand. This type of behavior is an essential part in creating novel transcription control elements to program gene expression. Transactivation of genes in plants, for example, has been made possible by integrating hybrid promoters alongside orthogonal TFs, and by using TFs and corresponding promoters that respond to tissue-specific ligands, gene expression can be localized to individual cellular compartments (Moore et al. 2006). Beyond rational constructs, there has been much success generating combinatorial libraries of hybrid promoters, which can be used to investigate how different regulatory elements drive or inhibit transcription as well as interact with one another (Cox et al. 2007). It has been demonstrated empirically that no strong physical constraints exist for associations between different TFs having overlapping or adjacent binding sites with respect to a core promoter (Guazzaroni and Silva-Rocha 2014). Currently, hybrid promoters find use in transcription engineering because various well-characterized regulatory regions can often be recombined to yield novel control mechanisms in a predictable manner. Their historical validation and robustness make hybrid promoters frequently used devices and provide scientists a method to investigate new functionalities inspired by natural parts.

2.2.2 Common Strategies for the Engineering of Prokaryotic Promoters

Similar to protein engineering, different methods from two main categories can be selected for the engineering of promoter sequences. The first approach is based on random mutagenesis, whereas the second one relies on rational engineering strategies. Additionally, methods that combine both strategies can be applied.

A survey of consensus promoter sequences obtained from genomic sequencing data can provide the engineer with more rational insight into the preferred bases at different positions, from which targets for randomization can be deduced (see Fig. 2.3). In *E. coli*, these include the −35 (TTGACA) and −10 (TATAAT) regions relative to the transcription start site (Mutalik et al. 2013), nucleotides surrounding the boxes (Blazeck and Alper 2013), and the spacer in between (Jensen and Hammer 1998; Solem and Jensen 2002; Hammer et al. 2006; De Mey et al. 2007). Site-specific variability is generally introduced by PCR with degenerate oligonucleotide primers, which provide a much richer library than error-prone PCR. Coussement et al. (2014) recently reported a simple and efficient method to

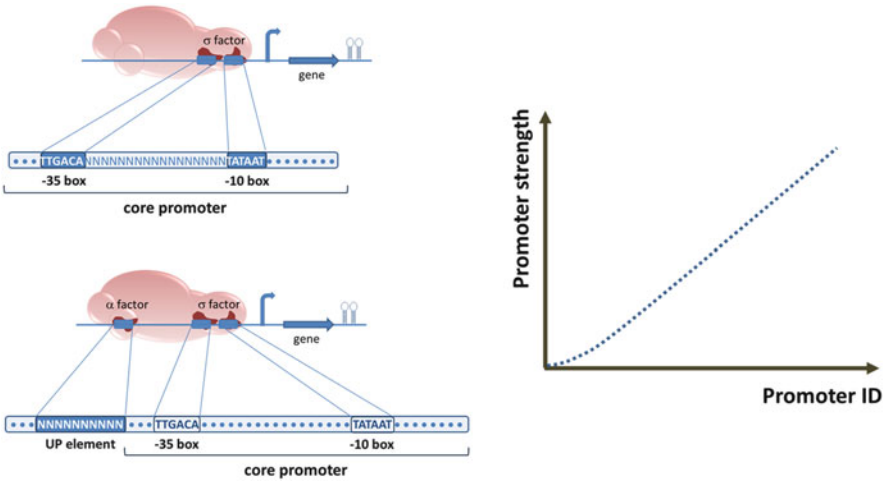


Fig. 2.3 Construction of promoter libraries. In prokaryotes the spacer sequence between the −10 box and −35 box (*top left*) is typically randomized yielding a promoter library with a broad range of expression strength (*right*). Sequences outside of the core promoter can also influence promoter strength via upstream promoter (UP) elements (*bottom left*)

assemble promoter libraries using degenerate oligonucleotides directly via Gibson assembly.

Strategies such as these have enabled the engineering of promoters over wide ranges of strength varying by several thousandfold in relative expression levels. 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. The use of a constitutive promoter as a template for library generation is often preferable with a view toward later applications, since inducible systems tend to not be economical on an industrial scale. In addition, one can use an exogenous promoter template if a more orthogonal system with high expression is desired (Tyo et al. 2011).

Engineering of promoters frequently aims for the creation of very strong promoters. However, even for recombinant protein production but especially for balancing of (synthetic) metabolic pathways, it is often not desirable to use the strongest promoters available. Instead, it can be much more helpful to have a library of promoters with continuously increasing strength on hand frequently obtained by the randomization of spacer sequences which can be attained by a single PCR step and leading to synthetic promoter libraries with a large percentage (50–90 %) of variants with different expression levels (Alper et al. 2005; Hammer et al. 2006).

Studies of Hammer et al. showed that randomization of the spacer sequences of bacterial promoters can lead to 400-fold changes in promoter activity, making the spacer regions attractive targets for mutagenesis (Jensen and Hammer 1998). One option for spacer modifications is to use PCR primers with randomized spacer sequences and homology regions to the target gene (Solem and Jensen 2002).

Advances in the technologies for chemical DNA synthesis provide highly diversified oligonucleotide sequences, which can be especially useful for the simple and fast generation of prokaryotic promoter libraries (Ruth and Glieder 2010). Cheap double-stranded DNA blocks and even long single-stranded oligonucleotides provided by several companies worldwide easily cover whole bacterial promoter regions or operons.

Alternatively, mutagenic PCR can be performed, resulting in promoters with a wide, e.g., 200-fold range of expression levels. Drawbacks of mutagenesis by PCR are the low percentage of functional promoters (around 0.1 %), which implies tedious screening processes, and the high homology of obtained promoter variants, which may reduce the genetic stability (Alper et al. 2005; Hammer et al. 2006).

Alper et al. (2005) also demonstrated the utility of the red-colored compound lycopene for engineering promoters driving expression of upstream enzymes in the methyl erythritol phosphate (MEP) pathway used in terpenoid biosynthesis in *E. coli*. In this case, the productivity of an entire pathway is easily measured using a colorimetric reporter, and furthermore by utilizing an accurate contextual screening system, they were able to engineer a range of different strength promoters while also optimizing a metabolic pathway.

The bias of all available random mutagenesis methods is a serious disadvantage for short DNA stretches such as bacterial promoters in comparison to DNA synthesis and saturation mutagenesis. This was also demonstrated in a combined approach at the MIT. Random mutagenesis by error-prone PCR (ep-PCR) was performed for the construction of synthetic libraries of the P_L-λ promoter. Stephanopoulos et al. developed a statistical method to predict the effect of a single mutation in library variants, which contain several different mutations and tested the applicability of the method on the P_L-λ promoter. The promoter variants were analyzed by fluorescence measurements using flow cytometry and revealed positions, which correlate significantly with promoter strength. Site-directed mutagenesis was performed to target these positions as well as statistically insignificant sites and combinations thereof. Seven of eight mutants showed the expected phenotype, which was predicted by the statistical method. This technology can ease the identification of targets for rational mutagenesis of biomolecules (Jensen et al. 2006).

Mutations within TFBSs and in the consensus sequence of the −35 and −10 regions as well as changes in the length of the spacer between them often lead to reduced promoter activity (Jensen and Hammer 1998; Hammer et al. 2006). The low expression of these variants can give insight into the structure of the promoter, since their mutations can reveal regions, which are responsible for efficient expression or binding of transcription factors (Remans et al. 2005; Blazeck and Alper 2013).

Since it is the arrangement and type of TFBSs that essentially define the architecture of a promoter, a systematic study was performed by Cox et al. to analyze the effect of different transcriptional regulators on expression. A combinatorial library of *E. coli* promoters was created by dividing the sequence of the promoters in three units: the distal region upstream of the −35 box, the core region

between the −10 and −35 box, and the proximal region downstream of the −10 box. Various combinations of four selected operators for transcriptional activators and repressors were incorporated in the units and they were randomly assembled by complementary overlaps. Analysis of these combinatorial variants allowed the authors to identify heuristic rules for the engineering of promoters, concerning the limits of regulation, number, and location of operators (Cox et al. 2007).

In a pioneering study by Kagiya et al., the generation of a prokaryotic promoter was achieved by random mutagenesis even though the starting eukaryotic 212-bp piece of genomic DNA from HeLa cells displayed no promoter activity at all. Within four rounds of ep-PCR, they obtained a strong bacterial promoter and demonstrated that synthetic prokaryotic promoters can be not only improved but also created in a relatively simple and fast way (Kagiya et al. 2005). However, it remains unknown if an enrichment of −10 and −35 like motifs or a change in DNA structure caused the activation of this eukaryotic genomic DNA fragment into a strong *E. coli* promoter.

No promoter is useful without reliable and reproducible expression characteristics in different contexts. This is often challenging when changing background expression strains or incorporating different genes upstream or downstream of a given promoter. These unwanted effects can be mitigated by properly insulating a promoter using buffering sequences 5′ and 3′ to the core promoter region. It is well known that UP elements can increase promoter expression several hundredfold by recruitment of core polymerase subunits, but the 20–30 nucleotides past the transcription start sites can also have a significant impact on promoter clearance and are thus important to consider when engineering an insulated promoter. Davis et al. (2011) have shown that by incorporating insulating DNA sequences from −105 up to +55 bp, they were able to negate the influence of UP elements as well as downstream inhibitory sequences on transcription efficiency, thus demonstrating the ability to engineer promoters with robust expression in a variety of contexts. Similarly, it can be of equal importance to keep the translation initiation rate constant for the resulting transcript, which can be achieved by using bicistronic domains, an architecture that couples translation of a gene of interest to an upstream miniature cistron, effectively normalizing the amount of gene expression regardless of variations in 5′ secondary structure (Mutalik et al. 2013). The wealth of mutagenesis data and library information has made it possible to begin rationally predicting promoter sequences using empirical data alongside thermodynamic models (Brewster et al. 2012). Taken together, the state of the art not only offers predictive models for engineering new promoters but also provides existing expression constructs capable of reproducibly driving gene expression over a wide range.

2.2.3 Common Methods for the Engineering of Eukaryotic Promoters

The structure of eukaryotic promoters is much more complex compared to prokaryotic promoters, making their rational engineering more challenging (Ruth and

Glieder 2010) and random mutagenesis was applied successfully (in combination with screening). One way to circumvent this problem is the application of random mutagenesis. The group of G. Stephanopoulos created mutants of the strong and constitutive TEF1 promoter from *S. cerevisiae* by ep-PCR and obtained activities ranging from 8 % to 120 %. The authors confirmed by real-time PCR that the variations are caused by different transcript levels and that they are independent of the integration mode (plasmid or promoter replacement cassettes) and the carbon source (Nevoigt et al. 2006).

However, also for eukaryotic promoters, random mutagenesis approaches by ep-PCR and saturation mutagenesis predominantly resulted in mutants with decreased expression levels. Rational methods are more likely to facilitate the increase of promoter activity (Blazeck and Alper 2013). Especially the assembly of synthetic hybrid promoters has proven to be very successful. The design of these hybrid promoters takes advantage of the architecture of eukaryotic promoters, which consist of a core promoter and various upstream activating or repressing sequences (UAS/URS). Examples for the engineering of hybrid promoters among others can be found in *Pichia pastoris* (Hartner et al. 2008), *Y. lipolytica* (Blazeck et al. 2011), and *S. cerevisiae* (Blazeck et al. 2012) as well as mammalian promoters (Gehrke et al. 2003).

By adding UAS sites from different genes to the core promoter of P_{GPD}, Alper et al. created new promoter variants, which showed up to 2.5-fold higher mRNA levels compared to the native P_{GPD}, the strongest constitutive yeast promoter (Blazeck et al. 2012). These results show that it is possible to create synthetic promoters, which clearly exceed the strength of the strongest native yeast promoters.

For the design of synthetic hybrid promoters, the specific characteristics of different organisms have to be taken into account. In *Yarrowia lipolytica* it is, for example, possible to insert more than 20 copies of tandem UAS (Blazeck et al. 2011), whereas this would be hardly possible in *S. cerevisiae*. The efficient homologous recombination machinery of this yeast limits the number of identical UAS that can be maintained in a genetically stable manner (Blazeck et al. 2012).

Eukaryotic promoter libraries, which span a wide range of expression levels, were also achieved by fusion of Gal promoter-derived binding sites for transcription of UAS from P_{GAL1} to the minimal promoters of P_{LEU} and P_{CYC}. Therefore, binding sites of the transcriptional activator Gal4p were fused to constitutive core promoters in different combinations. Additional fine-tuning was achieved by adapting the distance between the UAS and the transcriptional start site. As a result, a galactose-inducible promoter library was created with continuously increasing strength. Furthermore, addition of UAS_{CLB} and UAS_{CIT} elements (from the mitotic cyclin CLB2 gene and the mitochondrial citrate synthase CIT1) led to a linear derepression of P_{GAL} under glucose-repressive conditions. Thereby, leaky hybrid promoters were created, which show low levels of protein expression in glucose-containing media. Compared to the strong increases seen for weak promoters, addition of UAS to the strong P_{GAL} increased the transcript level "just" by 15 %. With their work Blazeck et al. could demonstrate that synthetic hybrid promoters can not only be

used to obtain a wide dynamic range of expression but also for the establishment of new synthetic regulation mechanisms (Blazeck et al. 2012).

Although the methods described here are based on diverse engineering approaches, they are eventually all influencing transcription by TFBS effects. This is achieved through addition, removal, or modification of TFBSs either as individual elements or as larger parts in case of promoter fusions. As a consequence, the direct and systematic modification of TFBSs seems to represent the most straightforward and efficient approach and a very interesting target for further studies (Blazeck and Alper 2013). Applications of this approach had been demonstrated before, for example, by Hartner et al. (2008). Since TFBS are most times only known for a few intensively studied model promoters, the authors used sequence homology of putative TFBSs for targeted deletions and insertions. Co-occurring point mutations lead to additional unexpected effects.

Hartner et al. determined putative TFBSs within the strong, methanol-inducible P_{AOX1} promoter of *Pichia pastoris (Komagataella phaffi)*. Their localization was achieved through computational predictions followed by deletion studies. By duplication and deletion of these TFBSs, a first-generation promoter library with activities between 6 % and 160 % of the native promoter was created. The new toolbox was not only tested with a reporter protein but also applied for expression of industrial enzymes. Repressing, derepressing, and inducible conditions were tested and at least 12 *cis*-acting elements were identified, which influence transcriptional regulation. Fusing these elements to basal promoters allowed the construction of short, synthetic promoters with different activities and regulation profiles. Promoter variants were constructed, with increased activity without the need of methanol addition. Short, synthetic promoters with such a regulation profile (Ruth et al. 2010) are well suited for conditions with carbon starvations in batch, fed-batch, or continuous cultivations. This approach, based on mutations within TFBSs, turned out to be highly successful for the generation of promoter variants with different strengths (Hartner et al. 2008). Some of the identified potential TFBS have been experimentally verified by others later (Kranthi et al. 2009).

Recently, Vogl et al. (2014) designed the first fully synthetic core promoter in *P. pastoris* and applied it for engineering and characterization of the P_{AOX1} core promoter. Since core promoters provide no or only very low basal transcription levels, they were fused to the upstream activating sequence of the *P. pastoris* AOX1 promoter (UAS_{AOX1}) to obtain high and easily quantifiable eGFP expression. The approach was based on a minimal consensus sequence that was obtained from the alignment of four different randomly chosen core sequences from natural *P. pastoris* promoters, which showed almost no sequence similarity. This first-generation synthetic core promoter sequence was functional in yeast but only to a very low degree. It was used as the basis for a second-generation core promoter, which was obtained by further incorporation of common core promoter sequence elements. The resulting synthetic core promoter showed at least some significant activity if fused to UAS_{AOX1}. Subsequently the native P_{AOX1} core promoter was engineered by replacing certain stretches with the sequence of the synthetic core promoter. The resulting library of synthetic variants spanned a range of 10–117 %

of the wild-type P_{AOX1} and can be used for the fine-tuning of protein expression (Vogl et al. 2014).

In Norway a different approach for the engineering of P_{AOX1}, based on the random introduction of point mutations, was used. Berg et al. selected for altered Zeocin resistance and were able to develop promoter variants with drastically increased tolerance under glucose-repressed conditions as well as under methanol-induced conditions (Berg et al. 2013). However, the effect of plasmid copy amplification of the ARS-based multicopy plasmids in *Pichia pastoris* was not discussed. The possible increase in promoter activity is much higher for the ep-PCR-based approach compared to the more rational engineering method by Vogl et al., but it requires the screening of a much higher amount of variants. The P_{AOX1} core promoters used in the two approaches differ in their length and reflect the versatile guiding principles according to which core promoters are defined. Due to the diverse promoter architectures, it is challenging to find universal rules for the definition of the exact length of core promoters. If a TATA box is present within the sequence, the 5′ end of the core promoter is usually adjusted to this motif; alternatively the length of the core promoter can be determined experimentally.

A major limitation that hampers the de novo design of fully synthetic promoters is the incomplete understanding of how *cis*-regulatory motifs effect gene expression. Detailed and systematic analysis of thousands of designed promoters revealed the influence of several parameters on expression. The effect of changes in the number, affinity, orientation, position, and organization of TFBSs and nucleosome-disfavoring sequences was assessed and measured. It turned out that the orientation of the TFBSs influenced only 8 % of the tested TFs and that the effect of sequence context can be substantial but is not as important as, e.g., single base-pair mutations in TFBSs. As it is intuitively expected, increasing the distance between the transcriptional start site and the TFBSs decreases the effect of activators and repressors. Interestingly, a 10-bp periodic relationship between the position of the TFBS and expression was identified, so that even small changes in the location of binding sites can have large effects (Sharon et al. 2012) and reflect the 10–12 bases, which correspond to one helical turn. This confirmed the importance of the three-dimensional orientation of bound TFBs in relation to other binding factors, which are needed for transcription.

In a recent study, the group of Segal et al. aimed to unravel the connection between core promoter sequences and promoter activities in yeast and humans. They thereby identified k-mer and base content sequence features, which are predictive for highly active yeast promoters. These sequences are positioned within close proximity of the transcription start site, i.e., 75 bp upstream and 50 bp downstream (Lubliner et al. 2013). The findings of these studies can prove to be very useful for future promoter engineering and can provide an important basis for the design of fully synthetic promoter variants.

A completely different approach can be applied for the generation of synthetic promoters in higher eukaryotes. Instead of known TFBSs a library of random 18-bp DNAs was fused to a minimal promoter with a TATA box and an initiator element. Thereby, over 100 DNA sequences with functional *cis*-acting motifs were

identified, which enhance expression of the minimal promoter in neuroblastoma cell line Neuro2A. Database searches led to the identification of several known as well as novel sequence motifs (Edelman et al. 2000). In metazoans very high transcription levels can be successfully reached with a synthetic super core promoter, which consists of four core promoter motifs: the TATA box, initiator, motif ten, and downstream promoter element (Juven-Gershon et al. 2006).

In summary these studies demonstrated the broad applicability and high value of synthetic promoters for engineering of gene expression throughout different classes of organisms and how close we are right now toward the challenging goal of computer-based ab initio design of fully functional synthetic, strong, and tunable promoter parts for prokaryotes as well as for eukaryotic hosts.

2.3 Terminators

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 toward the RNAP (Platt 1986). The engineer should not underestimate the importance of transcription termination, as read-through transcription may disrupt the careful regulation of downstream systems, which could include plasmid copy number control elements or other ORFs (Mairhofer et al. 2013). For example, a combination of multiple terminators is required to efficiently halt the T7 RNAP and prevent read through (Mairhofer et al. 2015). 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. 2013).

In bacteria, where transcription and translation occur simultaneously, terminators can also be used to program transcription through attenuation mechanisms. Best known in the context of the *E. coli* trp operon, attenuation is the process during which a stretch of RNA conditionally forms either a terminator or an anti-terminator depending on environmental conditions (Yanofsky 1981; Naville and Gautheret 2010). In this case, the trp attenuator displays the terminator conformation as a result of ribosomes stalling due to tryptophan starvation.

Various studies (Pfleger et al. 2006; Cambray et al. 2013; Chen et al. 2013) provided more insight into the principles underlying transcriptional termination and the influence of 3′UTRs in prokaryotes. In eukaryotes and fungal systems, on the other hand, these mechanisms are not completely understood. However, the substantial influence of terminators on gene expression and their applicability for metabolic engineering was recently demonstrated for transcriptional terminators in *S. cerevisiae*. Depending on the terminator, a 13-fold dynamic range of expression levels of a fluorescent reporter gene was obtained compared to the construct lacking a terminator. The authors found out that the variations in the transcript and protein levels were mainly caused by changes in mRNA half-life (Curran et al. 2013). The results so far indicate that synthetic terminators and 3′UTRs are

a so far almost untouched but very promising tuning knob for transcriptional regulation of gene expression.

2.4 Ribozymes

Enzymes have been mainly seen as those biomolecules responsible for catalysis, but the discovery of RNA catalysts, so-called ribozymes, has opened a new view (Ramesh and Winkler 2014).

Ribozymes are RNA molecules with enzyme-like characteristics and activities, which are capable of breaking and forming covalent bonds. They have been identified for the first time in 1982 in *Tetrahymena thermophila*, where autocatalytic RNA rearrangements have been described. The authors discovered intrinsic RNA splicing activity, which occurs without the help of enzymes or small nuclear RNAs (Kruger et al. 1982).

Several features characteristic for enzymes are also true for ribozymes. First of all, both of them are able to accelerate reaction rates, they can use cofactors, and they can be regulated by the binding of allosteric effectors. Furthermore, the formation of specific tertiary structures and active sites is important for catalysis by enzymes as well as ribozymes (Doudna and Cech 2002).

Ribozymes such as self-spicing introns play an essential role in the RNA world hypothesis. This theory describes RNA molecules, which are capable of their own assembly and self-replication by recombination and mutation, as the starting point of evolution. After developing enzymatic functions through RNA cofactors, the synthesis of enzymes started based on RNA templates and the RNA core of the ribosome. The created proteins would then outperform ribozymes and predominate. Eventually DNA was constructed to provide a double-stranded, stable, linear form of information storage (Walter 1986). Despite several objections, such as the low stability and high complexity of RNA molecules as well as the rarity and small repertoire of reactions catalyzed by RNAs, the RNA world hypothesis kept high relevance (Bernhardt 2012).

The central role of RNA molecules is still illustrated, e.g., by the ribosome where rRNA is responsible for the catalytic peptidyl transferase reaction (Nissen 2000; Doudna and Cech 2002).

The most abundant and very well-studied small endonucleolytic ribozyme is the hammerhead ribozyme. It was the first ribozyme discovered and was found in subviral plant pathogens for cleavage of multimeric replication intermediates (Prody et al. 1986). Later on, the hammerhead ribozyme was found to occur in over 50 eukaryotic genomes, mainly in repetitive DNA sequences or introns. The ribozymes of the various organisms differ greatly in their sequences and length and seem to have evolved independently (Seehafer et al. 2011). Although self-cleaving ribozymes can vary largely in their structures and catalytic strategies, they can perform the same self-cleaving reaction of 5′–3′ phosphodiester bonds or the reverse ligation (Fedor 2009). Their architectures and active sites are unique and

allow efficient general acid-base and electrostatic catalysis (Ferré-D'Amaré and Scott 2010).

Other important ribozymes that perform site-specific RNA self-cleavage are the hepatitis delta virus (HDV), hairpin, *Neurospora* Varkud satellite (VS), and *glmS* ribozymes. For cleavage they utilize base-pairing and alignment interactions between the target sequence and the cleavage site in the active center. In contrast to that, members of group I and II self-splicing introns use different mechanisms involving nucleophilic attacks and metal-ion catalysis to form mature transcripts by self-cleavage and ligation (Doudna and Cech 2002).

Interestingly, Bartel et al. found out that it is also possible for a single RNA sequence to assume two completely different ribozyme folds and consequently catalyze two different reactions (Schultes and Bartel 2000).

The mechanisms and characteristics of the different small self-cleaving ribozymes have been reviewed by Ferré-D'Amaré and Scott (2010). Recently, the list has been extended by the discovery of another member called twister RNA in many bacteria and eukaryotes (Roth et al. 2014; reviewed in Ramesh and Winkler 2014).

Additional to the natural function of cleavage of phosphodiester bonds, ribozymes can catalyze an impressive variety of reactions and they can do so even without the help of cofactors. In vitro-selected ribozymes can catalyze the formation of amide bonds, Michael adducts, nucleotides or coenzyme A, and so on (Doudna and Cech 2002). Ribozymes can furthermore catalyze the transfer of activated amino acids to tRNA. A covalent aminoacyl-ribosome intermediate is involved in charging of the tRNA (Jäschke 2001).

Recently, the research group of C. Voigt applied ribozymes as "insulator" parts in synthetic circuits to reduce the effect of the genetic context. The ribozymes do so by cleaving the 5′UTR of the mRNA, thereby generating a constant 5′ mRNA context. In this way, ribozymes can reduce the coupling effects between the promoters and 5′UTRs and improve the predictability of layered circuits with mathematical models (Lou et al. 2012; Nielsen et al. 2013).

Lately, the computational design of highly specific small-molecule-sensing allosteric ribozymes was reported. The ribozymes can be created by fusing an aptamer for the desired target molecule to an extended or minimal version of the hammerhead ribozyme. The aptamer modules are tunable and provide therefore the possibility to design tailored functions. Conservation of important tertiary interactions between the stems I and II of the hammerhead ribozyme allowed to create high-speed molecular switches, which are very specific for their ligands and serve as YES or NOT logic gates. There are several potential applications of such ribozymes as molecular sensors for regulation of gene expression, high-throughput screening arrays, or antibacterial drug discovery (Penchovsky 2013).

Very recently flanking ribozyme sequences next to gRNAs were also successfully applied as an alternative to RNA polymerase III-driven expression of gRNAs for CRISPR/Cas9-mediated genome engineering. This specific processing allowed to use strong polymerase II-dependent promoters for gRNA expression (Gao and Zhao 2014).

In Chap. 3 about genome engineering methods by A. Weninger, M. Killinger, and T. Vogl, the applications of guide RNAs, essential for genome modifications based on the CRISPR/Cas9 system (Haurwitz et al. 2010), are described in more detail. The use of guide RNAs is gaining increasing popularity due to their versatile applications (Künne et al. 2014) and their convenient availability on gBlocks and on vectors in combination with the T7 promoter or as ready-to-use building blocks for direct expression.

2.5 Riboswitches

Ribozymes are involved in essential cellular functions such as translation and RNA processing. A different class of RNAs, which regulate downstream gene expression, are so-called riboswitches. They do so mainly by influencing translation initiation or premature termination of transcription. The class of metabolite-sensing riboswitches couples the detection of specific ligands to ribozyme activity (Ramesh and Winkler 2014; Winkler 2005). A well-studied member of this group is the *glmS* riboswitch, which triggers self-cleavage upon binding of glucosamine-6-phosphate. The following degradation of the cleaved products by RNases turned out to be different in *E. coli* compared to other bacteria. In the end, expression of the *glmS* gene is reduced by regulation of mRNA stability (Collins et al. 2007; Ramesh and Winkler 2014; Winkler et al. 2004).

Riboswitches are *cis*-acting regulatory RNAs, which are binding to intracellular metabolites and thereby regulating gene expression. They are structural elements, which are usually occurring in the 5′ UTRs of mRNAs (Tucker and Breaker 2005). Research efforts have been mainly focused on riboswitches in bacteria, although they are also occurring in other organisms, e.g., thiamine pyrophosphate binding riboswitches in plants and fungi (Tucker and Breaker 2005; Kubodera et al. 2003; Sudarsan et al. 2003).

A detailed review about bacterial riboswitches and their role in regulation of gene expression and possible applications has been published by Winkler and Breaker (2005).

The structure of riboswitches can be divided into two main parts, namely, the aptamer and the expression platform, which is located downstream of the aptamer. The aptamer domain contains the sequence that binds to the metabolite and shows a very high degree of sequence conservation even within diverse organisms. Binding between the target metabolite and the aptamer causes a conformational change in the expression platform domain. These changes in conformation result then in different expression levels. The expression platforms show a high diversity with regard to their sequence, length, and structure. The fact that riboswitches are most of the time located upstream of the genes, which are coding for the synthesis or transport of the metabolite they are binding, can be exploited for the identification of target metabolites and the function of new genes (Winkler and Breaker 2005).

Based on the principle of riboswitches, Durand et al. have recently applied aptamers as biosensors for the detection of small ligands. These so-called

aptaswitches fold into a hairpin structure upon binding of the ligand. When no ligand is present, the aptamer is in its unfolded state. This structural switch, which depends on the absence or presence of the ligand, allows the application of aptaswitches as biosensors. In the folded state, a second hairpin recognizes the formed, apical loop and a kissing complex is formed through loop-loop interactions. The quantitative and specific detection of ligands by aptakiss-aptaswitch complexes was demonstrated successfully for GTP and adenosine. The development of such sensors based on hairpin aptamers can potentially be applied for any molecule with known hairpin aptamers, provided that the apical loop is not responsible for binding the ligand. The rationally designed, synthetic kissing loops could be combined with natural occurring kissing loops, which are involved in the regulation of different biological processes, and may in future also be useful for multiplexed analysis (Durand et al. 2014).

After the successful construction of riboswitches for translational regulation, they have also been engineered for transcriptional regulation. The theophylline aptamer was employed as sensor and the actuator part consisted of RNA sequences, which fold into functional intrinsic terminator structures. This concept allowed ligand-dependent regulation of gene expression by de novo design of synthetic riboswitches which influence transcriptional termination (Wachsmuth et al. 2013).

Recently, a novel approach was developed for the prediction of riboswitches in DNA sequences by a computational tool with high sensitivity and specificity called Denison Riboswitch Detector (Havill et al. 2014).

The knowledge and results of all these studies about ribozymes and riboswitches can provide the foundation for future intentions toward the design of synthetic, tailor-made riboswitches.

2.6 Small RNAs

2.6.1 Detection, Prediction, and Classification of Small RNAs

Small RNAs (sRNAs) play important and multifaceted roles in prokaryotes as well as eukaryotes. In prokaryotes, sRNAs are involved in the tagging of proteins destined for degradation and influence the activity of RNA polymerase and translation. The first bacterial sRNAs have been detected unintentionally by direct analysis of highly abundant sRNAs or during analysis of proteins or activities related to overexpression of genomic fragments. In order to get a more detailed and comprehensive insight into the role of sRNAs, studies for their systematic prediction were performed. A major challenge concerning the detection of sRNAs is that they lack conserved, characteristic sequence motifs that allow their identification (Wassarman et al. 2001). Furthermore, the discrimination between small, non-translated RNAs and random sequences is not possible solely based on secondary structural elements (Rivas and Eddy 2000; Argaman et al. 2001).

Three different approaches for sRNA prediction were applied simultaneously. Computational predictions based on transcription signals and genomic features of

already known sRNAs were used by Argaman et al. (Argaman et al. 2001). Wassarman and coworkers combined genome-wide computer searches using parameters identified in known sRNAs, genomic microarrays, and isolation of sRNAs associated with RNA-binding proteins (Wassarman et al. 2001). The identified sRNAs were experimentally confirmed and overlap with the small, noncoding RNAs predicted by Rivas et al. using a computational comparative genomic screen. Intergenic sequences of *E. coli* were analyzed and sequence data from four related bacterial strains were compared. Noncoding RNAs can have regulatory, structural, or catalytic roles, but in contrast to protein coding sequences, they lack inherent statistical biases and are therefore harder to predict. Position-specific mutational models have been applied to discriminate between probable coding regions, structural RNAs, and "other" sequences in pairwise alignments (Rivas et al. 2001). The sRNAs identified by the different methods varied in their length from 50 to several hundred nucleotides.

Apart from the noncoding, small RNAs described above, the term sRNAs is frequently used to refer to very short, usually 20–30-nucleotide-long RNAs, which are important for the regulation of gene expression and genome stability. Small RNAs can be divided into at least three classes. Depending on their mechanism, their localization within the cell, and the origin of the involved RNA molecule, they can be classified as short interfering RNAs (siRNAs), microRNAs (miRNAs), or PIWI-interacting RNAs (piRNAs) (Moazed 2009).

siRNAs and miRNAs are typically 21–25 nucleotides long, whereas piRNAs are with 24–31 nucleotides on average a bit longer. piRNAs are important components for defense mechanisms against parasitic DNA sequences and may also play a role in silencing of homologous genes. All three of them seem to be involved in posttranscriptional gene silencing (PTGS) as well as chromatin-dependent gene silencing (CDGS), which can be further divided into transcriptional and co-transcriptional gene silencing. These mechanisms illustrate that gene silencing by sRNAs can occur on the level of mRNA translation or stability and on the chromatin and DNA level. Interestingly, these mechanisms and their effect on chromatin regulation seem to be highly conserved among eukaryotes, except from *S. cerevisiae* (Moazed 2009). A very detailed description of the RNA processing pathways, the origin of different small RNA classes, and their role in chromatin silencing can be found in the review of D. Moazed.

2.6.2 Small RNA Processing Pathways

Initially, RNA interference (RNAi) was described for *C. elegans*, where injected dsRNA caused silencing of homologous host genes in the animal as well as their progeny. Notably, the effect of single-stranded sense and antisense RNA on gene expression was much lower compared to dsRNA (Fire et al. 1998). Injection of the dsRNA into the extracellular body cavity led to the spreading of the interference throughout a broad region of the organism. The interference effect was also observed when *C. elegans* larvae fed on *E. coli* bacteria, which express the

corresponding dsRNA (Timmons and Fire 1998). In later studies with HeLa cells, it was shown that short single-stranded 5′-phosphorylated antisense siRNAs can trigger gene silencing as well, since they are able to enter the mammalian RNAi pathway in vitro and in vivo (Martinez et al. 2002).

A ribonuclease III enzyme, called Dicer, is responsible for the generation of siRNAs and miRNAs by cleavage of precursor double-stranded RNAs (dsRNAs). Effector complexes, the so-called RNA-induced silencing complex (RISC) and RNA-induced transcriptional silencing complex (RITS), are involved in base pairing of the sRNA with the homologous target sequence (Hammond et al. 2000; Verdel et al. 2004; reviewed in Moazed 2009). Base pairing with the target mRNA can result in cleavage or degradation of that mRNA, so-called RNAi (Fire et al. 1998; Martinez et al. 2002).

The essential components of the silencing complexes are the so-called Argonaute proteins, which are binding to the guide sRNAs. The family of Argonaute proteins can be split into two clades, those resembling *Arabidopsis* AGO1 and those more similar to *Drosophila* Piwi. The Argonaute proteins are highly basic and contain two different domains, the C-terminal PIWI domain and PAZ domains, named after the Piwi, Argo, and Zwille/Pinhead proteins containing this domain (Cerutti et al. 2000; Carmell et al. 2002). The PIWI domain, which resembles ribonuclease H, is responsible for binding the sRNA at its 5′ end and provides the endonuclease activity. The PAZ domain, on the other hand, is involved in binding the 3′ end of the sRNA and probably in positioning the recognition and cleavage of the mRNA target (Song et al. 2004; Zamore and Haley 2005). The mechanism of target RNA cleavage is not only seen as the siRNA and RNAi mode but also as an important way of gene silencing by plant miRNAs as well as sometimes animal and viral miRNAs (reviewed in Zamore and Haley 2005).

piRNAs, on the other hand, associate with Argonaute proteins of the Piwi clade and are mostly targeting transposable elements of metazoan genomes. piRNAs recognize and repress these transposable elements and also memorize them. They are much more diverse than miRNAs and it seems that they can originate from any sequence that is located within a piRNA cluster region and processed via multiple enzymatic steps. The piRNA clusters provide information about foreign genes that have to be silenced and explain why piRNAs can be seen in a way as an immune system (Stuwe et al. 2014).

More information about the RNAi mechanism and its application for targeted gene knockdown can be found in Chap. 3 about genome engineering methods by A. Weninger, M. Killinger, and T. Vogl.

The general mechanism of the miRNA pathway of plants and animals is conserved and involves the RNaseIII enzymes Dicer and Drosha, although the latter of them is not occurring in plants (Moazed 2009). These dsRNA-specific endonucleases are responsible for processing the long, widely unstructured precursor RNAs (pre-miRNAs) to mature, single-stranded miRNAs. They do so by cutting out ~70-nt-long hairpin structures of the pre-miRNA. There are certain criteria by which miRNAs can be identified and distinguished from other small RNAs, such as confirmation of their expression by hybridization assays, their

structure and phylogenetic conservation, and their accumulation due to reduced Dicer function. In contrast to miRNAs, siRNAs originate from dsRNAs with hundreds or thousands of nucleotides in length and are created by successive cleavage (Ambros et al. 2003; Kim 2005).

miRNAs and siRNAs can cause gene silencing by suppressing mRNA translation or cleavage of the mRNA of the target gene (Zeng et al. 2003; reviewed in Rana 2007). Cleavage of fully complementary mRNA target sites was previously seen as a characteristic of siRNA-induced RNAi. Downregulation of expression, on the other hand, was seen as a characteristic of miRNAs. Following studies however indicated that miRNAs and siRNAs are functionally interchangeable and able to use the same mechanisms for mRNA degradation and mRNA translation inhibition. It turned out that the main feature that determines which mechanism is carried out is the degree of complementarity with the target mRNA. Fully complementary sequences cause mRNA cleavage, whereas mismatches result in the formation of central bulges and consequently in translational inhibition (Zeng et al. 2003).

An alternative mechanism for gene silencing by miRNAs is based on enhancing mRNA degradation. It is independent of slicer activity and requires only partial base pairing. This mechanism emphasizes the importance of mRNA stability in miRNA pathways (Bagga et al. 2005).

Interestingly, only six or seven nucleotides of the sRNAs are decisive for the main binding specificity of an sRNA. This part is therefore called "seed sequence" (Yekta et al. 2004). The 5′ end of the sRNA is contributing disproportionally to the binding of the target RNA, whereas the first nucleotide of the sRNA seems to remain unpaired (reviewed in Zamore and Haley 2005).

The interaction of these tiny sRNAs with their target relies on binding by Argonaute family proteins and is different from the mechanism of antisense oligonucleotide-target RNA pairing. The recognition sites of the sRNAs occur randomly every ~4000–65,000 nt. Upon binding of the target, the sRNA directs cleavage of a phosphodiester bond in the target RNA between the nucleotides corresponding to the middle of the guide sRNA. This cleavage requires binding of the appropriate Argonaute protein and of most of the sRNA nucleotides to the target RNA as well as the formation of at least one turn of an A-form helix. As a result, cleavage is more specific than sRNA binding itself (reviewed in Zamore and Haley 2005).

miRNAs are frequently targeting key transcription factors important for cellular identities. Due to the fact that expression of the miRNAs is regulated by transcription factors, they can be used to design diverse feedback loops (Stuwe et al. 2014).

2.6.3 Functions and Applications of Small RNAs

The wide range of functions of miRNAs in various regulatory pathways is outstanding, especially considering their tiny length of ~22 nt (Kim 2005). Their ability to act as posttranscriptional repressors by specific binding to the 3′UTRs of their target mRNA is just one example (Reinhart et al. 2000; reviewed in Ambros 2004). A

miRNA of *Drosophila* was found to be involved in regulation of apoptosis, cell proliferation, and tissue formation in a temporally and spatially regulated manner (Brennecke et al. 2003). Animal miRNAs are furthermore involved in hematopoiesis and neuronal patterning (reviewed in Kim 2005 and Ambros 2004). A miRNA from *C. elegans* named *lsy-6* has been shown to control the left/right asymmetric expression of genes in two chemosensory neurons. The miRNAs regulate the laterality of the chemosensory system of the nematode in a sequential and asymmetrical way. This sensory system enables the worm to discriminate between different attractive and repellent external, chemical stimuli. The miRNA produced from the *lsy-6* gene functions by repression of a downstream transcription factor, the so-called COG-1 transcription factor, through binding to a partially complementary sequence within the 3′UTR sequence of the *cog-1* mRNA (Johnston and Hobert 2003; Chang et al. 2004; reviewed in Ambros 2004).

Small RNAs are involved in regulation of gene expression and also genome stability. They have been shown to direct chromatin-modifying complexes to specific chromosome regions through interactions with nascent chromatin-bound ncRNAs (Moazed 2009).

Recent studies with flies and worms demonstrated that small RNAs can be involved in cellular memory and transgenerational inheritance, either in cooperation with chromatin modifications or independently (Stuwe et al. 2014).

Furthermore, it was shown recently that the plant RNAi machinery can be exploited by *Botrytis cinerea* to transfer "virulent" sRNA into the host cells. This fungal pathogen causes the gray mold disease, which can lead to severe impairments of many important agricultural crops. Bc-sRNAs can bind to the AGO proteins of the *Arabidopsis* or tomato plants and cause gene silencing of host genes with complementary target sequences. Detailed analysis of the effected genes revealed that the Bc-sRNAs target predominantly host immunity genes. Host gene silencing was not observed when the complementary sequences of the target genes were mutated and when AGO1 of the plant was knocked out. Suppression of host immunity genes was also abolished when the DCL genes of *B. cinerea*, which are involved in sRNA processing, were knocked out. These results support the suggested hijacking mechanism, by which sRNAs of the pathogen can achieve infection through suppression of host immunity genes (Weiberg et al. 2013).

2.7 Long Noncoding RNAs

In contrast to short- and mid-sized RNAs, long noncoding RNAs (lncRNAs) are more than 200 bp long. lncRNAs include transcribed ultraconserved regions (T-UCRs) as well as large intergenic noncoding RNAs (lincRNAs). They are involved in up- and downregulation of gene expression and chromatin architecture and in tumorigenesis and different neurological and cardiovascular diseases. The

locations, functions, and characteristics of the different ncRNA classes are described in more detail in several reviews (Esteller 2011; Wahlestedt 2013).

2.8 Aptamers and Adaptamers

In order to create a generic way for the formation of aptamers, which are binding to two target proteins, James et al. mixed two engineered aptamers. These two aptamers efficiently formed hybrid molecules, so-called adaptamers, which are able to bind two ligands simultaneously. The system was tested for the binding of streptavidin and a second target protein and widens the applicability of streptavidin-biotin-based detection systems (Tahiri-Alaoui et al. 2002).

Aptamers, the basic building blocks of adaptamers, are DNA and RNA molecules that are very selectively binding their target molecules. Their name is derived from the Latin word "aptus," fitting, referring to a nucleotide polymer, which fits to its target (Ellington and Szostak 1990). The development of aptamers was achieved by in vitro selection studies where random sequence pools were evolved for high binding affinities to target ligands, using the so-called SELEX (systematic evolution of ligands by exponential enrichment) procedure (Tuerk and Gold 1990; Ellington and Szostak 1990; Hermann and Patel 2000).

Aptamers display an outstanding versatility regarding possible target molecules, which include proteins, drugs, whole cells, or small organic molecules and metal ions. A major advantage of aptamers is their high affinity, which permits their use for biomedical applications like targeted drug delivery or analytics. A very interesting research field focuses on the combination of aptamers with nanoparticles, which are frequently used for bioimaging in cancer diagnostics and treatment. Thereby, the specific binding of the aptamer to the target molecule improves the binding of the nanoparticle. Nevertheless, aptamers for target molecules in medicine are rare and their field of application is therefore restricted (Reinemann and Strehlitz 2014).

The prominent specificity of aptamers is grounded on their highly optimized three-dimensional structures for recognition of their target molecule. A single methyl group difference is enough for theophylline-binding RNA aptamers to bind their target theophylline 10,000 times stronger than caffeine (Jenison et al. 1994). Several different types of interactions contribute to the molecular recognition. Stacking and hydrogen-bonding interactions are, e.g., involved in the complex formation between aptamers and flat, aromatic ligands (Hermann and Patel 2000). Further interactions, which are important for the high specificity of aptamer binding, are based on molecular shape complementarity. Structural electrostatic complementarity arises from positively charged ligands and negatively charged RNA molecules (Hermann and Patel 2000; Tor et al. 1998).

Small molecules and their RNA aptamers have been used successfully for the regulation of eukaryotic gene expression in living cells. To this end, small-molecule

aptamers were inserted into the 5′ untranslated region of a mammalian β-galactosidase mRNA and expressed in Chinese hamster ovary cells. In absence of the corresponding drug, no effect on expression was observed, whereas addition of the drug binding the aptamer inhibited β-galactosidase activity by more than 90 % (Werstuck and Green 1998).

The ability of aptamers to bind to bacterial cell surfaces was exploited in combination with quantum dots for the detection of bacteria. Therefore, the fluorescence emission of quantum dots was measured, which shifts upon binding to bacterial surfaces via DNA aptamers. The aptamers accomplished the role of antibodies, which can be used for the same application as well but they are significantly larger (Dwarakanath et al. 2004).

Research focused on aptamers allowed insight into intermolecular recognition and showed that they are very valuable and promising tools for molecular sensors and switches (Hermann and Patel 2000).

Furthermore, another kind of adaptamers can be an extremely useful tool for genome engineering, as it was shown, e.g., for the disruption of genes in *S. cerevisiae*. Here, the term adaptamer is used for primers with specific 5′ fusion tags, which allow the generic combination of DNA elements by PCR, due to the annealing of the attached adaptamers. A set of intergenic adaptamers is commercially available from Research Genetics containing primers with such 5′ sequence tags, which are not homologous to endogenous yeast DNA. The method (see Fig. 2.4) starts with the PCR amplification of the intergenic regions flanking the gene, which should be knocked out, with intergenic adaptamers. An appropriate selectable marker is PCR amplified in form of two overlapping fragments by adding the complement, reverse adaptamer tags. The two intergenic fragments and the marker fragments are fused by PCR. Thereby, two fusion segments are obtained which are co-transformed in yeast, where they recombine with genomic DNA and consequently disrupt the selected gene. Direct repeats, which are flanking the selectable marker, facilitate the removal and future reuse of the marker by recombination. The disruption of genes based on PCR and adaptamers provides a fast, efficient, and versatile tool, which can be used to study any gene disruption of interest and to increase the knowledge about gene functions in yeast (Erdeniz et al. 1997; Reid et al. 2002).

Genome modifications based on adaptamers and bipartite gene-targeting substrates were successfully implemented also in *Aspergillus nidulans*, a filamentous fungus, which shows mainly random integration of foreign DNA. The method applied by Mortensen et al. is very flexible and can reduce the amount of primers and PCRs needed for genome modifications and therefore the costs. Other advantages are a low number of false positives and the possibility to recycle the selectable marker, so that multiple genome modifications can be performed (Nielsen et al. 2006).

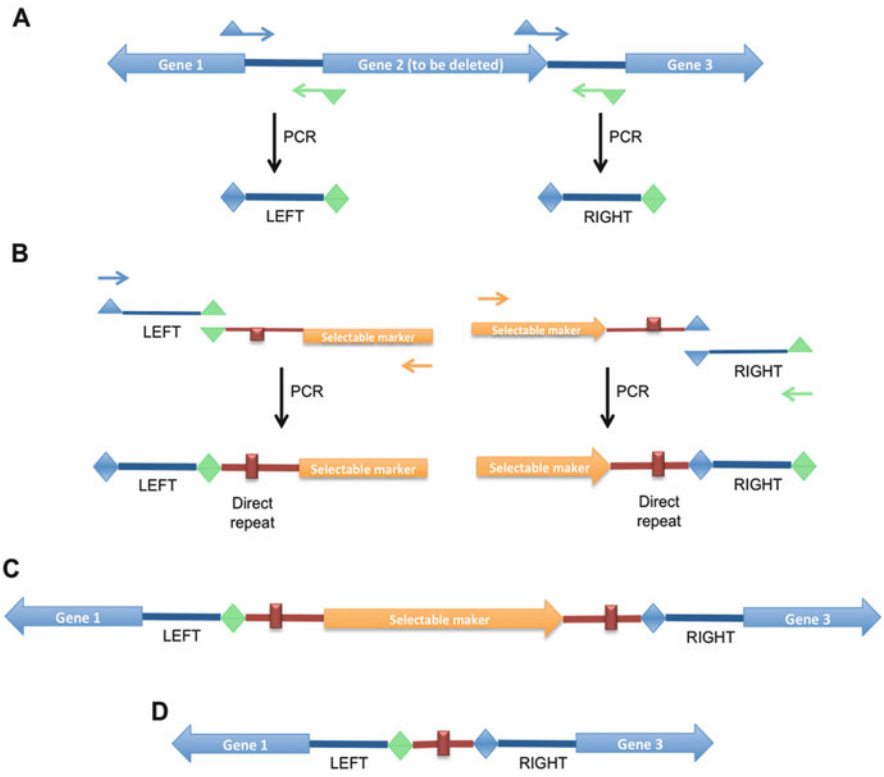


Fig. 2.4 Adaptamer-directed gene disruptions. In (**a**) the PCR, amplification of the intergenic regions flanking the gene, which is going to be disrupted in the genome, is shown. Adaptamers, depicted as *blue* and *green arrows* with *triangles*, are used to add adaptamer tags to the intergenic regions. The obtained PCR products are combined with a selectable marker (e.g., *K. lactis* URA3) by overlap extension PCR as illustrated in (**b**). Transformation of the two fusion DNA fragments results in recombination with genomic DNA and gene disruption as illustrated in (**c**). The original gene is replaced by the selectable marker and flanking direct repeats. Upon recombination of the direct repeats, the selectable marker is kicked out, resulting in the genome structure shown in (**d**). In the case of *K. lactis* URA3, marker-free constructs can be selected on 5-FOA medium, allowing the reuse of the marker for further genetic modifications. Figure adapted from Reid et al. (2002)

2.9 DNA Barcodes

DNA sequences can be employed as "barcodes," which facilitate on the one hand the assignment of unknown specimens to species and on the other hand the discovery and identification of new, otherwise inaccessible species. The mitochondrial cytochrome *c* oxidase I (COI) gene turned out to be a suitable reference for species identification based on COI profiles (Hebert et al. 2003; Frézal and Leblois 2008).

The great capability of DNA synthesis, far beyond the size of expression cassettes or plasmids, was demonstrated in 2008 when the chemical synthesis of a whole genome was published by the J. Craig Venter Institute. In the course of the synthesis, assembly and cloning of the 582,970-bp genome of the bacteria *Mycoplasma genitalium* short "watermark" sequences were inserted. These watermarks were inserted at intergenic sites to minimize biological effects and they enabled the clear differentiation between the synthetic and the native genome (Gibson et al. 2008). The first complete chemical synthesis of a bacterial genome represented an important milestone in synthetic biology (Gibson et al. 2008). Two years later the genomes of *Mycoplasma genitalium* and two other bacteria were cloned in the yeast *S. cerevisiae* as single-DNA molecules (Benders et al. 2010). Recently, the first total synthesis of a functional designer eukaryotic chromosome was achieved (Annaluru et al. 2014).

In addition to their use the labeling or identification of chromosomes or genomes watermarks or barcodes can also be employed on a smaller scale, e.g., for the labeling of plasmids and DNA sequences, e.g., in next-generation sequencing experiments. For example, unique 20-bp-long "molecular barcodes" have been furthermore employed for the identification of *S. cerevisiae* deletion strains (Giaever 2002). Barcodes can be added by primers and used to identify, e.g., hits of a promoter library by 454 pyrosequencing (Kinney et al. 2010) or for deep sequencing of barcoded mRNAs (Patwardhan et al. 2009; Melnikov et al. 2012).

2.10 DNA Machines

The reason why it is possible to build machines made from DNA lies in the highly specific interactions between complementary nucleotides. As a consequence, two-dimensional and complex three-dimensional DNA structures can be constructed based on the base sequences and the formation of branched helices (Seeman 2003; Bath and Turberfield 2007; Seeman 2010). An important characteristic of these nanoscale architectures is their self-assembling nature. DNA molecules can be used as scaffolds for the periodic assembly of molecules with possible applications for memory devices and DNA-based computing (Seeman 1998). The structures, which can be formed, are becoming more and more complex and advanced from cubes (Chen and Seeman 1991) and octahedrons (Shih et al. 2004) to multifaceted DNA origami structures such as five-pointed stars (Rothemund 2006).

The next step toward nanorobotics was the development of dynamic nanodevices from static DNA structures. These include, e.g., boxes and pinching devices, which can be used to detect molecules with an extremely large-sized range from metal ions to whole proteins (Kuzuya and Ohya 2014).

It is important to keep in mind that the nanomechanical movements of the devices are defined by their nucleotide sequence. As a consequence, DNA nanomachines are programmable and useful for highly diverse applications. Very interesting examples are sequence-dependent rotatory devices which function in a

cyclic manner (Yan et al. 2002), DNA walkers (Tian et al. 2005; Sherman and Seeman 2004; Shin and Pierce 2004; Yin et al. 2004), and DNA tweezers (Landon et al. 2012; Yurke et al. 2000). Their movements range from relatively simple conformational changes like opening/closure or rotation to complex walking step sequences (Tian et al. 2005).

However, a major limitation of early nanomachines was that in contrast to macroscale machines, they required human interference after each step (Sherman and Seeman 2004; Shin and Pierce 2004). Subsequently, autonomous machines, in the sense of self-contained devices, which are independent of human interference, have been established. An example for such a device is the nanomotor from Mao et al., which consumes chemical energy for autonomous motion. It can walk in two directions, thereby destroying its track. Compared to protein-based motors, which move along straight tracks, their DNA counterparts are very slow but more versatile (Tian et al. 2005).

The applications of nanomachines are highly diverse and range from sensors to optoelectronic devices and biopharmaceutical purposes. DNA origami "sheath," which imitates transcriptional suppressors, can be applied for controlling expression, whereas clamshell-like nanodevices allow the differentiation of cell lines by logic gates (Endo et al. 2012; Douglas et al. 2012; Kuzuya and Ohya 2014).

Recently, a DNA nanorobot was developed, which can transport cargo loads to specific cells and unload its charge after conditional, triggered activation and structural reconfiguration. Its function is controlled by different logical AND gates. The nanorobot is shaped as a hexagonal barrel and has two pairs of partially complementary lock strands. These lock strands contain an aptamer, which recognizes targets, such as cell line-specific antigens. Selective strand displacement causes the release of the loading at the target site. The applicability of these DNA nanorobots was demonstrated by the transport of fluorescent antibody fragments to the antigens on human cells. Unloading of the robot led to the fluorescent labeling of the specific cells (Douglas et al. 2012; Kuzuya and Ohya 2014).

DNA pliers have been shown to be some of the most versatile instruments of the DNA origami toolbox. They contain two 170-nm-long levers with a Holliday junction in between. These single-molecule beacons can be used for the detection of biomolecules by three different mechanisms. The first mechanism is based on pinching for the detection of target molecules, which are binding to ligands in the jaw. This method was demonstrated with biotin molecules serving as ligands and closing of the plier upon binding of streptavidin. In order to be able to detect molecules with weaker interactions compared to very strong protein-ligand interactions, a second zipping mechanism was developed. Zipping involves several elements in the levers, which are collectively binding together upon target addition, and allows the detection of, e.g., Na^+ ions. The reverse reaction is called unzipping and represents the third mechanism. Thereby, the initially closed plier is opened when target molecules, such as human microRNAs, are present (Kuzuya et al. 2011; Kuzuya and Ohya 2014).

Interestingly, this unzipping mechanism can also be used for the detection of specific binding modes such as the invasive binding of peptide nucleic acids in

DNA duplexes (Yamazaki et al. 2012). The transition between the opened and closed state of the pliers can be monitored in real time by labeling with fluorescent dyes or simply by agarose gel electrophoresis (Kuzuya and Ohya 2014).

DNA origami technology seems to have started a new epoch in structural DNA nanotechnology. In 2009, the first three-dimensional, hollow structures, comprising boxes, tetrahedrons, and prisms, were created. Advantages of DNA origami structures compared to conventional DNA nanomachines are their increased assembly yield and their ability to precisely assemble molecules with different functional groups. Furthermore, they are large enough to be detected by atomic force microscopy or transmission electron microscopy (Kuzuya and Ohya 2014).

At first sight, it does not seem to be logic to choose DNA as building material for machines since their catalytic capacity and structural versatility are lower compared to proteins or RNA. However, it is exactly this simplicity of DNA structures and interactions that facilitates researchers to predict their assembly and behavior and enables their use for nanomachines (Bath and Turberfield 2007).

2.11 DNA Walker

Precise intracellular transport along nanostructures represents a substantial difficulty, which was addressed by the construction of synthetic DNA walkers. The first DNA walkers have been designed in 2004 by the groups of Pierce et al., Seeman et al., and Reif et al. in parallel. The walkers, which were constructed at the beginning, moved in an inchworm-type gait, with one leg trailing the other (Sherman and Seeman 2004).

The group of Reif et al. designed a unidirectional and autonomous DNA motor, powered by ATP hydrolysis. The walker consists of a six-nucleotide DNA fragment, which is ligated to anchorages on a track and then released by a restriction endonuclease. Thereby, the walker may serve not only as a carrier of information but also of matter, such as nanoparticles (Yin et al. 2004).

The next step in the development of walking nanodevices consisted in bipedal DNA walkers, which are capable of moving forward by putting 1 f. in front of the other (Shin and Pierce 2004; Yin et al. 2004). The approach by Pierce et al. (see Fig. 2.5) consists of a walker, made of two partially complementary DNA strands with a double-stranded helix and two single-stranded legs. The legs can bind in an alternating manner to the protruding single-stranded branches of the track. Therefore, attachment fuel strands are used, which facilitate the anchoring by helix formation. After binding of both legs, the trailing leg is released from the track via displacement by the detachment fuel strand. The movement of the walker can be monitored by fluorescence measurements since the legs of the walker are marked with quenchers, whereas the ends of the branches are marked with various dyes. This allows real-time monitoring by multiplexed fluorescence quenching measurements (Shin and Pierce 2004).

Robots on single-molecule level represent an innovative and fascinating research area. A major challenge thereby is finding a way to store complex

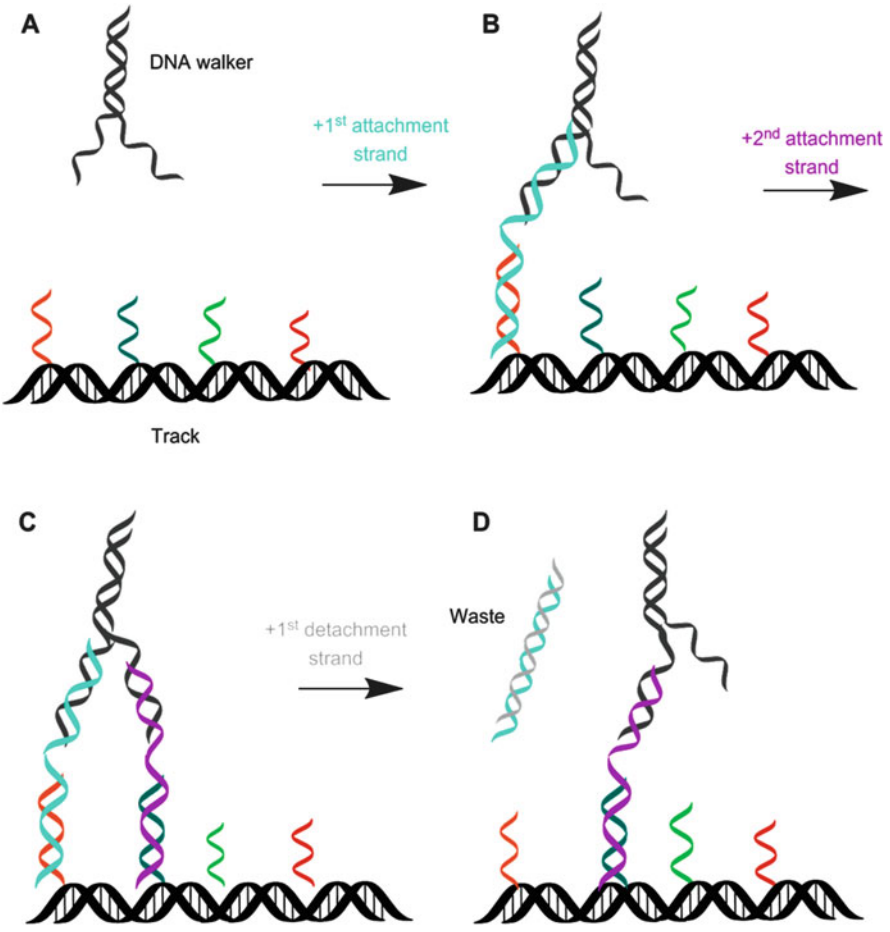


Fig. 2.5 Schematic drawing of the movement of a DNA walker. The *orange*, *dark green*, *green*, and *red* single-stranded branches represent dyes, and the *dark gray* strands of the walker represent quenchers for detection of walker locomotion. In (**a**), the unbound DNA walker is shown. Addition of the first attachment strand (*light blue*) results in the attachment of the walker to the first branch on the track, as depicted in (**b**). Upon addition of the second attachment strand (*pink*), the walker attaches to the first two branches with both legs before the first branch is released in form of duplex waste through addition of a detachment strand (*light gray*). Figure adapted from Shin and Pierce (2004)

information in individual molecules and to do a programming. In the examples mentioned before, the interaction of simple robots with their environment was utilized to create devices, which travel in a directional way along short, one-dimensional tracks. Lund et al. could show robotic behavior for so-called molecular spiders, made of an inert streptavidin molecule, which represents the body, and three legs consisting of deoxyribozymes, adapted from the 8-17 DNA enzyme. In contrast to the previously described one-dimensional movement,

spiders are able to move across two-dimensional DNA origami landscapes (Lund et al. 2010). These origami landscapes are self-assembling and consist of a long single-stranded scaffold and short oligonucleotide staple strands, which hold the scaffold in place (Rothemund 2006). The landscapes can be shaped as desired. Thus, they can be designed in such a way that the molecular spiders move across it, thereby performing a series of actions such as "start," "follow," "turn," and "stop." The movement of individual spiders was monitored in real time by super-resolution fluorescence video microscopy. The spider is positioned on a start site by a 20-base single-stranded DNA oligonucleotide and released by a single-stranded DNA trigger. Furthermore, the cofactor Zn^{2+} is added to facilitate the cleavage by the 8-17 deoxyribozyme. This enzyme cleaves at an RNA base within the substrate and leads to the formation of two shorter product fragments and the release of a leg, which can then bind to the next substrate. A crucial factor, which is essential for the concept of the molecular spider and provides a simple memory mechanism, is the lower enzyme affinity of the product compared to the substrate. When the deoxyribozyme of the leg of a spider binds to a place where it has been before, it dissociates faster from it than from a new substrate, where it stays bound longer before it finally cleaves it. Consequently, a spider, released at the boundary between products and substrates, moves toward the substrates and follows a linear, directional track during substrate cleavage (Lund et al. 2010).

Previous nanomotors have been mostly based on burn-the-bridge methods, which provide directionality by chemically damaging the traversed track, for example, a nanomotor driven by a nicking enzyme for the transport of DNA cargo (Bath et al. 2005). In contrast to this DNA motor and the DNA walkers described before (Sherman and Seeman 2004; Shin and Pierce 2004), spiders can take Brownian walks across already visited product sites until they run into the next substrates.

As another alternative to burn-the-bridge methods, a DNA motor, based on a bioinspired concept, was recently established, using mechanics-mediated symmetry breaking. The technology relies on local alignment with the track through binding of a pedal and achieves directionality by adjusting the size of the motor. A single action of leg dissociation is enough to drive the motor. The symmetric bipedal nanomotor is able to move continuously along a track with only two different footholds. The concept is designed to be generally applicable for DNA molecules, peptides, or synthetic polymers (Cheng et al. 2014).

The average step size of DNA walkers is around 2–5 nm (Sherman and Seeman 2004; Shin and Pierce 2004). The DNA nanomotor powered by nicking enzymes is moving with a speed of 0.1 nm s^{-1} (Bath et al. 2005). In comparison to that, molecular spiders have been shown to travel around 100 nm and exhibited mean speeds of 1–6 nm s^{-1}. Although a lot of progress was already made on the field of DNA walkers, there are still several factors, which limit their performance. The traveling distance of molecular spiders is restricted by dissociation and backtracking. Other shortcomings of this concept are that the substrate has to be recharged and that molecular spiders are slower and not as efficient as protein-based walkers. However, the programmability and predictability of DNA walkers make

them attractive research targets for nanoscale robotics with defined interactions with their environment (Lund et al. 2010).

References

Alper H et al (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102(36):12678–12683

Ambros V (2004) The functions of animal microRNAs. Nature 431(7006):350–355

Ambros V et al (2003) A uniform system for microRNA annotation. RNA (New York) 9 (3):277–279

Annaluru N et al (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344(6179):55–58

Argaman L et al (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr Biol 11(12):941–950

Bagga S et al (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell 122(4):553–563

Bath J, Turberfield AJ (2007) DNA nanomachines. Nat Nanotechnol 2(5):275–284

Bath J, Green SJ, Turberfield AJ (2005) A free-running DNA motor powered by a nicking enzyme. Angewandte Chemie 117(28):4432–4435

Benders GA et al (2010) Cloning whole bacterial genomes in yeast. Nucleic Acids Res 38 (8):2558–2569

Berg L et al (2013) Combinatorial mutagenesis and selection to understand and improve yeast promoters. BioMed Res Int 2013:926985

Bernhardt HS (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). Biol Direct 7:23

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

Brennecke J et al (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell 113(1):25–36

Brewster RC, Jones DL, Phillips R (2012) Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. PLoS Comput Biol 8(12), e1002811

Cambray G et al (2013) Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res 41(9):5139–5148

Carmell MA et al (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 16(21):2733–2742

Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem Sci 25 (10):481–482

Chang S et al (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature 430(7001):785–789

Chen JH, Seeman NC (1991) Synthesis from DNA of a molecule with the connectivity of a cube. Nature 350(6319):631–633

Chen Y-J et al (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 10(7):659–664

Cheng J et al (2014) A bioinspired design principle for DNA nanomotors: mechanics-mediated symmetry breaking and experimental demonstration. Methods (San Diego) 67(2):227–233

Collins JA et al (2007) Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev 21 (24):3356–3368
Comstock J (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. Proc Natl Acad Sci U S A 80:21–25
Coussement P et al (2014) One step DNA assembly for combinatorial metabolic engineering. Metab Eng 23:70–77
Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Syst Biol 3(145):145
Crook NC, Freeman ES, Alper HS (2011) Re-engineering multicloning sites for function and convenience. Nucleic Acids Res 39(14), e92
Curran KA et al (2013) 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
Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 39(3):1131–1141
De Mey M et al (2007) Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering. BMC Biotechnol 7:34
Doudna JA, Cech TR (2002) The chemical repertoire of natural ribozymes. Nature 418 (6894):222–228
Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. Science (New York) 335(6070):831–834
Durand G et al (2014) Riboswitches based on kissing complexes for the detection of small ligands. Angewandte Chemie 53(27):6942–6945 (International ed. in English)
Dwarakanath S et al (2004) Quantum dot-antibody and aptamer conjugates shift fluorescence upon binding bacteria. Biochem Biophys Res Commun 325(3):739–743
Edelman GM et al (2000) Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity. Proc Natl Acad Sci 97(7):3038–3043
Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822
Endo M et al (2012) Transcription regulation system mediated by mechanical operation of a DNA nanostructure. J Am Chem Soc 134(6):2852–2855
Erdeniz N, Mortensen UH, Rothstein R (1997) Cloning-free PCR-based allele replacement methods. Genome Res 7(12):1174–1183
Esteller M (2011) Non-coding RNAs in human disease. Nat Rev Genet 12(12):861–874
Fedor MJ (2009) Comparative enzymology and structural biology of RNA self-cleavage. Annu Rev Biophys 38:271–299
Ferré-D'Amaré AR, Scott WG (2010) Small self-cleaving ribozymes. Cold Spring Harb Perspect Biol 2(10):a003574
Fire A et al (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391(6669):806–811
Frézal L, Leblois R (2008) Four years of DNA barcoding: Current advances and prospects. Infect Genet Evol 8(5):727–736
Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol 56(4):343–349
Gehrke S, Jérôme V, Müller R (2003) Chimeric transcriptional control units for improved liver-specific transgene expression. Gene 322:137–143
Giaever EA (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418 (6896):387–391
Gibson DG et al (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science (New York) 319(5867):1215–1220
Guazzaroni M-E, Silva-Rocha R (2014) Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators. ACS Synth Biol 3(9):666–675

Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries–tuning of gene expression. Trends Biotechnol 24(2):53–55
Hammond SM et al (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404(6775):293–296
Hartner FS et al (2008) Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res 36(12), e76
Haurwitz RE et al (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science (New York) 329(5997):1355–1358
Havill JT et al (2014) A new approach for detecting riboswitches in DNA sequences. Bioinformatics (Oxford) 30(21):3012–3019
Hebert PDN et al (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270 (1512):313–321
Hermann T, Patel DJ (2000) Adaptive recognition by nucleic acid aptamers. Science 287 (5454):820–825
Jäschke A (2001) Artificial ribozymes and deoxyribozymes. Curr Opin Struct Biol 11(3):321–326
Jenison RD et al (1994) High-resolution molecular discrimination by RNA. Science (New York) 263(5152):1425–1429
Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64(1):82–87
Jensen K et al (2006) Identifying functionally important mutations from phenotypically diverse sequence data. Appl Environ Microbiol 72(5):3696–3701
Johnston RJ, Hobert O (2003) A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature 426(6968):845–849
Juven-Gershon T, Cheng S, Kadonaga JT (2006) Rational design of a super core promoter that enhances gene expression. Nat Methods 3(11):917–922
Kagiya G et al (2005) Generation of a strong promoter for Escherichia coli from eukaryotic genome DNA. J Biotechnol 115(3):239–248
Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 6 (5):376–385
Kinney JB et al (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc Natl Acad Sci U S A 107(20):9158–9163
Kranthi BV et al (2009) Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris. Biochim Biophys Acta 1789(6-8):460–468
Kruger K et al (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell 31(1):147–157
Kubodera T et al (2003) Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. FEBS Lett 555 (3):516–520
Künne T, Swarts DC, Brouns SJJ (2014) Planting the seed: target recognition of short guide RNAs. Trends Microbiol 22(2):74–83
Kuzuya A, Ohya Y (2014) Nanomechanical molecular devices made of DNA origami. Acc Chem Res 47(6):1742–1749
Kuzuya A et al (2011) Nanomechanical DNA origami "single-molecule beacons" directly imaged by atomic force microscopy. Nat Commun 2:449
Landon PB et al (2012) DNA zipper-based tweezers. Langmuir 28(1):534–540
Li M et al (2012) A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. Microb Cell Fact 11:19
Lou C et al (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nat Biotechnol 30(11):1137–1142
Lubliner S, Keren L, Segal E (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res 41(11):5569–5581

Lund K et al (2010) Molecular robots guided by prescriptive landscapes. Nature 465 (7295):206–210
Mairhofer J et al (2013) Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions. Appl Environ Microbiol 79(12):3802–3812
Mairhofer J et al (2015) Preventing T7 RNA polymerase read-through transcription – a synthetic termination signal capable of improving bioprocess stability. ACS Synth Biol 4(3):265–273
Martinez J et al (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110(5):563–574
Melnikov A et al (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30(3):271–277
Moazed D (2009) Small RNAs in transcriptional gene silencing and genome defence. Nature 457 (7228):413–420
Moore I, Samalova M, Kurup S (2006) Transactivated and chemically inducible gene expression in plants. Plant J 45(4):651–683
Mutalik VK et al (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10(4):354–360
Naville M, Gautheret D (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. Genome Biol 11(9):R97
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
Nielsen ML et al (2006) Efficient PCR-based gene targeting with a recyclable marker for Aspergillus nidulans. Fungal Genet Biol 43(1):54–64
Nielsen AAK, Segall-Shapiro TH, Voigt CA (2013) Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. Curr Opin Chem Biol 17 (6):878–892
Nissen P (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289 (5481):920–930
Patwardhan RP et al (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol 27(12):1173–1175
Penchovsky R (2013) Computational design and biosensor applications of small molecule-sensing allosteric ribozymes. Biomacromolecules 14(4):1240–1249
Pfleger BF et al (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24(8):1027–1032
Platt T (1986) Transcription termination and the regulation of gene expression. Annu Rev Biochem 55:339–372
Prody GA et al (1986) Autolytic processing of dimeric plant virus satellite RNA. Science (New York) 231(4745):1577–1580
Ramesh A, Winkler WC (2014) Metabolite-binding ribozymes. Biochim Biophys Acta 1839 (10):989–994
Rana TM (2007) Illuminating the silence: understanding the structure and function of small RNAs. Nat Rev Mol Cell Biol 8(1):23–36
Reid RJD et al (2002) Efficient PCR-based gene disruption in Saccharomyces strains using intergenic primers. Yeast (Chichester) 19(4):319–328
Reinemann C, Strehlitz B (2014) Aptamer-modified nanoparticles and their use in cancer diagnostics and treatment. Swiss Medical Weekly 144(January):w13908
Reinhart BJ et al (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403(6772):901–906
Remans T et al (2005) Functional promoter analysis using an approach based on an in vitro evolution strategy. BioTechniques 38(2):209–10, 212, 214–216
Rivas E, Eddy SR (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. Bioinformatics 16(7):583–605

Rivas E et al (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics. Curr Biol 11(17):1369–1373
Ross W et al (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J Bacteriol 180(20):5375–5383
Roth A et al (2014) A widespread self-cleaving ribozyme class is revealed by bioinformatics. Nat Chem Biol 10(1):56–60
Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440 (7082):297–302
Ruth C, Glieder A (2010) Perspectives on synthetic promoters for biocatalysis and biotransformation. Chembiochem 11(6):761–765
Ruth C et al (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. Syst Synth Biol 4(3):181–191
Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science (New York) 289(5478):448–452
Seehafer C et al (2011) From alpaca to zebrafish: hammerhead ribozymes wherever you look. RNA (New York) 17(1):21–26
Seeman NC (1998) DNA nanotechnology: novel DNA constructions. Annu Rev Biophys Biomol Struct 27:225–248
Seeman NC (2003) DNA in a material world. Nature 421(6921):427–431
Seeman NC (2010) Nanomaterials based on DNA. Annu Rev Biochem 79:65–87
Sharon E et al (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 30(6):521–530
Sherman WB, Seeman NC (2004) A precisely controlled DNA biped walking device. Nano Lett 4 (7):1203–1207, Available at: http://pubs.acs.org/doi/abs/10.1021/nl049527q
Shih WM, Quispe JD, Joyce GF (2004) A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. Nature 427(6975):618–621
Shin JS, Pierce NA (2004) A synthetic DNA walker for molecular transport. J Am Chem Soc 126 (35):10834–10835
Solem C, Jensen PR (2002) Modulation of gene expression made easy. Appl Environ Microbiol 68 (5):2397–2403
Song J-J et al (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science (New York) 305(5689):1434–1437
Stuwe E, Tóth KF, Aravin AA (2014) Small but sturdy: small RNAs in cellular memory and epigenetics. Genes Dev 28(5):423–431
Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA (New York) 9(6):644–647
Tahiri-Alaoui A et al (2002) High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands. Nucleic Acids Res 30(10), e45
Tian Y et al (2005) A DNAzyme that walks processively and autonomously along a one-dimensional track. Angewandte Chemie 44(28):4355–4358 (International ed. in English)
Timmons L, Fire A (1998) Specific interference by ingested dsRNA. Nature 395(6705):854
Tor Y, Hermann T, Westhof E (1998) Deciphering RNA recognition: aminoglycoside binding to the hammerhead ribozyme. Chem Biol 5(11):R277–R283
Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. Curr Opin Struct Biol 15(3):342–348
Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
Tyo KEJ, Nevoigt E, Stephanopoulos G (2011) Directed evolution of promoters and tandem gene arrays for customizing RNA synthesis rates and regulation. Methods Enzymol 497:135–155
Verdel A et al (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science (New York) 303(5658):672–676
Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 24(6):1094–1101

Vogl T et al (2014) Synthetic core promoters for Pichia pastoris. ACS Synth Biol 3(3):188–191
Wachsmuth M et al (2013) De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res 41(4):2541–2551
Wahlestedt C (2013) Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat Rev Drug Discov 12(6):433–446
Walter G (1986) Origin of life: the RNA world. Nature 319(20):618
Wassarman KM et al (2001) Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev 15:1637–1651
Weiberg A et al (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science (New York) 342(6154):118–123
Werstuck G, Green MR (1998) Controlling gene expression in living cells through small molecule-RNA interactions. Science 282(5387):296–298
Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Curr Opin Chem Biol 9(6):594–602
Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59:487–517
Winkler WC et al (2004) Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428(6980):281–286
Yamazaki T et al (2012) Clear-cut observation of PNA invasion using nanomechanical DNA origami devices. Chem Commun (Camb) 48(92):11361–11363
Yan H et al (2002) A robust DNA mechanical device controlled by hybridization topology. Nature 415(6867):62–65
Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. Nature 289 (26):751–758
Yekta S, Shih I-H, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. Science (New York) 304(5670):594–596
Yin P et al (2004) A unidirectional DNA walker that moves autonomously along a track. Angewandte Chemie 43(37):4906–4911 (International ed. in English)
Yurke B et al (2000) A DNA-fuelled molecular machine made of DNA. Nature 406 (6796):605–608
Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. Science (New York) 309 (5740):1519–1524
Zeng Y, Yi R, Cullen BR (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A 100(17):9779–9784

Key Methods for Synthetic Biology: Genome Engineering and DNA Assembly

3

Astrid Weninger, Manuela Killinger, and Thomas Vogl

Contents

3.1 Genome Engineering Technologies 102
3.1.1 Non-programmable Genome Engineering 107
3.1.2 RNA Interference and Translational Repression 110
3.1.3 Prokaryotic Genome Engineering 112
3.1.4 Universally Applicable Genome Editing Strategies 114
3.2 DNA Assembly Methods 119
3.2.1 Restriction Endonuclease-Based Cloning Methods 120
3.2.2 Restriction-/Ligation-Independent Assembly Strategies 125
References 130

Abstract

The advent of potent genome engineering techniques and efficient DNA assembly methods has greatly expanded our capacities to modify genomes and have fueled metabolic engineering and synthetic biology endeavours. Here we provide an overview on genome engineering methods ranging from non-programmable tools such as site-specific recombinases to highly specific, programmable strategies as CRISPR–Cas9. Advantages and limitations of the methods for the generation of custom-designed organisms are compared. Certain methods such as group II introns and recombineering are limited to prokaryotes, whereas zinc-finger nucleases (ZFN), transcription activator-like effector

A. Weninger and M. Killinger contributed equally.

A. Weninger • T. Vogl (✉)
Institute for Molecular Biotechnology, Graz University of Technology, Petersgasse 14/2, 8010, Graz, Austria
e-mail: aweninger@student.tugraz.at; thomas.vogl@aon.at

M. Killinger
Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14/4, 8010, Graz, Austria
e-mail: manuela.killinger@acib.at

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_3

nucleases (TALENs) and CRISPR–Cas9 are universally applicable for all domains of life. We summarize also popular DNA assembly methods, suitable for generating multigene pathways for metabolic engineering and artificial constructs for synthetic biology or entire genomes.

3.1 Genome Engineering Technologies

Synthetic biology is described as engineering-driven building of organisms and biological entities with beneficial functions (Silver et al. 2014; Heinemann and Panke 2006). In the last decade, the interest of biological engineering shifted from individual genes to entire genomes. For basic research and industrial application, the generation of rationally designed organisms cannot be achieved solely by random mutagenesis and selection, but requires advanced genome engineering techniques (Esvelt and Wang 2013; Conrad et al. 2011; Elena and Lenski 2003).

Genome engineering refers to any technology and method for genome-scale modification of organisms (Esvelt and Wang 2013). Molecular engineering tools to introduce random modifications and techniques for targeted genome engineering, which address a defined, desired genomic locus, are available. Some methods modify the genome in a self-sufficient manner, whereas others rely on the host organism's cellular machinery. Genome editing is a collective term for methods taking advantage of the DNA repair mechanisms of the host cell. Host-mediated DNA repair is achieved by two different strategies and depends on the nature of the DNA strand break. In case of a double strand break, the repair is either triggered by nonhomologous end joining (NHEJ) (Lieber 2010) or by homologous recombination (HR) (Capecchi 1989). Single strand breaks are solely repaired by HR. Both mechanisms allow the introduction of genome modifications and alteration of genomic conditions. In its natural role, HR promotes the exchange of endogenous DNA sequences, but for genome engineering applications, the mechanism is hijacked to exchange genetic information between endogenous sequences and an artificially constructed exogenous DNA fragment (Dudás and Chovanec 2004). NHEJ rejoins DNA ends without the requirement of a homologous template and is often accompanied by substitutions, insertions and deletions (indels) of nucleotides in the targeted region (Barnes 2001). Targeted single and double strand break-induced DNA repair increases the rate of homologous recombination by several orders of magnitude (up to 4000-fold in yeast) (Rouet et al. 1994; Smih et al. 1995; Caldecott 2008; Storici et al. 2003). Therefore systems introducing breaks at programmable positions in the genome are highly sought after.

Early genome engineering approaches showed low efficiency rates and relied on random integration or on a limited number of predefined genomic sites. Nowadays a vast range of prokaryotic and eukaryotic genome engineering techniques (Table 3.1 and Fig. 3.1) facilitate the generation of gene knockouts, gene delivery and the introduction of gene displacements or chromosomal rearrangements (Esvelt and Wang 2013; Carr and Church 2009).

Table 3.1 Detailed comparison of genome engineering methods

Genome engineering tool	Capacities	Target binding principle	Target length	Multiplexing	Targeting frequency	Programmable[a]	Off-targeting	Advantages	Drawbacks	Improvements/ Combinatorial methods	Applications	References
Site-specific recombinases	Integration, excision, inversion and translocation relying on pre-integrated recognition sites; Selection marker recycling; Landing pads and RMCE	Recombinase recognizes a specific recognition site	typically 34 bp	–	High	No, the recombinase is restricted to the recognition sequence	Minor effects	Not dependent on host cell (co-) factors or DNA replication	Recombination only occurs at preexisting recognition sites	Group II introns in prokaryotes (Enyeart et al. 2013)	Prokaryotes; Eukaryotes	Nern et al. (2011), Sternberg et al. (1981), Proteau et al. (1986), Turan et al. (2011), Sauer (1994)
Transposons	(Close to) random mutagenesis—used for gene delivery as alternative to viral vectors; Random mutagenesis	Transposons excise sequences, which are flanked by defined recognition sequences (TIR); (Close to) non-specific (re)integration	Depending on transposase either random, dinucleotides or short palindromic consensus sequences	–	Low	Moderate	Wild-type enzymes integrate close to random; fusion proteins to ZF and TALEN DNA BDs also show high off-targeting	Collection of transposons with differences in cargo capacity, integration preferences and species specificity; Useful for random mutagenesis –	Due to high off-targeting rate little suitable for targeted genome engineering	Fusions to zinc-finger DNA binding domains (Feng et al. 2010; Yant et al. 2007) and TALE DNA binding domains allow targeted genome engineering (Owens et al. 2013)	Prokaryotes; Eukaryotes	Ivics and Izsvák (2010), Grabundzija et al. (2010), Le Breton et al. (2006), Dumais Pope et al. (1994), Fernandes et al. (2001), Weil and Kunze (2000), Parinov et al. (2004) Rushforth et al. (1993), Bessereau et al. (2001), Cooley et al. (1988)
Group II introns	Targeted integration; Induction of ds breaks and host cell repair (HR, NHEJ)	Specific protein and RNA recognition	Predefined 5 bp is required by the protein + 14–16 bp variable RNA sequence	–	Variable	High, but the target sites are limited to protein recognition sites (5 bp)	No, minor effects	Ease of retargeting	Low efficiencies in eukaryotes	Combined with transposons (Enyeart et al. 2013) Reverse transcriptase-deficient group II introns with nuclease activity	Prokaryotes	Karberg et al. (2001), Frazier et al. (2003), Rawsthorne et al. (2006)

(continued)

Table 3.1 (continued)

Genome engineering tool	Capacities	Target binding principle	Target length	Multiplexing	Targeting frequency	Programmable[a]	Off-targeting	Advantages	Drawbacks	Improvements/ Combinatorial methods	Applications	References
Recombineering systems	In vivo cloning technique; Insertions, deletions, point mutations, gene replacement	ss donor DNA fragment is annealed to complementary sequence during DNA replication	35–50 bp (length of donor fragment)	Yes, MAGE (Wang et al. 2009) CAGE (Isaacs et al. 2011)	High	High	No	Independent from host cell recombination	Negative effect on host cell replication, limited size of the inserted DNA, dependent on host cell replication	MAGE (Wang et al. 2009) CAGE (Isaacs et al. 2011)	Prokaryotes	Sharan et al. (2009), Yu et al. (2000), Mosberg et al. (2012)
RNAi, translational repression	Knock-down of genes at posttranscriptional level	RNA guided protein complex binds to mRNA	21–23 bp	Yes	High	High	High, binding to similar RNA sequences	Fast and easily reprogrammable	No knockouts, incomplete knock-downs; dsRNA can active various cellular pathways	–	Higher eukaryotes	Bernstein et al. (2001), Misquitta and Paterson (1999), Elbashir et al. (2001), Paddison et al. (2002), Brummelkamp et al. (2002), Rivas et al. (2005)
Homing endonucleases	Introduction of ds break triggers host cell repair (HR, NHEJ)	Protein recognizes specific DNA target site	12–40 bp	–	High	Low	Minor effects		Limited capacity of available target sites; engineering for new target sites is complicated and time-consuming	Fusions of preexisting nuclease domains and specifically altered DNA binding domains (Arnould et al. 2006; Seligman et al. 2002; Doyon et al. 2006; Grizot et al. 2011)	Prokaryotes; Eukaryotes	Thierry and Dujon (1992), Jasin (1996), Gimble and Wang (1996), Argast et al. (1998)

Zinc finger nucleases	Introduction of ds or ss breaks and host cell repair (HR, NHEJ)	Assembly of single protein modules of which each recognizes three DNA bases	2 × 18 bp + spacer (~15 bp)	–	Variable, depends on target locus	Moderate	Low–medium	Focus of research for many years, findings were quickly adopted to TALEN technologies	Complicated reprogramming requires protein assembly and engineering	ZF nickase (Ramirez et al. 2012), heterodimeric ZFN (Miller et al. 2007), hyperactive nuclease variants (Gordley et al. 2009; Guo et al. 2010)	Prokaryotes; Eukaryotes	Liu et al. (1997), Beerli and Barbas (2002), Bhakta et al. (2013), Gonzalez et al. (2010)
TALEN	Introduction of ds or ss breaks and host cell repair (HR, NHEJ)	Assembly of protein modules of which each recognizes one DNA base	2 × 17 bp + spacer (~15 bp)	–	Variable, depends on target locus	High	No, minor effects	Almost every locus is targetable	Large protein size; Moderate construction time	Fusions to transcription activation and repression domains (TAL–TF) (Morbitzer et al. 2010)	Prokaryotes; Eukaryotes	Cermak et al. (2011), Weber et al. (2011a), Sanjana et al. (2012), Uhde-Stone et al. (2013)
CRISPR–Cas9	Introduction of ds or ss breaks and host cell repair (HR, NHEJ)	RNA-guided protein complex binds to DNA	20 bp + PAM (3 bp)	Yes	Variable, depends on target locus	High, restricted by PAM	High	Fast and easily reprogrammable	Large protein size, several gRNAs should be tested for one locus	*Fok*I–Cas9 fusion protein (Tsai et al. 2014)	Prokaryotes; Eukaryotes	Jinek et al. (2012), Gasiunas et al. (2012), Cong et al. (2013)

[a]Altering of the target site

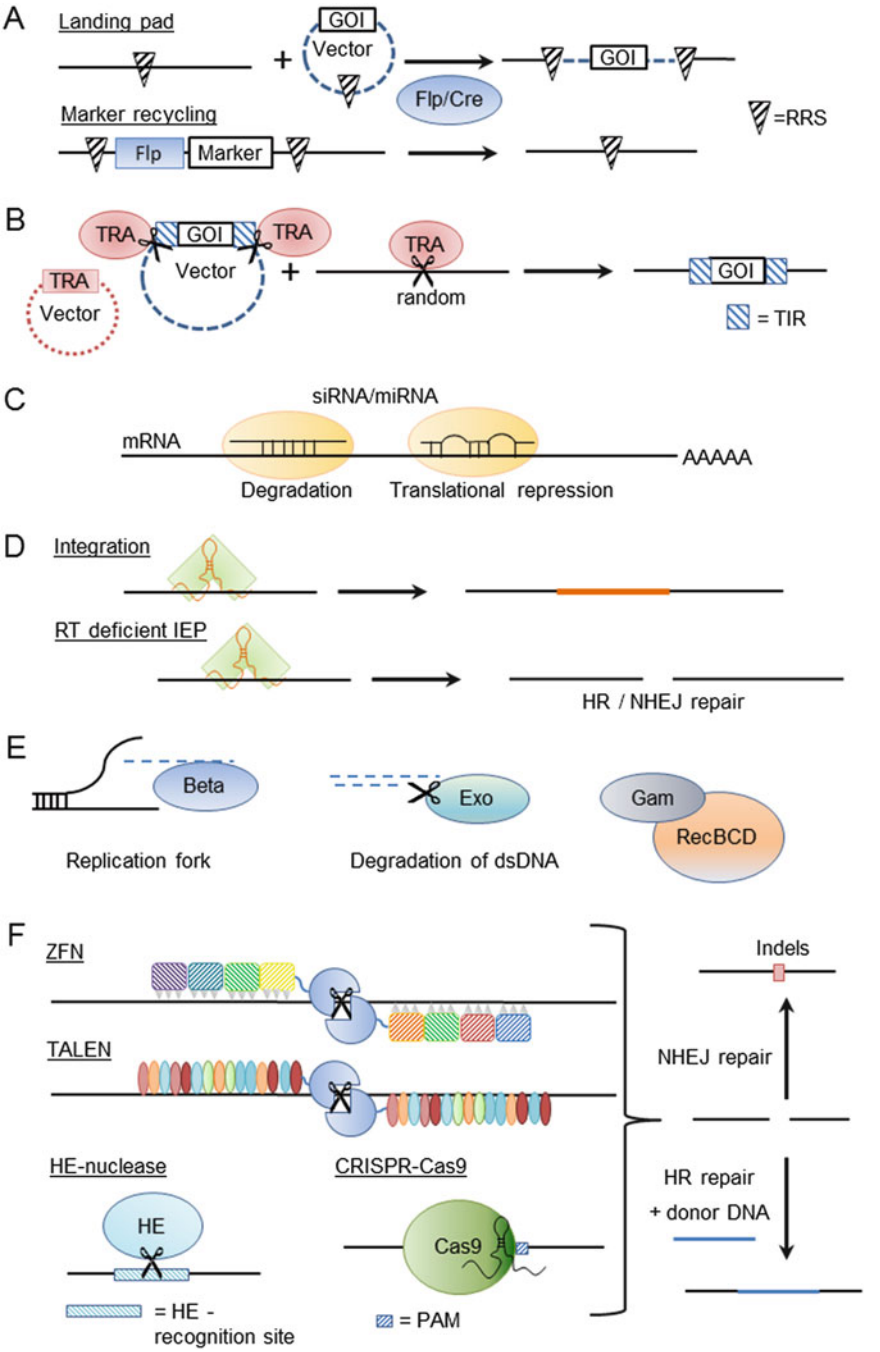


Fig. 3.1 Basic principles of genome engineering technologies. (**a**) Site-specific recombinases: Recombinase recognition sites (RRS) can be integrated as 'landing pad' into the host genome. The flippase/Cre integrase cuts at the RRSs in the genome and on the vector mediating the integration of the gene of interest. A reverse setup can be used for selection marker recycling

In the first section of this chapter, we discuss several genome engineering techniques by comparing their capabilities and limitations. Genome engineering techniques range from site-specific endonucleases (Cre/loxP, flippases) to mobile genetic elements such as transposons or ribozymes (group II Introns). Methods such as recombineering are only applicable for prokaryotes, whereas others are suitable for all domains of life. One of these techniques is genome editing, which takes advantage of engineered nucleases to cut a defined genome locus and thereby introduces a sequence change (Perez-Pinera et al. 2012). These nucleases possess either a specific affinity for a defined DNA sequence (homing endonucleases), are fused to DNA binding domains such as transcription activator-like effectors (TALEs) and zinc-finger domains or use a short guide RNA to specifically cut a DNA sequence of interest (CRISPR–Cas9).

3.1.1 Non-programmable Genome Engineering

3.1.1.1 Site-Specific Recombinases (Cre/loxP, FLP/FRT)

Site-specific recombinases catalyse DNA cleavage reactions between two identical recognition sites to inverse, excise or integrate a DNA fragment (Fig. 3.1a) (Nern et al. 2011). In contrast to programmable systems such as CRISPR–Cas9 (Jinek et al. 2012) or TALENs (Miller et al. 2011), the recognition sites of site-specific

Fig. 3.1 (continued) (Vogl et al. 2013). (**b**) Transposons: A vector expressing the transposase gene and a vector bearing the gene of interest (GOI) flanked by two transposon recognition sites (TIR) are transformed into the cell. The transposase excises the GOI and mediates its integration into the host genome. The integration occurs in a close to random manner. (**c**) RNAi: Short, processed miRNA and siRNAs guide the RNA-induced silencing complex (RISC) to bind complementary sequences on the mRNA. The degree of complementarity causes degradation or translational repression of the mRNA. (**d**) Group II introns: The intron-encoded protein (IEP, green rectangles) and the intron RNA form a complex, which scans the DNA. If base pairing between the RNA and a complementary DNA sequence can take place, the intron RNA is integrated in the genome by the IEP. In the case of a reverse transcriptase-deficient IEP, the RNA is not integrated in the genome; however a ds break is introduced, which has to be repaired by the cellular HR or NHEJ machinery. (**e**) Recombineering: The three enzymes Beta, Exo and Gam are involved in lambda Red recombineering. Beta is a ssDNA binding protein, which anneals the ss donor DNA to complementary sequences near the replication fork. If ds donor DNA molecules are transformed into the host cell, one strand is degraded by Exo. Gam inhibits the RecBCD nuclease, which is responsible for the degradation of dsDNA. (**f**) Different genome editing tools. All technologies introduce a ds DNA break that is repaired by the host endogenous repair machinery by either NHEJ or HR. Zinc-finger (ZFN) modules, of which each recognizes three DNA bases, are combined to recognize a DNA sequence of interest. A monomer of the catalytic domain of the *Fok*I nuclease is fused to the zinc finger array. If two arrays bind adjacent DNA sequences, *Fok*I can dimerize and introduce a strand break. TALE repeats recognize a single DNA base. The repeats are combined to a TALE array and a *Fok*I monomer is fused to the DNA binding domain. Again *Fok*I dimerization is required for cleavage. Homing endonucleases (HE) recognize a 12–40 bp recognition site and introduce a ds break. CRISPR–Cas9 takes advantage of a short guide RNA, requiring a PAM motif for binding

recombinases cannot be altered (non-programmable). The site-specific Cre [*C*auses *re*combination (Sternberg et al. 1981)] recombinase from bacteriophage P1 and the yeast-derived flippase (FLP) are the most extensively studied recombinases and are widely used for genome engineering (Sorrell and Kolb 2005). Cre recombination occurs between two consensus 34 bp DNA recognition sites, called loxP sites, and its applications date back to the early 1980s (Hamilton and Abremski 1984). However, cytotoxic effects due to off-target endonucleolytic activity with pseudo recombination sites in some eukaryotes shifted the interest to the FLP recombinase for eukaryotic site-specific recombination (Loonstra et al. 2001; Silver and Livingston 2014). FLP is naturally encoded on the 2-micron plasmid of *Saccharomyces cerevisiae* and promotes recombination between two identical, minimal 34 bp FLP recombination sites (FRT sites) (Proteau et al. 1986). Cre/FLP recombination does not depend on any supporting host cell factors such as topoisomerases or DNA replication machinery. Moreover the recombinases can enter the eukaryotic nucleus and perform recombination unimpaired by the chromatin structure (Sauer 1987). Recombination between two recognition sites organized in the same orientation leads to the excision of the flanked DNA, whereas recombination between target sites arranged in opposite orientations triggers the inversion of an intervening DNA fragment. If the recognition sites are located on different chromosomes, translocation events can take place (Craig 1988). Incorporation of DNA sequences only occurs if the target site is present on the donor vector and in the targeted genomic locus (Fig. 3.1a): After transformation of the recombinase expression vector and a donor plasmid into the cells, the recombinase cuts the genomic locus at the target site and also linearizes the donor vector. The donor sequence is incorporated at the target locus accompanied by a duplication of the FRT/loxP sites. A major disadvantage of this system is the requirement for pre-integration of the recognition site at the target site. Therefore, site-specific recombinases cannot be used to introduce targeted mutations in any desired locus (Cabaniols et al. 2009). Nevertheless, efficient bacterial genetic engineering was performed by combining the Cre/loxP recombinase and mobile group II introns. The introns were used to deliver the loxP sites to a specific genomic locus enabling the genomic modification by the site-specific recombinase (Enyeart et al. 2013). Another disadvantage is that the recognition sites of the recombinases are identical before and after the recombination, which facilitates the excision of the integrated fragment (Cabaniols et al. 2009). This drawback had been overcome by the use of poisoned half sites, which cannot be cleaved by the enzyme upon recombination (Schlake and Bode 1994).

Recombinase-mediated cassette exchange (RMCE) (Torres et al. 2011; Turan et al. 2011) is an integrase-based technique used for the nondisruptive insertion of a DNA cassette into a precharacterized genomic locus. A genomic destination termed ‘landing pad’ contains a selection marker cassette flanked by recombinase recognition sites. A circular donor vector containing an analogous cassette encoding the gene of interest is transformed into the host cells and is used to replace the resident cassette by the aid of the recombinase. RMCE is successfully used for predictable expression of heterologous genes in cell cultures and for the systematic generation

of transgenic animals in a selection marker-free environment (Baer and Bode 2001).

Site-specific recombinases show also a great promise for selection marker recycling [e.g. in *S. cerevisiae* (Sauer 1994) or *Pichia pastoris* (Vogl et al. 2013)] (Fig. 3.1a). Typically a knockout cassette, which contains a sequence identical to the 5′ region of the target, a selection marker cassette gene flanked by two FRT/loxP sites and a sequence identical to a region 3′ of the target, is transformed into the host and integration takes place by the cellular homologous recombination machinery. After verification of a positive transformant, the selection marker is excised by a site-specific recombinase (Sauer 1994).

3.1.1.2 Transposons

Transposons are mobile genetic elements, which change their position in the host genome by a self-mediated mechanism called transposition. Many naturally occurring DNA transposons consist of a transposase gene and its promoter flanked by terminal inverted repeats (TIRs) (Ivics and Izsvák 2010). The TIRs of DNA transposons function as DNA binding sites for the transposase, which on the one hand excises the transposon from one genomic locus and on the other hand cuts at a genomic integration site and mediates the reintegration of the transposon (Ivics and Izsvák 2010). In contrast to DNA transposons, which replicate mainly by this cut and paste strategy, RNA transposons are transcribed into RNA, reverse transcribed into DNA and then inserted into a new position applying a copy and paste mechanism (Wicker et al. 2007). The integration sites of transposons are either close to random (2 bp recognition) or occur at specific recognition sites or genomic hot spots (Ivics and Izsvák 2010; Grabundzija et al. 2010). Therefore transposons have been widely used for random mutagenesis in prokaryotic and eukaryotic host systems [e.g. (Le Breton et al. 2006; Dumais Pope et al. 1994; Fernandes et al. 2001; Weil and Kunze 2000; Parinov et al. 2004)]. For genome engineering purposes, a DNA sequence of interest is cloned between the TIRs on a plasmid vector from where it gets excised and further integrated in the chromosome by the transposase, expressed from a separate expression plasmid (Fig. 3.1b) (Ivics and Izsvák 2010). A collection of transposable elements with different integration preferences, cargo capacities and species specificities are available (Skipper et al. 2013). Widely used transposons include the commercially available, mariner-type transposon Sleeping Beauty (Ivics et al. 1997), the Ac/Ds system described by the Nobel Prize winner Barbara McClintock (McClintock 1950), the synthetically reconstructed Frog Prince (Miskey 2003) and the artificial piggyBAC transposon (Cary et al. 1989; Thibault et al. 2004). The integration site of these transposons ranges from AT dinucleotides to palindromic consensus sequences similar to restriction enzymes (Ivics and Izsvák 2010). Transposases with higher site specificity have been created by fusions to zinc-finger (ZF) (Feng et al. 2010; Yant et al. 2007) and TALE DNA binding domains (Owens et al. 2013). Although transposase fusion proteins have a high affinity for their intended target region, they also show off-targeting activity with a prevalent danger of dysregulating endogenous genes (Feng et al. 2010).

Transposition was one of the first genome editing tools used for insertional somatic and germ line transgenesis in mammalian cell lines (Rushforth et al. 1993; Bessereau et al. 2001; Cooley et al. 1988). Moreover transposons are often used as a delivery system for DNA as a less oncogenic alternative to viral vectors. In contrast to their viral counterparts, transposons can be maintained as plasmid DNA. Also, packing capacities are higher and the transfection protocols are less labour intensive and time-consuming (Ivics and Izsvák 2010). One of the major advantages of the transposition is its independency from cellular repair pathways, the stage of the cell cycle and the cell type (Gaj et al. 2013).

3.1.2 RNA Interference and Translational Repression

RNA interference (RNAi)-induced targeted gene knockdown by small interfering RNAs (siRNA) or microRNAs (miRNAs) is a rapid and inexpensive technique primarily applied in higher eukaryotes (Fig. 3.1c). In contrast to the previously described genome engineering methods, regulation takes place at the posttranscriptional level. It offers an alternative to genome engineering methods by taking advantage of short non-coding RNAs (miRNAs and siRNAs), which guide an RNA-induced silencing complex (RISC) to bind complementary sequences in the messenger RNA (mRNA). miRNAs and siRNAs derive from structurally different precursor molecules, but they are processed in a similar manner. Binding of the miRNA–RISC complex to the target mRNA decreases protein output; however, the molecular details and timing of how mRNA degradation and translational repression each contribute to this effect are still a matter of scientific debate. Several studies show that miRNAs can function as siRNAs and vice versa and that the mechanism of choice is highly dependent on the degree of complementarity of the RNA target (Doench et al. 2003; Zeng et al. 2003; Hutvágner and Zamore 2002; Karbiener et al. 2014).

The RNAi mechanism is naturally protecting the genome against mobile genetic elements such as transposons or viruses (Ratcliff et al. 1999) and is used as an important mechanism for regulating gene expression (He and Hannon 2004). In metazoan RNAi the primary siRNA transcripts are cleaved by a ribonuclease III termed Drosha (in animals, DcI1 in plants) into 70–80 bp precursor miRNAs (pre-miRNAs) (Lee et al. 2003). The double-stranded pre-miRNAs are then exported to the cytoplasm and processed by the multidomain ribonuclease Dicer into double-stranded siRNA. The 21–23 bp siRNAs have a phosphorylated 5′ terminus and a 2 nt 3′ overhang, which is required for the recognition of the siRNA by RISC (Bernstein et al. 2001; Knight and Bass 2001). Dicer passes the ds siRNA to the RISC, where the DNA duplex is unwound, and the passenger strand, which is the strand whose 5′ terminus has the thermodynamically more stable end of the duplex, gets degraded by the nuclease called Ago2 (in human, fly) (Matranga et al. 2005; Tomari et al. 2004). The single-stranded guide RNA (guide

strand) is incorporated in the RISC and guides the complex for sequence-specific mRNA cleavage, which occurs in the region complementary to the siRNA (Elbashir 2001).

In contrast to siRNAs, miRNAs are naturally synthesized from a 60–70 nt transcript, which folds into a stem loop precursor. These precursor molecules are processed by the nucleic Drosha and the cytosolic Dicer, similarly to siRNAs, into ds miRNAs of ~22 bp length (Bartel 2004; Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). The mature miRNA is incorporated into a ribonucleoprotein (RNP) complex comparable to the RISC (Mourelatos et al. 2002). miRNAs bind with mismatches and bulges to the target and cause direct translational repression or mRNA destabilization (Lim et al. 2005). A short miRNA region designated as 'seed', which comprises the positions 2–7 or 2–8 of the mature miRNA, shows perfect base pairing to the mRNA's 'seed match' sequence (Lewis et al. 2003). In general miRNAs and siRNAs use similar mechanisms for the repression of mRNA expression and the cleavage of mRNA. Although the binding sites for miRNAs in animals lie almost only in the 3′ UTR of the target gene, any position on the mRNA is mechanistically sufficient for miRNA binding and translation repression (Lytle et al. 2007).

In addition to the naturally occurring mi-/siRNAs, synthetic single-stranded (ss) and double-stranded (ds) RNAs, 21–22 nucleotide siRNAs and short hairpin RNAs (shRNAs) have been successfully used to induce RNA interference and specific gene silencing in numerous organisms (Bernstein et al. 2001; Misquitta and Paterson 1999; Elbashir et al. 2001; Paddison et al. 2002; Brummelkamp et al. 2002; Rivas et al. 2005). These molecules are either endogenously expressed in vivo or synthesized in vitro and delivered into the host organism (Donzé and Picard 2002). Transfected RNA is prone to degradation, whereas the stable integration of RNA-expressing vectors ensures a persistent gene silencing (Brummelkamp et al. 2002; Nikolova and Toncheva 2008).

Creating a clean, confirmed gene knockout is a tedious procedure in higher eukaryotes. RNAi is a popular method for posttranscriptional loss-of-function studies due to is simplicity, efficiency and reduction of cost compared to genome engineering techniques. siRNAs and miRNAs for RNAi can be quickly prepared and introduced into the host organism or cell, where they are used to perform functional studies and generate double- or triple-loss-of-function effects independently from molecular genetic technologies (Niwa and Slack 2007a). Nevertheless RNAi is accompanied by off-targeting effects, since it can activate dsRNA-responsive cellular pathways resulting in a dysregulation of host cell genes (Sledz et al. 2003). In addition, off-target binding to similar mRNA sequences alters the endogenous protein production levels and thereby limits the applicability of RNAi for basic research and clinical therapy (Sledz and Williams 2005; Jackson et al. 2006). Thus it is recommended to confirm RNAi effects with two or more dsRNAs/siRNAs/shRNAs targeting different regions in the gene of interest (Niwa and Slack 2007b).

3.1.3 Prokaryotic Genome Engineering

Genome engineering tools such as group II introns and recombineering are limited to use in prokaryotes. Although group II introns have been tested in eukaryotes, the efficiency is very low. Recombineering has not been applied in eukaryotes, because of the substantial differences in DNA replication.

3.1.3.1 Group II Introns: Targetrons

Group II introns are naturally found in bacterial genomes as well as in organellar genomes of some eukaryotes (Michel and Feral 1995) and have become commercially available for targeted genome engineering in prokaryotes ('Targetrons'). The most widely used Targetron is the *Lactococcus lactis* Ll.LtrB group II intron, which consists of a multifunctional intron-encoded protein (IEP) and an intron RNA that can be reprogrammed for a desired locus (Mills et al. 1997). IEP has reverse transcriptase, RNA splicing and DNA endonuclease activities.

Group II introns are mobile genetic elements that insert themselves into specific genomic sites by a catalytic mechanism called retrohoming, where the target sequence is recognized by specific base pairing with the intron RNA (Fig. 3.1c) (Lambowitz et al. 1999). The IEP and the intron form an RNA–protein complex (RNP), which scans DNA and specifically recognizes the target site by interactions of the IEP and the intron. The IEP recognizes a small number of fixed nucleotides (<5) similar to the protospacer adjacent motif (PAM) of CRISPR–Cas9 (see below) and triggers the unwinding of the DNA, which enables the intron RNA to specifically base pair to the 14–16 bp target sequence (Singh and Lambowitz 2001). These group II intron recognition sequences are large enough to be unique even in a complex genome (Guo 2000). Thereafter, the intron RNA inserts into one strand of DNA by a reverse splicing mechanism and the IEP cleaves the opposite strand and uses the cleaved 3′ end as a primer for reverse transcription of the intron RNA. Subsequently, the host cell repair enzymes integrate the intron cDNA into the genomic target (Eskes et al. 1997; Cousineau et al. 1998; Smith et al. 2005). Group II introns can be used for two types of DNA modifying reactions in the host genome: on the one hand the site-specific integration of the intron RNA and on the other hand the introduction of double strand breaks by reverse transcriptase deficient IEPs, which are repaired by the NHEJ/HR machinery of the host cell (Karberg et al. 2001). The integration reaction can also be applied to deliver additional genes within the intron as cargo into the target locus (Frazier et al. 2003; Rawsthorne et al. 2006).

Insertion efficiencies with antibiotic selection reached almost 100 %, whereas without selection disruption frequencies of 1–80 % have been obtained (Perutka et al. 2004). The main advantages of group II introns are the ease of retargeting and the highly site-specific integration. Although they have been primarily used for the modification of prokaryotic organisms [e.g. (Karberg et al. 2001; Frazier et al. 2003)], a few applications in eukaryotes were reported (Mastroianni et al. 2008). Drawbacks for the use in higher organisms are the low efficiency, the requirement of high Mg^{2+} concentrations for IEP to be active and the chromatin

structure of eukaryotic target DNA, which may impede access of the RNP complex (Lambowitz and Zimmerly 2011). However, injection of additional Mg^{2+} into the cells improved the site-specific integration and double strand break-induced recombination (Mastroianni et al. 2008). With an increased efficiency, group II intron-based systems may also become a valuable tool for eukaryotic genome engineering. They combine the low off-targeting rates similar to TALENs and zinc-finger nucleases and the flexible reprogramming comparable to CRISPR–Cas9.

3.1.3.2 Recombineering

Recombineering is used for engineering of bacterial chromosomes by bacteriophage proteins, which promote gene replacement of a linear single- or double-stranded DNA substrate upon introduction in the host cells (Sharan et al. 2009). Lambda Red is the most commonly used recombineering system amongst several types reported (Murphy 1998; Muyrers et al. 1999). Lambda Red recombineering involves three enzymes, Gam, Exo and Beta, of which Beta is the key factor in the recombination process. Beta is a ssDNA binding protein that anneals the single-stranded donor molecule to the complementary lagging strand near the replication fork, which leads to a permanent integration after one round of replication. Exo, a $5' \rightarrow 3'$ exonuclease, is only required for dsDNA integration for the generation of a single-stranded intermediate. Gam inhibits the endogenous RecBCD nuclease, which is responsible for the degradation of dsDNA, and thereby additionally increases the efficiency of dsDNA integration (Fig. 3.1e) (Ellis et al. 2001; Mosberg et al. 2010). Lambda Red integration is independent of endogenous host cell recombination enzymes and requires ss or ds donor substrate sequences with 35–50 bp identities to the target sequence (Muyrers et al. 1999). Lambda Red recombineering can be applied for gene replacements, deletions, insertions, inversions and point mutations in prokaryotic genomes. The development of recombineering systems is reminiscent of the efficient homologous recombination in *S. cerevisiae* upon transformation with linear DNA fragments (Eckert-Boulet et al. 2011). In prokaryotes, lambda Red-mediated integration of oligonucleotides takes place with a frequency of $\sim 5 \times 10^{-4}$ recombinants per viable cell, depending on the oligo length and concentration (Yu et al. 2000). This efficiency is too low for markerless gene insertion and the simultaneous modification of multiple genes (Mosberg et al. 2012). Recombination frequencies were increased by mutating the host recombination machinery, deleting endogenous nucleases, varying promoters and copy numbers of the lambda Red genes or engineering the lambda Red proteins (Murphy 1998; Mosberg et al. 2012; Muyrers et al. 2000; Nakayama and Ohara 2005).

Multiplex automated genome engineering (MAGE) by lambda Red-mediated gene replacement has been used for large-scale reprogramming and evolution of cells. A pool of oligos with degenerate sequences is repeatedly introduced into cells, which harbour the lambda Red recombineering mechanism. These oligos are incorporated depending on the similarity to genomic sequences in more than 30 % of the *E. coli* cell population every 2–2.5 h. This technique enables an automated, simultaneous modification of different genomic loci in a fast and reliable manner

(Wang et al. 2009). MAGE was used, for instance, to insert hexa-histidine sequences into genes coding for the entire translation machinery in *E. coli* in vivo. His-tag purification and reconstitution of the translation machinery allowed fully recovered in vitro protein synthesis (Wang et al. 2012a). Moreover it was applied to insert multiple T7 promoters into several genomic operons to optimize biosynthesis of aromatic amino acid derivatives (Wang et al. 2012b) and for genome-wide codon replacement (Isaacs et al. 2011).

Strains that were modified by MAGE (or that harbour other interesting mutations) can easily be combined by conjugative assembly genome engineering (CAGE). This technique is used to facilitate the large-scale assembly of genomes by merging genes of a donor and a recipient strain by bacterial conjugation (Isaacs et al. 2011).

3.1.4 Universally Applicable Genome Editing Strategies

3.1.4.1 Homing Endonucleases

Homing endonucleases are rare cutting endonucleases, which cleave dsDNA larger than 12 bp and thereby trigger the HR and NHEJ repair mechanisms of the cell (Thierry and Dujon 1992; Belfort and Roberts 1997). They are encoded by introns, self-splicing inteins and freestanding open reading frames, which is described by the prefixes 'I', for intron; 'PI', for protein insert; and 'F', for freestanding in the nomenclature of homing nucleases (Belfort and Roberts 1997; Dujon et al. 1989; Perler et al. 1994). Intron-encoded homing nucleases are considered as selfish genetic elements, because they recognize DNA sequences, which are identical to the sequences up- and downstream of the nuclease-encoding intron and thereby promote their own duplication and perpetuation of genetic information without providing advantages for the host organism (Edgell 2009). The homing process itself is initiated by a double strand break performed by the homing endonuclease (Fig. 3.1f) (Chevalier and Stoddard 2001). The homing sites, which are specifically recognized by the nucleases, are 12–40 bp long (Belfort and Roberts 1997; Lambowitz and Belfort 1993) and appear, assuming a random sequence organization, once per 7×10^{10} bp for an 18 bp sequence. Accordingly, an 18 bp homing site occurs statistically once in a mammalian-sized genome (Jasin 1996). In order to survive, homing nucleases are more tolerant to mutations in the recognition sequence and off-targeting. In contrast to restriction enzymes, minor aberrations to the recognition sequence lead to a decrease in the activity and do not completely abolish DNA cutting (Gimble and Wang 1996; Argast et al. 1998). Genetic and protein engineering techniques increased the quantity of available homing nucleases by fusing preexisting nuclease domains to create protein chimaeras and by directed evolution to specifically alter the DNA binding domain (Arnould et al. 2006; Seligman et al. 2002; Doyon et al. 2006; Grizot et al. 2011). The prefix H, for hybrid, or E, for engineered, was introduced to describe these synthetically modified enzymes (Roberts 2003; Chevalier et al. 2002).

3.1.4.2 Zinc-Finger Nucleases

Zinc-finger nucleases (ZFN) are synthetically engineered enzymes, which contain a) zinc-finger DNA binding domain fused to the non-specific cleavage domain of the type IIS restriction enzyme *Fok*I (Urnov et al. 2010). The zinc-finger binding domain was originally found in the *Xenopus laevis* transcription factor IIIA (TFIIIA), which contains tandem-arranged sequences of 30 amino acids (aa) including two pairs of cysteine and histidine residues. Each 30 aa unit ('module') binds one zinc atom using the cysteine and histidine residues and thereby forms a structure reminiscent of a finger, which specifically contacts three DNA bases (Miller et al. 1985; Pavletich and Pabo 1991). Many different zinc-finger motifs have been identified with binding affinities towards DNA, RNA, proteins and membrane-associated ZF domains (Laity et al. 2001). The *Fok*I nuclease consists of a DNA binding and a cleavage domain. The cleavage domain cuts DNA non-specifically if separated from the DNA binding domain, which is responsible for the specific DNA recognition (Li et al. 1992). The *Fok*I nuclease is only active as a dimer. Therefore cleavage by ZFNs is only achieved if two ZFNs are designed to bind adjacent DNA sequences with a spacer region of ~6 bp in between (Fig. 3.1f) (Bitinaite et al. 1998; Bibikova et al. 2001). Several studies show that the distance between the two ZF binding sites (~6 bp) and the size of the linker, which connects the nuclease and the DNA binding domain, are interdependent on each other (Händel et al. 2009; Shimizu et al. 2009).

The zinc-finger DNA binding domain can be individually designed to bind a sequence of interest by combining single zinc-finger modules or by altering the binding specificity of a given module by exchanging single aa residues (Pabo et al. 2001; Liu et al. 1997; Isalan and Klug 1998, Choo Y 1998). Typically three to six single zinc fingers of known specificity are linked together to specifically recognize a contiguous DNA sequence. Two three-finger arrays recognize an 18 bp DNA sequence, which is in theory long enough to address a unique locus in a mammalian genome (Carroll 2011). Until now, for nearly all of the 64 possible nucleotide triplets, specific ZF domains have been developed, which can be linked together to target a desired genomic sequence (Liu et al. 1997; Beerli and Barbas 2002; Bhakta et al. 2013; Gonzalez et al. 2010). Not all artificially designed fingers in a ZFN show equal binding affinities towards the target sequence. Thus, some ZFNs have affinities for related sequences other than the supposed target and others completely fail to bind the target sequence. To overcome these obstacles, multiple pairs of ZFNs are tested for a single target gene, and the number of fingers is often increased to improve the specificity of ZFN binding (Carroll 2011). In order to reduce the levels of unwanted, mutagenic NHEJ, zinc-finger nickases (ZFNickases) were created by the inactivation of the catalytic domain of one ZFN monomer. After binding of the ZFN to the target site, *Fok*I dimerizes and introduces a nick in the dsDNA (Ramirez et al. 2012). Further improvements include engineered ZFN variants, which only cleave DNA, when paired as a heterodimer, thereby reducing off-targeting by more than 40 % (Miller et al. 2007). Also hyperactivated nuclease variants were generated (Gordley et al. 2009). A highly efficient *Fok*I nuclease derivate 'Sharkey', which had been created by multiple rounds of cycling

mutagenesis and DNA shuffling, showed a more than 15-fold increase in activity compared to the conventional *Fok*I cleavage domain and a three- to sixfold improvement in targeted ZFN mutagenesis (Guo et al. 2010).

The first genome modification using ZFNs was described more than 15 years ago, and since then it has remained the most effective and versatile genome editing technique for many years (Carroll 2011). ZFN technology for the first time allowed programmable genome engineering by combining small, single protein modules. Drawbacks of ZFN include the laborious and time-consuming design of the DNA binding domain and the inaccessibility of some genomic target sites for ZFN cleavage due to the chromatin status. ZFN de novo design requires screening of various ZF-libraries, which are often not publicly available (Sun et al. 2012). Nowadays TALENs have been widely applied as an alternative to ZFNs, providing some advantage: compared with ZFNs, TALENs can target a larger range of sequences, are easier to design and display higher rates of cleavage (Sun et al. 2012; Yan et al. 2013).

3.1.4.3 TALE Nuclease Technologies

Transcription activator-like effector (TALE)-mediated genome engineering takes advantage of the DNA binding domain of bacterial effector proteins, which influence the expression of plant genes to aid bacterial infection after injection into host cells. For genome engineering purposes, the DNA binding domain of TALEs can be redesigned to bind a desired genomic sequence. The most common TALE applications are TALE nucleases (TALENs), which provide a powerful tool to introduce mutations in the genome of the host organism, and TAL–transcription factors (TAL–TFs), which are used to enhance or silence the expression of a gene of interest (Gaj et al. 2013; Mussolino and Cathomen 2012).

Plant pathogen-mediated promoter activation by bacterial TALE proteins was first described in 2007 (Kay et al. 2007; Römer et al. 2007). Until now more than 100 members of the TALE family have been described, all consisting of similar structural elements (Boch and Bonas 2010; Boch 2011): C-terminal nuclear localization signals (NLSs), an N-terminal type III secretion and membrane translocation signal, a transcriptional activation domain (AD) and a DNA binding domain of 1.5–33.5 (in most cases 17.5) highly similar repeated domains (Fig. 3.1f) (Römer et al. 2007; Van den Ackerveken et al. 1996; Bogdanove et al. 2010). Each repeat consists of 30–42 nearly identical amino acids, where only the 12th and the 13th amino acids are highly variable. These two amino acids, called 'repeat variable diresidues (RVDs)', recognize a single DNA base in the target DNA. Some highly specific RVDs recognize unique DNA bases, whereas others differentiate purine or pyrimidine bases. There exist also non-specific repeats, which bind any DNA base (Boch et al. 2009; Moscou and Bogdanove 2009). RVDs are further classified into weak, intermediate and strong according to their efficiency to bind a DNA base (Streubel et al. 2012). TALE–DNA recognition and activity are influenced by the number and position of DNA repeat mismatches, the number of repeats present and the specificity of the single repeats (Scholze and Boch 2010). Certain genomic

regions are most probably inaccessible for TALE binding because of the condensed chromatin status (Zhang et al. 2014).

TAL effectors for specific gene activation were used with native or heterologous transcriptional activation domains (Morbitzer et al. 2010). TALE fusions with VP16 and the VP64 transcription activators of the herpes simplex virus successfully activated transcription in human cells (Zhang et al. 2011; Tremblay et al. 2012). TAL–TF-mediated gene activation can lead to a more than 20-fold increase of the transcription level (Morbitzer et al. 2010).

Similar to ZFNs, TAL effectors have been fused to nucleases to introduce targeted double strand breaks for genome engineering. *Fok*I is the most prominent nuclease used for TALEN research (Miller et al. 2011; Christian et al. 2010; Li et al. 2011). Since *Fok*I is only functional as a dimer, two specific TAL effectors are required to bind opposing sites of the target DNA in a certain distance, which allows *Fok*I to dimerize and to introduce the strand break. The optimal length of this spacer region is between 10 and 30 bp and depends on the TALEN architecture and the number of repeats (Miller et al. 2011; Li et al. 2011; Cermak et al. 2011; Mussolino et al. 2011). The requirement for two TALE domains, which bind in close proximity and enable the dimerization of *Fok*I, causes minimal off-target effects and thereby reduces the risk of cytotoxicity (Mussolino et al. 2014).

One of the most laborious and time-consuming aspects is the assembly of TALE repeats, because of their highly repetitive nature. Several hierarchical ligation strategies have been described and optimized by different research groups in terms of time efficiency, specificity and overall convenience (Weber et al. 2011a; Sanjana et al. 2012; Uhde-Stone et al. 2013). A ligation-independent method for high-throughput TALE assembly takes advantage of 10–30 bp long, non-palindromic, single-stranded overhangs, which anneal in a highly specific manner (Schmid-Burgk et al. 2013). Automated assembly makes TALENs fast and easily available, but is accompanied by relatively high costs (Reyon et al. 2012). Independent of the assembly method for reprogramming TALENs, retargeting CRISPR–Cas9 can be achieved in a much faster and cheaper way. The large size of TALENs as well as ZFNs may limit their delivery by size-restricted vectors (Gaj et al. 2013). Forthcoming research may also include TALE fusion proteins with several domains such as methyl- and acetyltransferases, deacetylases or deaminases to study transcriptional regulation (Bogdanove and Voytas 2011).

3.1.4.4 CRISPR–Cas9 Technology

Since the elucidation of the biological function of clustered regularly interspaced short palindromic repeats (CRISPR) in 2007 (Barrangou et al. 2007; Marraffini and Sontheimer 2008) and mechanistic understanding of the RNA-guided DNA nuclease Cas9 in 2012 (Jinek et al. 2012; Gasiunas et al. 2012), the number of publications has exploded suggesting a new standard strategy for genome editing. CRISPR–Cas9 uses a short, non-coding, 'guide' RNA (gRNA) to direct a nuclease to specifically cut a DNA sequence of interest (Fig. 3.1f). In contrast to genome editing techniques described above, the genomic target locus can be changed by simply varying a 20 bp sequence of the gRNA, instead of cumbersomely

reprogramming DNA binding domains, giving researchers high flexibility and an accelerated throughput (Gasiunas and Siksnys 2013).

Natural Function in Adaptive Immunity in Bacteria

This method originates from bacteria and archaea, where it provides an adaptive immunity against invading nucleic acids such as phage or plasmid DNA (Mojica et al. 2000; Jansen et al. 2002). The term CRISPR describes a genomic locus consisting of multiple, short, palindromic sequences (typically 24–37 bp), which are interspaced by short sequences of extrachromosomal origin (phage and plasmid DNA) called spacers (27–72 bp) (Grissa et al. 2007; Mojica et al. 2005). Usually more than one CRISPR locus is found in a bacterial or archaeal genome (Barrangou and Horvath 2009). These CRISPR loci are flanked by a 300–500 bp leader sequence on one side, and a set of CRISPR-associated (Cas) genes are located adjacent to the CRISPR array. The function of the leader sequence is suggested on the one hand to promote the transcription of the CRISPR array and on the other hand to play an important role in the recognition and acquisition of incoming spacers (Brouns et al. 2008). Until now there were three CRISPR–Cas systems identified (types I, II, III), which are further divided into various subtypes (Westra et al. 2012; Makarova et al. 2011).

CRISPR–Cas9-Based Targeted Genome Editing

Engineered CRISPR–Cas9 nowadays serves as one of the most promising, advanced methods in targeted genome editing. In contrast to other CRISPR types, which require multiple proteins, the endonuclease Cas9 from the type II A CRISPR system is solely responsible for DNA recognition, binding and cleavage. Cas9, originally isolated from *Streptococcus pyogenes*, uses two separate nuclease domains each of which cuts one DNA strand. Deletion of the catalytic activity of one nuclease domain leads to nicking of the DNA (Cas9 nickase). For genome engineering purposes, the Cas9 endonuclease and a short artificially designed guide RNA (gRNA) are co-expressed in the host cells (Fig. 3.1f). The gRNAs consist of an ~80 bp structural part and a ~20 bp variable part. The structural part of the gRNA folds into a stem loop, whereas the variable part binds to the complementary target DNA. Upon binding of Cas9 to the DNA, the strands get unzipped at a sequence called protospacer adjacent motif (PAM, located on the DNA), and base pairing of the gRNA to the complementary DNA strand takes place. Sequence-specific cleavage of Cas9 is dependent on the PAM as well as on the complementarity between the gRNA and the DNA target (Jinek et al. 2012). The CRISPR–Cas9 target regions and the design of gRNAs are restricted to the PAM sequence. *S. pyogenes* Cas9 tolerates the nucleotides NGG and NAG as PAM (Hsu et al. 2013), but PAM sequences differ amongst the various CRISPR types and Cas proteins (Haft et al. 2005; Shah et al. 2013).

Multiplexed genome engineering can easily be achieved by expressing several gRNAs on a single vector (Cong et al. 2013; Sakuma et al. 2014). CRISPR–Cas9-mediated transcription regulation (CRISPRi) was achieved by the use of a

catalytically inactive Cas9 mutant, which is fused to transcriptional activation and repression domains in eukaryotes (Gilbert et al. 2013), a strategy similar to TAL–TFs.

A main drawback of the CRISPR–Cas9 system is the high off-targeting rate. Cas9 off-target sites can harbour up to five mismatches and the cleavage rates are similar to the on-target region (Fu et al. 2013; Cradick et al. 2013). Single-base mismatches up to 11 bp upstream of the PAM completely abolish DNA cleavage, whereas mismatches further upstream retain Cas9 activity (Cong et al. 2013). Undesired off-targeting can be reduced by the use of the Cas9 nickase and truncated gRNAs (Ran et al. 2014; Fu et al. 2014). Dimeric Cas9–*Fok*I fusion proteins reduce the likelihood that a suitable target site will occur more than once in a mammalian-sized genome. These dimeric RNA-guided *Fok*I nucleases (RFN) combine the ease of reprogramming of the CRISPR–Cas9 system and precise targeting similar to TALEN or ZFN (Tsai et al. 2014).

Genome engineering techniques for synthetic biology are used for a broad range of applications starting from basic research use in biotechnology up to medical applications. Their huge potential is confirmed by a rapidly growing list of organisms and targeted genes. However, continuous improvements and diversification of currently available techniques are required to overcome their limitations and bottlenecks. Future prospects might include further improvements of existing technologies by increasing their specificities, reducing toxicity effects as well as finding novel technologies and combinatorial methods. The collection of methods for genome engineering mentioned here may serve as a guideline to identify the most appropriate method to address a certain application of interest.

3.2 DNA Assembly Methods

In recent years two major enabling technologies contributed to the success of synthetic biology. On the one hand, rapid developments in DNA synthesis provided new opportunities for synthetic biologists (Carr and Church 2009). Recently, synthetic genes have gained increasing importance for routine cloning due to affordable prizes and because they allow to engineer host cells for various applications (Ellis et al. 2011). On the other hand, novel DNA assembly methods allowed the combinatorial construction of pathways and cellular networks up to the assembly of an entire synthetic genome (Gibson et al. 2010). A wide variety of quick and reliable DNA assembly methods is available, yet the assembly method may vastly influence the final construct in both architecture and diversity (Ellis et al. 2011). Here we provide an overview of currently established DNA assembly techniques that can be divided into two groups: methods based on restriction endonucleases and restriction-/ligation-independent assembly strategies. The latter techniques can either be used in vitro (e.g. PCR-based methods and ligation-dependent and ligation-independent methods) or in vivo, carried out in either yeast or *Bacillus* species (Fig. 3.2).

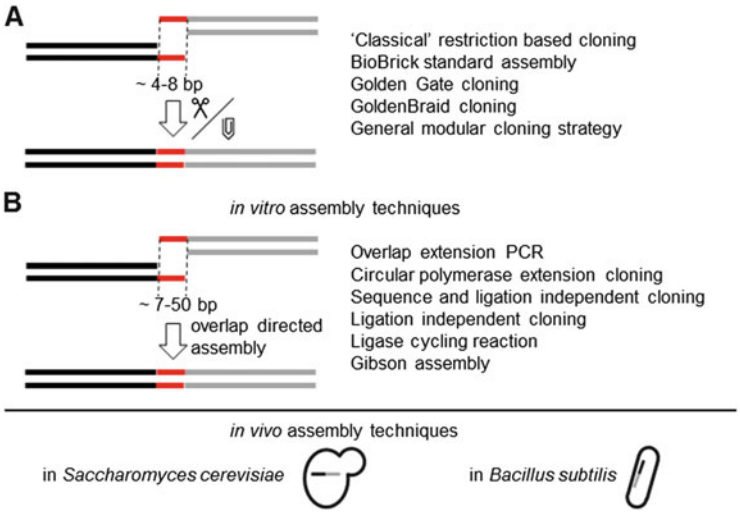


Fig. 3.2 Overview of DNA assembly methods. (**a**) Restriction endonuclease (RE)-based assembly methods relying on short overhangs by RE digestion followed by ligation are shown. (**b**) Overview of the overlap-based cloning methods relying on longer sequences (<50 bp) highlighting in vitro and in vivo methods which rely on the endogenous recombination machinery of either yeast or bacteria

3.2.1 Restriction Endonuclease-Based Cloning Methods

Using restriction endonucleases (REs) is a popular routine method for DNA cloning. Double-stranded DNA is cleaved at specific sites by REs producing either sticky (overhangs) or blunt ends. After digestion, fragments can be combined in the desired order and/or inserted into a plasmid backbone using a DNA ligase (Ellis et al. 2011). DNA fragments can be shuffled between vectors given that suitable restriction sites are present. Moreover, restriction cloning is applied for DNA analysis and DNA forensics (Ellis et al. 2011). Assembly techniques of this group include the use of 'classical' restriction cloning based on type II REs (Pingoud et al. 2005), which recognize palindromic sequences (Fig. 3.3a). Standardized methods for DNA assembly such as BioBrick standard assembly (Knight 2008) apply a set of type II restriction endonucleases to enable greater modularity and the assembly of multiple parts in a defined order. Seamless cloning (Lu 2005), however, is not possible using type II enzymes, as the recognition site remains as a 'scar' between the joined fragments. Type IIS REs cut outside of a non-palindromic recognition site in a variable sequence that can be defined by experimental design. Thereby customized overlapping regions can be designed that allow seamless fusions of multiple parts in a predefined order (Fig. 3.3b). Methods employing type IIS REs include Golden Gate cloning (Engler et al. 2009) and its descendent methods GoldenBraid (Sarrion-Perdigones et al. 2011) and General Modular Cloning Strategy (MoClo) (Weber et al. 2011b).

Fig. 3.3 Cleavage patterns of type II (**a**) and type IIS (**b**) restriction endonucleases. *Eco*RI and *Bsa*I are shown as examples for each group. The recognition site is highlighted by a *rectangle*, the cleavage pattern and overhangs created are indicated by *dashed lines*. Note that *Bsa*I has a non-palindromic recognition sequences and cuts outside in a variable sequence

3.2.1.1 Standardized Assembly Methods Using Type II REs

The BioBrick (BB) standard assembly relies on restriction cloning that meets the specifications of the BioBrick standard established by Knight et al. (Knight 2008). BB standard biological parts are defined units of DNA, such as promoter or terminator sequences, coding sequences and ribosome binding sites. Parts designed according to the BioBrick standard can be combined with any other BB in a standardized assembly process (Knight 2008). The BB standard assembly employs two pairs of REs that produce compatible ends. A standard BB part is flanked by the recognition sites of these enzymes in a defined manner (Knight 2008): *Eco*RI (outer cutter) and *Xba*I (inner cutter) cut upstream and *Spe*I (inner cutter) and *Pst*I (outer cutter) downstream of the part. In order to join standard parts to already existing BioBrick vectors, each part is digested at one inner and an outer site. For instance, the insert part is digested with *Xba*I and *Pst*I, while the vector containing already another part, e.g. a coding gene, is restricted using *Spe*I and *Pst*I. The *Spe*I and *Xba*I overhangs are compatible and readily hybridize, thus joining the vector part and the insert part, while the *Pst*I overhangs on the other end join the insert part fragment with the vector backbone. The ligation of the compatible overhangs leaves a scar in that neither *Spe*I nor *Xba*I will cleave. This scar is typical for the BioBrick standard assembly and results from every assembly reaction. The standard assembly produces a new BB, though: On the assembly product, the arrangement of the *Eco*RI, *Spe*I, *Xba*I and *Pst*I restriction sites is the same as on the BB standard parts for its construction. In theory, an indefinite number of BB standard parts can be concatemerized using this approach (Knight 2008). The assembly of BioBrick standard parts is easy and straight forward and can be performed in any lab equipped for restriction cloning. BB standard assembly is a low-cost method with a pool of genetic parts available free of intellectual property rights from the Registry of Standard Biological Parts (http://parts.igem.org). This institution collects BB parts and distributes them to registered members and iGEM teams on demand under http://parts.igem.org/Help:Requesting_Parts. However, the application of BB assembly is hampered by the requirement for the standardized restriction sites to be unique. Depending on the reading frame, the *Spe*I/*Xba*I scar sequence may lead to a frameshift or act as a translation termination signal. Phillips and Silver addressed these shortcomings and suggested appropriate improvements in parts design by reducing the scar region to only 6 nt (Phillips and Silver 2006). Also Knight and colleagues improved their own method in the BioBrick-2 standard assembly (Knight 2008) by reducing the scar region to 6 nt, to enable in frame

cloning comparable to Phillip and Silver. Changes to assembly protocols and the location and design of the restriction sites decreased the compatibility to the original BB standard biological parts from the parts registry. The BglBrick (BBb) assembly (Anderson et al. 2010) is an alternative to BioBrick standard assembly relying on the compatible RE pairs *Bam*HI + *Bgl*II and *Eco*RI + *Xho*I. These combinations leave also only a 6 bp scar. Parts designed according to the BBb standard are not compatible with other standard assembly methods and therefore make up an additional assembly standard of their own. The strict dependence on unique restriction sites again limits its applicability for some cloning purposes (Lee et al. 2011).

3.2.1.2 Type IIS RE-Based Cloning Methods

Golden Gate Cloning

Golden Gate cloning (GG) applies only a single type IIS RE, *Bsa*I, for the enzymatic digest of DNA fragments to generate custom ssDNA overhangs at the 5′ and 3′ termini (Engler et al. 2008). In contrast to type II REs that cut within their palindromic recognition site, type IIS REs use separate recognition and cleavage sites. The recognition sequence is non-palindromic and determines the orientation of the cleavage reaction. For instance, *Bsa*I binds to a 6 bp recognition sequence and cleaves downstream of it creating 4 bp overhangs (Zhenyu and Huang-Yong 2002) (Fig. 3.3b). The sequence of the cleavage site is random and can be designed such that compatible ends for ligation are generated after restriction (Engler et al. 2008). Simply put type IIS REs can be conceived as sticky-end cutters with free-to-choose overhangs. In this way, the directional cloning of several fragments is feasible, and all fragments can be digested with the same RE (Engler et al. 2008). The GG assembly starts with two DNA molecules, the vector and an insert. On the vector, *Bsa*I sites flank the desired integration site for the insert. Their cleavage direction points into opposite directions and away from each other such that the intervening sequence carrying the *Bsa*I recognition sites is excised. The digestion produces a vector with terminal overhangs complementary to those of the insert but lacking the *Bsa*I recognition sites. The insert is flanked by *Bsa*I sites as well; however, their orientation is pointed towards each other. Restriction liberates the insert with terminal overhangs compatible to the vector. The complementary overlaps anneal and are joined by DNA ligase. The whole process is an isothermal, one-pot reaction, including all DNA fragments, buffers and enzymes (Engler et al. 2008). It is possible to combine the RE and the ligase in one reaction because the ligation product does not contain *Bsa*I recognition sites. GG cloning features a fast reaction time of 30 min and the assembly of up to nine fragments at a time is possible (Engler et al. 2009). However, as with any RE-based assembly, the sequences of interest must not contain *Bsa*I sites. Golden Gate assembly does not allow recycling of already assembled parts. As for other synthetic biology strategies, there is computer-aided design (CAD) software available for GG cloning, for example, ‘j5’ (Hillson et al. 2012), which aids in the design of the fragments to reduce costs and time during the cloning process.

GoldenBraid

GoldenBraid (GB) (Fig. 3.4) (Sarrion-Perdigones et al. 2011) is a multipartite assembly strategy based on Golden Gate cloning. The method employs two type IIS REs with standardized restriction sites in a specific order. As outlined in the previous section, GG cloning products cannot be reused in another GG assembly. They lack the *Bsa*I recognition sites that are removed during the assembly. To relieve this constraint, the recognition site for a second type IIS RE, *Bsm*BI, is added to the destination vector for the first GG assembly. This second RE facilitates the release of the ligation product from the first GG cloning. To ease multipartite cloning GoldenBraid assembly includes a loop design, the 'braid', that limits the necessary compatible overlaps to only three per RE: The product vectors of the first assembly become the destination vectors for the second assembly and vice versa. Compatible overlaps in two vectors per level facilitate the assembly of two inserts in a predefined order. The vectors for the two assembly levels contain both *Bsa*I and *Bsm*BI sites, however, in inverted orientations. Two *Bsm*BI sites pointing inward flank two *Bsa*I sites pointing outward on the first level destination vector, while on the second level vector two inward *Bsa*I sites flank two outward *Bsm*BI sites. In a one-pot reaction including two first and a second level vector and *Bsm*BI as well as a ligase, the inserts are liberated from the first level vectors and combined in the second level vector (Fig. 3.4). The iterative character of the braid design is caused by the alternation between the two REs. Different selection markers on the first and second level vectors allow for counter-selection to improve the efficiency in each step. GoldenBraid can theoretically continue indefinitely; it is limited only by the DNA amount the host can take up and by the stability of the resulting construct. However, the DNA sequences to be ligated must not contain recognition sites of the REs used in the process. Additionally, the availability of destination vectors for different organisms is limited and the method is quite time-consuming due to the need of counter-selection after every assembly step. Recently, Golden Braid 2.0 has been introduced as an assembly method relying on parts that are interchangeable in between work groups, similar to BioBrick standard parts. This new method is available only to plant bioengineers so far (Sarrion-Perdigones et al. 2013).

General Modular Cloning Strategy

General Modular Cloning Strategy (MoClo) is also based on the use of type IIS REs (Weber et al. 2011b) and is similar to GoldenBraid. The restriction enzymes applied in this technique are *Bsa*I and *Bpi*I and are not used in an alternating fashion, but rather limited to different stages of the assembly. *Bsa*I is employed during the first stage of the MoClo assembly where modular DNA parts are assembled into multipart constructs. *Bsm*BI is used at a later stage for the combination of the multipart constructs into more complex assemblies (Weber et al. 2011b). An extension to even more complex designs requires the application of flanking end linkers, which provide *Bsa*I sites for a braid-like assembly mode (Werner et al. 2012). MoClo is a fast and reliable method for the scarless assembly of up to six multipart constructs. Again, the sequences of interest must be optimized to

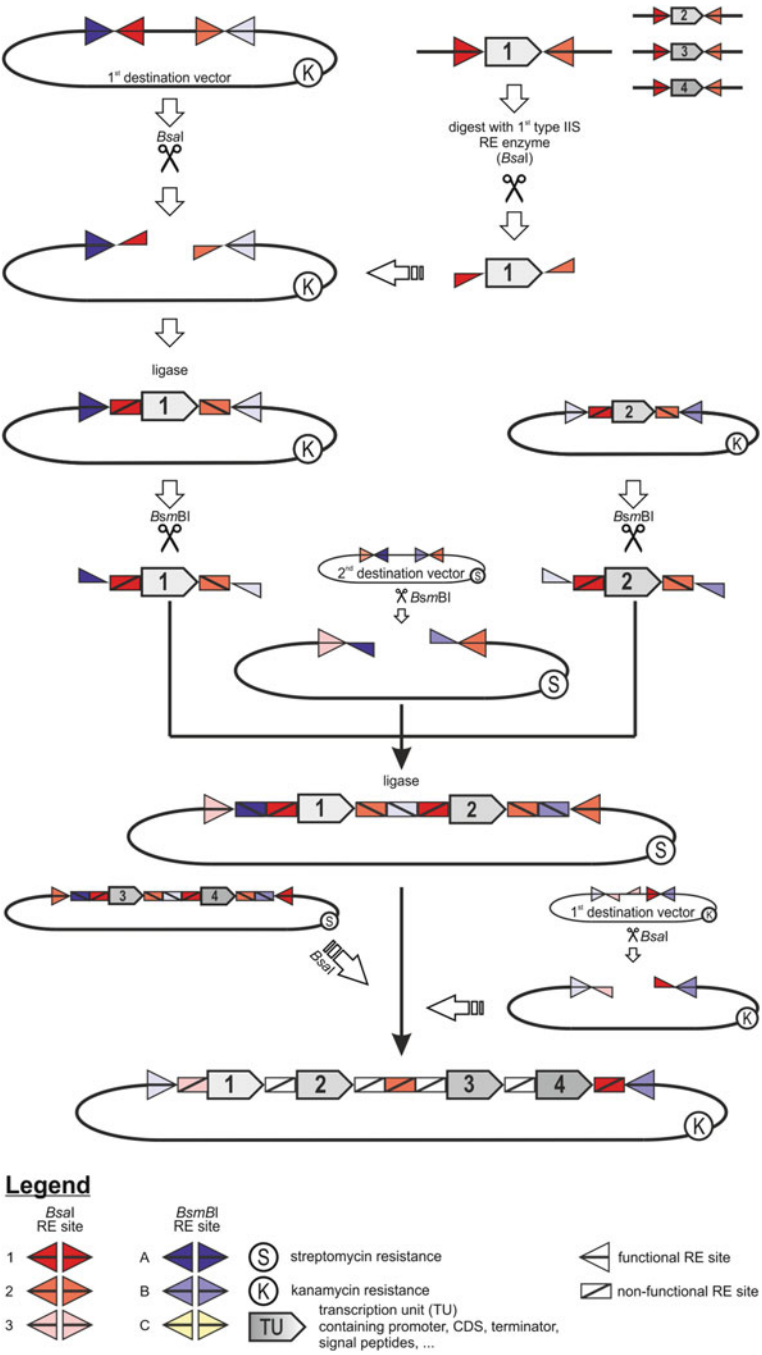


Fig. 3.4 GoldenBraid assembly. The method is based on restriction digest with type IIS restriction enzymes. The cleavage pattern of these enzymes is illustrated in Fig. 3.3b. The restriction enzymes thereby have specific recognition sites and cut the dsDNA at a specified cleavage site. This site is located either downstream or upstream of the recognition site, as indicated by the

avoid restriction by the applied REs. Additionally, different recipient vectors are necessary for the two stages of MoClo as well as its extension to braiding.

Overall, all assembly strategies based on restriction endonucleases share the same advantages: They are usually very reliable in the generation of the final assembly product and easily applicable in standard molecular biology laboratories. However, the assembly of a higher number of DNA fragments can include several steps, which is time-consuming. Additionally, the sequence limitations due to the specific unique restriction sites can limit the application of these methods or further increase the work load required for sequence modifications.

3.2.2 Restriction-/Ligation-Independent Assembly Strategies

Sequence-independent assembly strategies do not rely on restriction/ligation for the fusion of DNA fragments, but rather on the linkage of overlapping homologies. Homologous sequence overlaps are required for all in vitro and in vivo recombination techniques. Nevertheless, the length/design of the homology regions, the means to complete the assembly and the method of stabilization of the assembled products vary. In the following sections several important methods are introduced, providing an overview of the basic underlying principles.

3.2.2.1 Overlap Extension PCR

Overlap extension PCR (OE PCR), also called splicing by overlap extension PCR (SOE) (Horton et al. 1989), is a method initially developed for site-directed mutagenesis (Ho et al. 1989; Higuchi et al. 1988). Terminal overlaps allow DNA fragments to anneal to each other in the desired order. The overhangs are added as primer extensions to the fragments to be joined in a primary PCR reaction. The fragments are then linked and the final product is amplified in a successive PCR reaction. Originally, OE PCR was used for the site-specific mutagenesis of linear DNA fragments (Higuchi et al. 1988) or the construction of hybrid genes (Horton et al. 1989), which were then inserted into a vector backbone. Unless the linear DNA fragment is directly used for transformation of the host, a subcloning step into a plasmid using other methods is necessary. However, even this issue has recently

Fig. 3.4 (continued) orientation of the arrowhead. GoldenBraid method requires that parts, indicated in grey, are initially inserted in specific vector backbones carrying *Bsa*I restriction sites. Depending on the specific overlap generated by the type IIS restriction, the resulting homologous overlaps are marked in shades of *red*. Additionally, the vector containing the *Bsa*I restriction sites carries the one of the two selection markers, indicated by S or K. The plasmids are then introduced into the assembly reaction together with another vector backbone carrying the second selection maker. Upon restriction digest with the second enzyme *Bsm*BI, indicated in shades of *blue*, two inserts are liberated from the primary assembly vector and constituted in a directional assembly into the second assembly backbone. Correct products of the reaction can be identified by counter-selection for the second marker. Further iterative character of the braid design is caused by the alternation between the two REs and the respective vector backbones

been addressed by Bryskin et al. (Bryksin and Matsumura 2010), who described the application of OE PCR for the synthesis of recombinant plasmids. Overall, OE PCR is a simple, fast and efficient DNA assembly method which can be easily applied in any standard molecular biology lab. Nevertheless, the efficiency and fidelity of the DNA polymerase used is a critical factor for the robustness and reliability of the technique. A proofreading DNA polymerase is mandatory and the final assembly product should be sequenced.

3.2.2.2 Circular Polymerase Extension Cloning

Circular polymerase extension cloning (CPEC) (Quan and Tian 2009) can be regarded as an extension of the classical OE PCR. Contrary to the linear assembly product resulting from OE PCR, the product of the CPEC assembly reaction is a circular plasmid, which can be used directly for the transformation of hosts.

Up to four fragments can be assembled into a backbone vector. The insert fragments contain overlapping homologies to their adjacent neighbours, and the first and the last fragments carry terminal overlapping homologies to the insertion locus on the vector backbone (Quan and Tian 2009). This fragment architecture causes the circularization of the assembly product. Usually, the insert fragments are PCR amplified using primers that introduce the overlapping sequence homologies. The vector backbone is linearized at the insertion locus, and all fragments must be purified before assembly. Analogous to OE PCR, the CPEC assembly proceeds through an initial denaturation and an ensuing elongation step. During the denaturation the homologous stretches are exposed on the single-stranded DNA and anneal to guide the assembly of the insert fragments with the backbone vector. The 3′ ends of the hybridized single-stranded DNA fragments act as primers for the synthesis of the complementary DNA strand. The elongation stops as soon as the 5′ end of an annealed fragment is encountered. In this way, a nicked circular dsDNA product is generated in which the number of the nicks corresponds to the number of the assembled DNA fragments. CPEC does not involve an amplification of the final assembly product. To prevent sequence errors during the assembly process, a proofreading polymerase is required. Depending on the number of inserts, up to 20 denaturation, annealing and elongation cycles are needed to generate sufficient amounts of the assembly product (Quan and Tian 2009).

Prolonged overlap extension PCR (POE PCR) (You and Zhang 2014) exploits the same principle as CPEC except that overlapping homologies are introduced to the vector backbone by PCR with appropriate primers (You and Zhang 2014). CPEC and POE PCR are simple, fast and efficient DNA assembly methods. Their products need no further subcloning and can be directly transformed into the host cells. The equipment needed for the application of the methods is standard in molecular biology labs. Similar to OE PCR, the yield and robustness of these methods depend on the used DNA polymerase. The efficiency of the method decreases with the number of assembly fragments and their size difference (De Kok et al. 2014).

3.2.2.3 Sequence- and Ligation-Independent Cloning and Seamless Ligation Cloning Extract

Sequence- and ligation-independent cloning (SLIC) (Li and Elledge 2007) relies on the 3′ exonuclease activity of the T4 DNA polymerase for the generation of homologous ssDNA overhangs, which act as anchor points for overlap-directed assembly. The resulting DNA molecule in its nicked form is then directly transformed into *E. coli* without an additional ligation step. The endogenous DNA repair machinery of the cell further stabilizes the assembly product and closes the gaps, resulting in the desired recombinant DNA molecule (Li and Elledge 2012). Up to 10 fragments can be combined seamlessly in a one-pot reaction.

An extension of SLIC was introduced recently by Zhang et al. (2012) called seamless ligation cloning extract (SLiCE) generating a link between in vitro and in vivo assembly of DNA parts. The introduction of homologous regions at the termini of the assembly fragments during a preliminary PCR step is facilitated by PCR, similar to SLIC. The actual assembly of the DNA parts is then promoted by a cell extract from *recA*1-deficient *E. coli* strains such as DH10B or JM109 (Zhang et al. 2012). This reaction step mimics in vivo homologous recombination in *E. coli* cells and results in an assembly product stable enough for the transformation into *E. coli* cells. The host DNA repair mechanism closes the recombinant DNA molecule. Both SLIC and SLiCE are fast and reliable assembly methods. However, the assembly of more than two fragments has not been demonstrated by SLiCE so far.

3.2.2.4 Ligation-Independent Cloning

Ligation-independent cloning (LIC) has been established for a long time in molecular biology following various approaches. All methods are similar to the SLIC reaction. They rely on the generation of homologous ssDNA overhangs, which hybridize in in vitro recombination events. The reaction products can be transformed directly into the host. The ssDNA overhangs can be generated by several methods, which specify the nomenclature of the LIC assembly strategy.

LIC PCR employs synthetic, C-free overlapping homologies, which are introduced into the assembly fragments by PCR with specific primers. The first 12 nucleotides at the 5′ end of the primer can be any sequence provided that the base C is excluded. The double-stranded PCR products are incubated with T4 DNA polymerase and dGTP. The 3′ exonuclease activity of the T4 DNA polymerase chews back the lower strand from the 3′ terminus to expose the synthetic overlapping homology at the 5′ terminus of the upper strand. The enzyme cannot go beyond the base G due to the presence of dGTP in the mixture, which results in 5′ single-stranded overhangs of defined length. In this regard, LIC PCR is an advanced variation of SLIC, which controls the overhang length solely by the incubation time with the T4 DNA polymerase (Li and Elledge 2012). The fragments resulting from the sequence-guided strand degradation anneal by their exposed 5′ synthetic overlapping homologies and a nicked assembly product are formed (Aslanidis and de Jong 1990). Due to the synthetic, C-free overlapping homologies, this method is not completely sequence independent.

The uridine-specific excision reagent (USER) cloning/fusion approach exploits a similar principle for the exposure of overlapping homologies, yet they are quite short. For the PCR of the assembly fragments, PCR primers are used which contain a deoxyuridine at the eighth position from their 5′ end. The PCR products are incubated with a commercial deoxyuridine-excising enzyme mix, which introduces single strand nicks at the position of the deoxyuridine. The small 7 nt tail is lost due to insufficient hydrogen bonding, and the exposed 8 nt overlapping homology can be used for hybridization (Salomonsen et al. 2014). Since only single deoxyuridine is added, this method allows quasi scarless DNA assemblies.

Longer single-stranded overhangs may be generated by using nicking endonucleases. The introduction of a recognition site for a nicking endonuclease within the 5′ end of the overlapping homology of the assembly fragment (Wang et al. 2013) serves to introduce a defined single strand break. The recognition site is positioned such that a 10–15 nt tail is generated that is removed by a denaturation step prior to hybridization. The recognition sites are integral parts of the overlapping homologies; therefore, the technique does not facilitate scarless assemblies.

All described techniques are PCR-based, fast and efficient and result in more or less scarless assembly products. Besides the specific primers, which contain non-canonical bases in the case of USER, a proofreading polymerase is required for sequence error minimization.

3.2.2.5 Ligase Cycling Reaction

The ligase cycling reaction (LCR) is an overlap-based method for the assembly of multiple fragments into a stable circular DNA molecule (De Kok et al. 2014). Contrary to the assembly strategies described before, LCR does not rely on PCR or any kind of amplification for the introduction of the homologous regions necessary for the assembly. As no DNA sequences must be introduced into the assembly parts, LCR is a truly scarless assembly method for multiple fragments. ‘Bridging oligos’ (De Kok et al. 2014), short ssDNA molecules comprising two arms that are complementary to the ends of the adjacent parts are used to mediate the assembly. The length of these bridging oligos is determined by the melting temperature (Tm) of the two complementary arms (approximately 60 °C). The bridging oligo and the linear assembly fragments are combined in a one-pot reaction containing thermostable ligase, DMSO and other buffer components (De Kok et al. 2014), and the mix is subjected to several heating/cooling cycles. During the heating to 94 °C, the dsDNA parts denature, freeing the single strands for the annealing to the bridging oligo. During the cooling step, one complementary arm of the specific DNA linker anneals to the end of the first assembly fragment, and the other complementary arm anneals to the start of the adjacent fragment. This brings the ends of both fragments in close proximity so that the thermostable ligase can join them. In the following denaturation step, the bridging oligo dissociates from the ligated DNA strand, freeing it for the successive annealing of other complementary strands. During the subsequent cooling step, the complementary strands anneal to

the ligated DNA fragment and the thermostable ligase can close the nick on the second strand, resulting in a stable dsDNA molecule. LCR was compared with other assembly methods for which identical overlaps were employed. One study showed that LCR outperforms CPEC and Gibson assembly for the assembly of multiple fragments while it was more time efficient than yeast homologous recombination, which facilitates the assembly of up to 20 DNA parts (De Kok et al. 2014). Since the method was published recently, more data has to be collected to validate this information. As LCR does not involve any amplification steps, the method unlikely generates mutations but also does not yield high amounts of DNA.

3.2.2.6 Gibson Isothermal Assembly

Gibson isothermal assembly resembles SLIC except that it employs a DNA ligase to stabilize the assembly product. Fifteen to 80 bp of overlapping homology (depending on the melting temperature) is sufficient for a reliable assembly (Gibson et al. 2009). Equimolar amounts of linear DNA fragments are combined in a one-pot reaction. The Gibson reaction mix contains a 5′ exonuclease, a proofreading DNA polymerase, a DNA ligase (Gibson and Smith 2010) as well as appropriate buffers and dNTPs. During the isothermal reaction at 50 °C, initially, all enzymes are active. The 5′ exonuclease degrades the terminal regions of the linear fragments generating single-stranded overhangs that can anneal. However, since the 5′ exonuclease is not thermostable, its activity decreases during an incubation time of 60 min. Expectedly, this thermo-instability avoids extensive chew back of the 5′ termini. Since the chew back generated by the 5′ exonuclease is unspecific, the resulting single-stranded overhangs have different lengths, and upon annealing of the homologous sequences, gaps may be form. These are filled up by the proofreading polymerase. The ligase joins the nicked (circular) intermediate generating a double-stranded assembly product. Overall, Gibson assembly is a fast and efficient DNA assembly method, which can be easily applied in all standard molecular biology labs. Again, the efficiency of the method decreases with the number of fragments used in the assembly and the size difference between those (De Kok et al. 2014).

3.2.2.7 In Vivo Assembly Strategies

In vivo DNA assembly techniques rely on the homologous recombination mechanisms of the cloning host for the generation of the assembly product. Overlapping homologies play a vital role and their design is of special importance for the assembly of very large DNA constructs such as artificial chromosomes or whole genomes (Gibson et al. 2008). The best known host for in vivo assembly strategies is the yeast *S. cerevisiae*, in which up to 39 different linear DNA parts were correctly linked in one assembly (Gibson 2009). The first method reported for cloning in yeast was transformation-associated recombination (TAR) (Larionov et al. 1996). TAR depends on the endogenous ability of the yeast cells to perform high levels of intermolecular recombination between homologous and diverged DNA during transformation of the cells. To facilitate this reaction, the insert DNA

has to be co-transformed with linearized TAR cloning vectors, which contain insert-specific target sequences at the termini. These target sequences can then be used by the cell internal homologous recombination machinery to create an assembly product. Even though this method was initially applied for the integration of assembly products into the genome of *S. cerevisiae*, the use of yeast artificial chromosomes (YACs) facilitates the generation of large, autonomously replicating constructs of several 100 kb. Recently homologous recombination in *Bacillus subtilis* was employed for the assembly of the 16.3 kb mouse mitochondrial genome in a successive fashion (Itaya et al. 2008). The advantage of these techniques is not only the reliability for assemblies with multiple fragments. A limiting factor is time: Homologous recombination assembly in yeast requires 36–72 h from the beginning to the final product (De Kok et al. 2014); in *B. subtilis* it is slightly shorter.

Overall, synthetic biology depends on the assembly of DNA parts for various purposes, which has been demonstrated by the continuous increase in assembly methods over the last years. Currently, it is possible to choose from a vast variety of techniques to determine the best possible solution for a specific assembly experiment.

Acknowledgements A.W. and M.K. have received funding from the Innovative Medicines Initiative Joint Undertaking under grant agreement n°115360, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution. M.K. has been supported by the Austrian BMWFW, BMVIT, SFG, Standortagentur Tirol and ZIT through the Austrian FFG-COMET-Funding Programme. T.V. and M.K. gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz) for funding. We would like to thank Christina Glantschnig for the fruitful discussion and careful reading of the RNAi section. The authors gratefully acknowledge support from NAWI Graz.

References

Anderson JC, Dueber JE, Leguia M et al (2010) BglBricks: a flexible standard for biological part assembly. J Biol Eng 4:1. doi:10.1186/1754-1611-4-1

Argast GM, Stephens KM, Emond MJ, Monnat RJ (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J Mol Biol 280:345–353. doi:10.1006/jmbi.1998.1886

Arnould S, Chames P, Perez C et al (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458. doi:10.1016/j.jmb.2005.10.065

Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074

Baer A, Bode J (2001) Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. Curr Opin Biotechnol 12:473–480. doi:10.1016/S0958-1669(00)00248-2

Barnes DE (2001) Non-homologous end joining as a mechanism of DNA repair. Curr Biol 11: R455–R457

Barrangou R, Horvath P (2009) The CRISPR system protects microbes against phages, plasmids. Microbe 4:224–230

Barrangou R, Fremaux C, Deveau H et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. doi:10.1126/science.1138140
Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297
Beerli RR, Barbas CF (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20:135–141. doi:10.1038/nbt0202-135
Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. Nucleic Acids Res 25:3379–3388
Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366. doi:10.1038/35053110
Bessereau JL, Wright A, Williams DC et al (2001) Mobilization of a Drosophila transposon in the *Caenorhabditis elegans* germ line. Nature 413:70–74. doi:10.1038/35092567
Bhakta MS, Henry IM, Ousterout DG et al (2013) Highly active zinc-finger nucleases by extended modular assembly. Genome Res 23:530–538. doi:10.1101/gr.143693.112
Bibikova M, Carroll D, Segal DJ et al (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21:289–297. doi:10.1128/MCB.21.1.289-297.2001
Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A 95:10570–10575
Boch J (2011) TALEs of genome targeting. Nat Biotechnol 29:135–136. doi:10.1038/nbt.1767
Boch J, Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 48:419–436. doi:10.1146/annurev-phyto-080508-081936
Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512. doi:10.1126/science.1178811
Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333:1843–1846. doi:10.1126/science.1204094
Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol 13:394–401. doi:10.1016/j.pbi.2010.04.010
Brouns SJJ, Jore MM, Lundgren M et al (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321:960–964. doi:10.1126/science.1159689
Brummelkamp TR, Bernards R, Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. Science 296:550–553. doi:10.1126/science.1068999
Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48:463–465. doi:10.2144/000113418
Cabaniols J-P, Mathis L, Delenda C (2009) Targeted gene modifications in drug discovery and development. Curr Opin Pharmacol 9:657–663. doi:10.1016/j.coph.2009.05.001
Caldecott KW (2008) Single-strand break repair and genetic disease. Nat Rev Genet 9:619–631
Capecchi MR (1989) Altering the genome by homologous recombination. Science 244:1288–1292. doi:10.1126/science.2660260
Carr PA, Church GM (2009) Genome engineering. Nat Biotechnol 27:1151–1162. doi:10.1038/nbt.1590
Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188:773–782. doi:10.1534/genetics.111.131433
Cary LC, Goebel M, Corsaro BG et al (1989) Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology 172:156–169
Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82. doi:10.1093/nar/gkr218
Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29:3757–3774
Chevalier BS, Kortemme T, Chadsey MS et al (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905

Christian M, Cermak T, Doyle EL et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761. doi:10.1534/genetics.110.120717

Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. doi:10.1126/science.1231143

Conrad TM, Lewis NE, Palsson BØ (2011) Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol 7:509. doi:10.1038/msb.2011.42

Cooley L, Kelley R, Spradling A (1988) Insertional mutagenesis of the Drosophila genome with single P elements. Science 239:1121–1128

Cousineau B, Smith D, Lawrence-Cavanagh S et al (1998) Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell 94:451–462

Cradick TJ, Fine EJ, Antico CJ, Bao G (2013) CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res 41:9584–9592. doi:10.1093/nar/gkt714

Craig NL (1988) The mechanism of conservative site-specific recombination. Annu Rev Genet 22:77–105. doi:10.1146/annurev.ge.22.120188.000453

De Kok S, Stanton LH, Slaby T et al (2014) Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 3:97–106. doi:10.1021/sb4001992

Doench JG, Petersen CP, Sharp PA (2003) siRNAs can function as miRNAs. Genes Dev 17:438–442. doi:10.1101/gad.1064703

Donzé O, Picard D (2002) RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase. Nucleic Acids Res 30:e46. doi:10.1093/nar/30.10.e46

Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128:2477–2484. doi:10.1021/ja057519l

Dudás A, Chovanec M (2004) DNA double-strand break repair by homologous recombination. Mutat Res 566:131–167. doi:10.1016/j.mrrev.2003.07.001

Dujon B, Belfort M, Butow RA et al (1989) Mobile introns: definition of terms and recommended nomenclature. Gene 82:115–118

Dumais Pope C, Dhand L, Cianciotto NP (1994) Random mutagenesis of *Legionella pneumophila* with mini-Tn10. FEMS Microbiol Lett 124:107–111. doi:10.1111/j.1574-6968.1994.tb07269.x

Eckert-Boulet N, Rothstein R, Lisby M (2011) Cell biology of homologous recombination in yeast. Methods Mol Biol 745:523–536. doi:10.1007/978-1-61779-129-1_30

Edgell DR (2009) Selfish DNA: homing endonucleases find a home. Curr Biol 19:R115–R117. doi:10.1016/j.cub.2008.12.019

Elbashir SM (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 15:188–200. doi:10.1101/gad.862301

Elbashir SM, Harborth J, Lendeckel W et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498. doi:10.1038/35078107

Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4:457–469

Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98:6742–6746. doi:10.1073/pnas.121164898

Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 3:109–118. doi:10.1039/c0ib00070a

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647. doi:10.1371/journal.pone.0003647

Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4, e5553. doi:10.1371/journal.pone.0005553

Enyeart PJ, Chirieleison SM, Dao MN et al (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. Mol Syst Biol 9:685. doi:10.1038/msb.2013.41

Eskes R, Yang J, Lambowitz AM, Perlman PS (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. Cell 88: 865–874

Esvelt KM, Wang HH (2013) Genome-scale engineering for systems and synthetic biology. Mol Syst Biol 9:641. doi:10.1038/msb.2012.66

Feng X, Bednarz AL, Colloms SD (2010) Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. Nucleic Acids Res 38:1204–1216. doi:10.1093/nar/gkp1068

Fernandes PJ, Powell JA, Archer JA (2001) Construction of Rhodococcus random mutagenesis libraries using Tn5 transposition complexes. Microbiology 147:2529–2536

Frazier CL, San Filippo J, Lambowitz AM, Mills DA (2003) Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. Appl Environ Microbiol 69:1121–1128. doi:10.1128/AEM.69.2.1121

Fu Y, Foden JA, Khayter C et al (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826. doi:10.1038/nbt.2623

Fu Y, Sander JD, Reyon D et al (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol 32:279–284. doi:10.1038/nbt.2808

Gaj T, Gersbach CA, Barbas CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31:397–405. doi:10.1016/j.tibtech.2013.04.004

Gasiunas G, Siksnys V (2013) RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? Trends Microbiol 1–6. doi:10.1016/j.tim.2013.09.001

Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109:E2579–E2586. doi:10.1073/pnas.1208507109

Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res 37:6984–6990. doi:10.1093/nar/gkp687

Gibson DG, Smith HO (2010) US Patent 7,776,532 B2. Method for in vitro recombination

Gibson DG, Benders GA, Andrews-Pfannkoch C et al (2008) Complete chemical synthesis, assembly and cloning of a mycoplasma genitalium genome. Science 319:1215–1220

Gibson DG, Young L, Chuang R et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345. doi:10.1038/nmeth.1318

Gibson DG, Glass JI, Lartigue C et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56. doi:10.1126/science.1190719

Gilbert LA, Larson MH, Morsut L et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154:442–451. doi:10.1016/j.cell.2013.06.044

Gimble FS, Wang J (1996) Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. J Mol Biol 263:163–180. doi:10.1006/jmbi.1996.0567

Gonzalez B, Schwimmer LJ, Fuller RP et al (2010) Modular system for the construction of zinc-finger libraries and proteins. Nat Protoc 5:791–810

Gordley RM, Gersbach CA, Barbas CF (2009) Synthesis of programmable integrases. Proc Natl Acad Sci U S A 106:5053–5058. doi:10.1073/pnas.0812502106

Grabundzija I, Irgang M, Mátés L et al (2010) Comparative analysis of transposable element vector systems in human cells. Mol Ther 18:1200–1209. doi:10.1038/mt.2010.47

Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinf 8:172. doi:10.1186/1471-2105-8-172

Grizot S, Duclert A, Thomas S et al (2011) Context dependence between subdomains in the DNA binding interface of the I-CreI homing endonuclease. Nucleic Acids Res 39:6124–6136. doi:10.1093/nar/gkr186

Guo H (2000) Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science 289:452–457. doi:10.1126/science.289.5478.452

Guo J, Gaj T, Barbas CF (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol 400:96–107. doi:10.1016/j.jmb.2010.04.060

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 1:e60. doi:10.1371/journal.pcbi.0010060

Hamilton DL, Abremski K (1984) Site-specific recombination by the bacteriophage P1 lox-Cre system: cre-mediated synapsis of two lox sites. J Mol Biol 178:481–486. doi:10.1016/0022-2836(84)90154-2

Händel EM, Alwin S, Cathomen T (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol Ther 17:104–111. doi:10.1038/mt.2008.233

He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5:522–531

Heinemann M, Panke S (2006) Synthetic biology-putting engineering into biology. Bioinformatics 22:2790–2799. doi:10.1093/bioinformatics/btl469

Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16:7351–7367

Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1:14–21. doi:10.1021/sb2000116

Ho SN, Hunt HD, Horton RM et al (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59

Horton RM, Hunt HD, Ho SN et al (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68

Hsu PD, Scott DA, Weinstein JA et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827–832. doi:10.1038/nbt.2647

Hutvágner G, Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex. Science 297:2056–2060. doi:10.1126/science.1073827

Isaacs FJ, Carr PA, Wang HH et al (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333:348–353. doi:10.1126/science.1205822

Isalan M, Klug A, Choo Y (1998) Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. Biochemistry 37:12026–12033. doi:10.1021/bi981358z

Itaya M, Fujita K, Kuroki A, Tsuge K (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nat Methods 5:41–43

Ivics Z, Izsvák Z (2010) The expanding universe of transposon technologies for gene and cell engineering. Mob DNA 1:25. doi:10.1186/1759-8753-1-25

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Jackson AL, Burchard J, Schelter J et al (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA 12:1179–1187. doi:10.1261/rna.25706

Jansen R, van Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43:1565–1575

Jasin M (1996) Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet 12:224–228

Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. doi:10.1126/science.1225829

Karberg M, Guo H, Zhong J et al (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol 19:1162–1167. doi:10.1038/nbt1201-1162

Karbiener M, Glantschnig C, Scheideler M (2014) Hunting the needle in the haystack: a guide to obtain biologically meaningful microRNA targets. Int J Mol Sci 15:20266–20289. doi:10.3390/ijms151120266

Kay S, Hahn S, Marois E et al (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 318:648–651. doi:10.1126/science.1144956

Knight T (2008) Draft standard for biobrick BB-2 biological parts. http://hdl.handle.net/1721.1/45139. Accessed 20 Dec 2014

Knight SW, Bass BL (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. Science 293:2269–2271. doi:10.1126/science.1062039

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294:853–858. doi:10.1126/science.1064921

Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol 11:39–46

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622. doi:10.1146/annurev.bi.62.070193.003103

Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol 3:a003616. doi:10.1101/cshperspect.a003616

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS (1999) Group I and group II ribozymes as RNPs: clues to the past and guides to the future. Cold Spring Harb Monogr Ser 37:451–486

Larionov V, Kouprina N, Graves J et al (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. Proc Natl Acad Sci U S A 93:491–496

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294:858–862. doi:10.1126/science.1065062

Le Breton Y, Mohapatra NP, Haldenwang WG (2006) In vivo random mutagenesis of Bacillus subtilis by use of TnYLB-1, a mariner-based transposon. Appl Environ Microbiol 72:327–333. doi:10.1128/AEM.72.1.327-333.2006

Lee RC, Ambros V (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294:862–864. doi:10.1126/science.1065329

Lee Y, Ahn C, Han J et al (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425:415–419. doi:10.1038/nature01957

Lee TS, Krupa RA, Zhang F et al (2011) BglBrick vectors and datasheets: a synthetic biology platform for gene expression. J Biol Eng 5:12. doi:10.1186/1754-1611-5-12

Lewis BP, Shih I, Jones-Rhoades MW et al (2003) Prediction of mammalian microRNA targets. Cell 115:787–798. doi:10.1016/S0092-8674(03)01018-3

Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256. doi:10.1038/nmeth1010

Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. Methods Mol Biol 852:51–59. doi:10.1007/978-1-61779-564-0_5

Li L, Wu LP, Chandrasegaran S (1992) Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci U S A 89:4275–4279

Li T, Huang S, Zhao X et al (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 39:6315–6325. doi:10.1093/nar/gkr188

Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181–211. doi:10.1146/annurev.biochem.052308.093131

Lim LP, Lau NC, Garrett-Engele P et al (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769–773. doi:10.1038/nature03315

Liu Q, Segal DJ, Ghiara JB, Barbas CF (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc Natl Acad Sci U S A 94:5525–5530

Loonstra A, Vooijs M, Beverloo HB et al (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proc Natl Acad Sci U S A 98:9209–9214. doi:10.1073/pnas.161269798

Lu Q (2005) Seamless cloning and gene fusion. Trends Biotechnol 23:199–207. doi:10.1016/j.tibtech.2005.02.008

Lytle JR, Yario TA, Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci U S A 104:9667–9672. doi:10.1073/pnas.0703820104

Makarova KS, Haft DH, Barrangou R et al (2011) Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol 9:467–477

Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322:1843–1845. doi:10.1126/science.1165771

Mastroianni M, Watanabe K, White TB et al (2008) Group II intron-based gene targeting reactions in eukaryotes. PLoS One 3:e3121. doi:10.1371/journal.pone.0003121

Matranga C, Tomari Y, Shin C et al (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123:607–620. doi:10.1016/j.cell.2005.08.044

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci 36:344–355. doi:10.1073/pnas.36.6.344

Michel F, Feral J (1995) Structure and activities of group II introns. Annu Rev Biochem 64:435–461. doi:10.1146/annurev.bi.64.070195.002251

Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. EMBO J 4:1609–1614

Miller JC, Holmes MC, Wang J et al (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25:778–785. doi:10.1038/nbt1319

Miller JC, Tan S, Qiao G et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148. doi:10.1038/nbt.1755

Mills DA, Manias DA, McKay LL, Dunny GM (1997) Homing of a group II intron from *Lactococcus lactis* subsp. lactis ML3. J Bacteriol 179:6107–6111

Miskey C (2003) The frog prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucleic Acids Res 31:6873–6881. doi:10.1093/nar/gkg910

Misquitta L, Paterson BM (1999) Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. Proc Natl Acad Sci U S A 96:1451–1456

Mojica FJM, Díez-Villaseñor C, Soria E, Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. Mol Microbiol 36:244–246. doi:10.1046/j.1365-2958.2000.01838.x

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 60:174–182. doi:10.1007/s00239-004-0046-3

Morbitzer R, Römer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 107:21617–21622. doi:10.1073/pnas.1013133107

Mosberg JA, Lajoie MJ, Church GM (2010) Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. Genetics 186:791–799. doi:10.1534/genetics.110.120782

Mosberg JA, Gregg CJ, Lajoie MJ et al (2012) Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. PLoS One 7:e44638. doi:10.1371/journal.pone.0044638

Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501. doi:10.1126/science.1178817

Mourelatos Z, Dostie J, Paushkin S et al (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 16:720–728. doi:10.1101/gad.974702

Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol 180:2063–2071

Mussolino C, Cathomen T (2012) TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol 23:644–650. doi:10.1016/j.copbio.2012.01.013

Mussolino C, Morbitzer R, Lütge F et al (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res 39:9283–9293. doi:10.1093/nar/gkr597

Mussolino C, Alzubi J, Fine EJ et al (2014) TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. Nucleic Acids Res 42:6762–6773. doi:10.1093/nar/gku305

Muyrers JP, Zhang Y, Testa G, Stewart AF (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res 27:1555–1557

Muyrers JP, Zhang Y, Buchholz F, Stewart AF (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. Genes Dev 14:1971–1982. doi:10.1101/gad.14.15.1971

Nakayama M, Ohara O (2005) Improvement of recombination efficiency by mutation of red proteins. Biotechniques 38:917–924. doi:10.2144/05386RR02

Nern A, Pfeiffer BD, Svoboda K, Rubin GM (2011) Multiple new site-specific recombinases for use in manipulating animal genomes. Proc Natl Acad Sci U S A 108:14198–14203. doi:10.1073/pnas.1111704108

Nikolova D, Toncheva D (2008) RNA interference – regulations and application in oncology. J Cancer Mol 4:67–77

Niwa R, Slack FJ (2007a) The evolution of animal microRNA function. Curr Opin Genet Dev 17:145–150. doi:10.1016/j.gde.2007.02.004

Niwa R, Slack FJ (2007b) Ins and outs of RNA interference analysis. Cell Press, Cambridge, pp 34–36

Owens JB, Mauro D, Stoytchev I et al (2013) Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. Nucleic Acids Res 41:9197–9207. doi:10.1093/nar/gkt677

Pabo CO, Peisach E, Grant RA (2001) Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem 70:313–340. doi:10.1146/annurev.biochem.70.1.313

Paddison PJ, Caudy AA, Bernstein E et al (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev 16:948–958. doi:10.1101/gad.981002

Parinov S, Kondrichin I, Korzh V, Emelyanov A (2004) Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. Dev Dyn 231:449–459. doi:10.1002/dvdy.20157

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252:809–817. doi:10.1126/science.2028256

Perez-Pinera P, Ousterout DG, Gersbach CA (2012) Advances in targeted genome editing. Curr Opin Chem Biol 16:268–277. doi:10.1016/j.cbpa.2012.06.007

Perler FB, Davis EO, Dean GE et al (1994) Protein splicing elements: inteins and exteins-a definition of terms and recommended nomenclature. Nucleic Acids Res 22:1125–1127

Perutka J, Wang W, Goerlitz D, Lambowitz AM (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. J Mol Biol 336:421–439. doi:10.1016/j.jmb.2003.12.009

Phillips IE, Silver PA (2006) A new biobrick assembly strategy designed for facile protein engineering. http://hdl.handle.net/1721.1/32535. Accessed 20 Dec 2014

Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol Life Sci 62:685–707. doi:10.1007/s00018-004-4513-1

Proteau G, Sidenberg D, Sadowski P (1986) The minimal duplex DNA sequence required for site-specific recombination promoted by the FLP protein of yeast in vitro. Nucleic Acids Res 14: 4787–4802. doi:10.1093/nar/14.12.4787

Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441. doi:10.1371/journal.pone.0006441

Ramirez CL, Certo MT, Mussolino C et al (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic Acids Res 40:5560–5568. doi:10.1093/nar/gks179

Ran FA, Hsu PD, Lin C et al (2014) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. doi:10.1016/j.cell.2013.08.021.Double

Ratcliff F, MacFarlane S, Baulcombe D (1999) Gene silencing without DNA. RNA-mediated cross-protection between viruses. Plant Cell 11:1207–1216

Rawsthorne H, Turner KN, Mills DA (2006) Multicopy integration of heterologous genes, using the lactococcal group II intron targeted to bacterial insertion sequences. Appl Environ Microbiol 72:6088–6093. doi:10.1128/AEM.02992-05

Reyon D, Tsai SQ, Khayter C et al (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30:460–465. doi:10.1038/nbt.2170

Rivas FV, Tolia NH, Song JJ et al (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12:340–349. doi:10.1038/nsmb918

Roberts RJ (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res 31:1805–1812. doi:10.1093/nar/gkg274

Römer P, Hahn S, Jordan T et al (2007) Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science 318:645–648. doi:10.1126/science.1144958

Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14:8096–8106

Rushforth AM, Saari B, Anderson P (1993) Site-selected insertion of the transposon Tc1 into a *Caenorhabditis elegans* myosin light chain gene. Mol Cell Biol 13:902–910. doi:10.1128/MCB.13.2.902.Updated

Sakuma T, Nishikawa A, Kume S et al (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci Rep 4:5400. doi:10.1038/srep05400

Salomonsen B, Mortensen U, Halkier B (2014) USER-derived cloning methods and their primer design. In: Valla S, Lale R (eds) DNA cloning assembly methods SE – 5. Humana Press, Totowa, pp 59–72

Sanjana NE, Cong L, Zhou Y et al (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7:171–192. doi:10.1038/nprot.2011.431

Sarrion-Perdigones A, Falconi EE, Zandalinas SI et al (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS One 6:e21622. doi:10.1371/journal.pone.0021622

Sarrion-Perdigones A, Vazquez-Vilar M, Palací J et al (2013) GoldenBraid2.0: a comprehensive DNA assembly framework for plant synthetic biology. Plant Physiol. doi:10.1104/pp.113.217661

Sauer B (1987) Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol 7:2087–2096

Sauer B (1994) Recycling selectable markers in yeast. Biotechniques 16:1086–1088

Schlake T, Bode J (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33:12746–12751. doi:10.1021/bi00209a003

Schmid-Burgk JL, Schmidt T, Kaiser V et al (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. Nat Biotechnol 31: 76–81. doi:10.1038/nbt.2460

Scholze H, Boch J (2010) TAL effector-DNA specificity. Virulence 1:428–432. doi:10.4161/viru.1.5.12863

Seligman LM, Chisholm KM, Chevalier BS et al (2002) Mutations altering the cleavage specificity of a homing endonuclease. Nucleic Acids Res 30:3870–3879

Shah SA, Erdmann S, Mojica FJM, Garrett RA (2013) Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol 10:891–899. doi:10.4161/rna.23764

Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc 4:206–223

Shimizu Y, Bhakta MS, Segal DJ (2009) Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. Bioorg Med Chem Lett 19:3970–3972. doi:10.1016/j.bmcl.2009.02.109

Silver DP, Livingston DM (2014) Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. Mol Cell 8:233–243. doi:10.1016/S1097-2765(01)00295-7

Silver PA, Way JC, Arnold FH, Meyerowitz JT (2014) Synthetic biology: engineering explored. Nature 509:166–167. doi:10.1038/509166a

Singh NN, Lambowitz AM (2001) Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. J Mol Biol 309:361–386. doi:10.1006/jmbi.2001.4658

Skipper KA, Andersen PR, Sharma N, Mikkelsen JG (2013) DNA transposon-based gene vehicles – scenes from an evolutionary drive. J Biomed Sci 20:92. doi:10.1186/1423-0127-20-92

Sledz CA, Williams BRG (2005) RNA interference in biology and disease. Blood 106:787–794. doi:10.1182/blood-2004-12-4643

Sledz CA, Holko M, de Veer MJ et al (2003) Activation of the interferon system by short-interfering RNAs. Nat Cell Biol 5:834–839

Smih F, Rouet P, Romanienko PJ, Jasin M (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic Acids Res 23:5012–5019. doi:10.1093/nar/23.24.5012

Smith D, Zhong J, Matsuura M et al (2005) Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. Genes Dev 19:2477–2487. doi:10.1101/gad.1345105

Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. Biotechnol Adv 23:431–469. doi:10.1016/j.biotechadv.2005.03.003

Sternberg N, Hamilton D, Hoess R (1981) Bacteriophage P1 site-specific recombination: II. Recombination between loxP and the bacterial chromosome J Mol Biol 150:487–507. doi:10.1016/0022-2836(81)90376-4

Storici F, Durham CL, Gordenin DA, Resnick MA (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc Natl Acad Sci U S A 100:14994–14999. doi:10.1073/pnas.2036296100

Streubel J, Blücher C, Landgraf A, Boch J (2012) TAL effector RVD specificities and efficiencies. Nat Biotechnol 30:593–595. doi:10.1038/nbt.2304

Sun N, Abil Z, Zhao H (2012) Recent advances in targeted genome engineering in mammalian systems. Biotechnol J 7:1074–1087. doi:10.1002/biot.201200038

Thibault ST, Singer M, Miyazaki WY et al (2004) A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nat Genet 36:283–287. doi:10.1038/ng1314

Thierry A, Dujon B (1992) Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. Nucleic Acids Res 20:5625–5631. doi:10.1093/nar/20.21.5625

Tomari Y, Matranga C, Haley B et al (2004) A protein sensor for siRNA asymmetry. Science 306:1377–1380. doi:10.1126/science.1102755

Torres R, García A, Payá M, Ramirez JC (2011) Non-integrative lentivirus drives high-frequency Cre-mediated cassette exchange in human cells. PLoS One 6, e19794. doi:10.1371/journal.pone.0019794

Tremblay JP, Chapdelaine P, Coulombe Z, Rousseau J (2012) Transcription activator-like effector proteins induce the expression of the frataxin gene. Hum Gene Ther 23:883–890. doi:10.1089/hum.2012.034

Tsai SQ, Wyvekens N, Khayter C et al (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. doi:10.1038/nbt.2908

Turan S, Galla M, Ernst E et al (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. J Mol Biol 407:193–221. doi:10.1016/j.jmb.2011.01.004

Uhde-Stone C, Gor N, Chin T et al (2013) A do-it-yourself protocol for simple transcription activator-like effector assembly. Biol Proced Online 15:3. doi:10.1186/1480-9222-15-3

Urnov FD, Rebar EJ, Holmes MC et al (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646

Van den Ackerveken G, Marois E, Bonas U (1996) Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. Cell 87:1307–1316

Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr Opin Biotechnol 24:1094–1101. doi:10.1016/j.copbio.2013.02.024

Wang HH, Isaacs FJ, Carr PA et al (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898. doi:10.1038/nature08187

Wang HH, Huang P-Y, Xu G et al (2012a) Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. ACS Synth Biol 1:43–52. doi:10.1021/sb3000029

Wang HH, Kim H, Cong L et al (2012b) Genome-scale promoter engineering by coselection MAGE. Nat Methods 9:591–593. doi:10.1038/nmeth.1971

Wang RY, Shi ZY, Guo YY et al (2013) DNA fragments assembly based on nicking enzyme system. PLoS One 8:e57943. doi:10.1371/journal.pone.0057943

Weber E, Gruetzner R, Werner S et al (2011a) Assembly of designer TAL effectors by golden gate cloning. PLoS One 6:e19722. doi:10.1371/journal.pone.0019722

Weber E, Engler C, Gruetzner R et al (2011b) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6, e16765. doi:10.1371/journal.pone.0016765

Weil CF, Kunze R (2000) Transposition of maize Ac/Ds transposable elements in the yeast *Saccharomyces cerevisiae*. Nat Genet 26:187–190

Werner S, Engler C, Weber E et al (2012) Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng Bugs 3:38–43. doi:10.4161/bbug.3.1.18223

Westra ER, Swarts DC, Staals RHJ et al (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu Rev Genet 46:311–339. doi:10.1146/annurev-genet-110711-155447

Wicker T, Sabot F, Hua-Van A et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Yan W, Smith C, Cheng L (2013) Expanded activity of dimer nucleases by combining ZFN and TALEN for genome editing. Sci Rep 3:2376. doi:10.1038/srep02376

Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:e50. doi:10.1093/nar/gkm089

You C, Zhang Y-HP (2014) Simple cloning and DNA assembly in *Escherichia coli* by prolonged overlap extension PCR. Methods Mol Biol 1116:183–192. doi:10.1007/978-1-62703-764-8_13

Yu D, Ellis HM, Lee EC et al (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A 97:5978–5983. doi:10.1073/pnas.100127597

Zeng Y, Yi R, Cullen BR (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A 100:9779–9784. doi:10.1073/pnas.1630797100

Zhang F, Cong L, Lodato S et al (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153. doi:10.1038/nbt.1775

Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 40:e55. doi:10.1093/nar/gkr1288

Zhang M, Wang F, Li S et al (2014) TALE: a tale of genome editing. Prog Biophys Mol Biol 114: 25–32. doi:10.1016/j.pbiomolbio.2013.11.006

Zhenyu Z, Huang-Yong X (2002) US Patent 6723546 B2. Method for cloning and expression of BsaI restriction endonuclease and BsaI methylase in *E. coli*. 27

Protein Building Blocks and the Expansion of the Genetic Code

4

Birgit Wiltschi

Contents

4.1 Protein Building Blocks 144
4.1.1 Canonical Amino Acids 144
4.1.2 Noncanonical Amino Acids 146
4.2 Assembling the Building Blocks into (Poly)peptides 151
4.2.1 Non-ribosomal Peptide Synthesis 151
4.2.2 The Genetic Code 152
4.2.3 Ribosomal Translation 155
4.2.4 Natural Exceptions of the Standard Genetic Code 156
4.2.5 Changing the Genetic Code 160
4.2.6 Incorporation of ncAAs in In Vitro Expression Systems 173
4.2.7 Engineering the Genetic Code: In Vivo Residue-Specific Incorporation of ncAAs 174
4.2.8 Expansion of the Genetic Code: Site-Specific Incorporation of ncAAs 183
References 198

Abstract

The proteins of all known organisms are built of a set of 20 canonical amino acids prescribed by the genetic code. Many more amino acids occur in nature but they are excluded from ribosomal translation. Nevertheless, nature exploits their vast chemical diversity for the production of highly bioactive peptides by non-ribosomal biosynthesis routes. The extraordinarily rich structural and functional repertoire of the noncanonical amino acids holds great promise for the future of protein engineering, yet we have only just begun to tap the cornucopia of noncanonical building blocks for the biosynthesis of synthetic proteins.

This chapter provides a broad overview of canonical as well as noncanonical amino acids as building blocks for proteins and peptides. It recapitulates the

B. Wiltschi (✉)
Austrian Centre of Industrial Biotechnology – ACIB, Petersgasse 14, 8010 Graz, Austria
e-mail: birgit.wiltschi@acib.at

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_4

genetic code and its natural deviations. The structures of selected naturally occurring noncanonical amino acids are listed referencing their source and biosynthesis pathways where known. The principles of current approaches to engineer and expand the genetic code are described. Numerous examples illustrate their application in protein engineering, and they are complemented by a compilation of the noncanonical building blocks involved.

Abbreviations

aaRS	Aminoacyl-tRNA synthetase
BOC	Bioorthogonal conjugation
BONCAT	Bioorthogonal noncanonical amino acid tagging
cAA	Canonical amino acid
CAGE	Conjugative assembly genome engineering
CFP	Cyan fluorescent protein
CuAAC	Copper(I)-catalyzed azide–alkyne cycloaddition
FUNCAT	Fluorescent noncanonical amino acid tagging
GFP	Green fluorescent protein
IEDDA	Inverse electron-demand Diels–Alder reaction
MAGE	Multiplex automated genome engineering
*Mm*PylRS/*Mm*tRNA$_{CUA}^{Pyl}$	Orthogonal PylRS/tRNA$_{CUA}^{Pyl}$ pair from *Methanosarcina mazei*
ncAA	Noncanonical amino acid
ncAARS	ncAA-specific aminoacyl-tRNA synthetase
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthesis
o-pair	Orthogonal pair
PFAA	Polyfluorinated amino acid
PTM	Posttranslational modification
SCR	Sense codon recoding
SCS	Stop codon suppression
SpAAC	Strain-promoted Huisgen 1,3-dipolar cycloadditions between azides and cyclooctynes
SPI	Supplementation-based incorporation

4.1 Protein Building Blocks

4.1.1 Canonical Amino Acids

All currently known living organisms use amino acids as the basic building blocks of their peptides and proteins. Although more than 500 amino acids occur naturally (Hunt 1985; Wagner and Musso 1983; Walsh et al. 2013), only a subset of 20 is

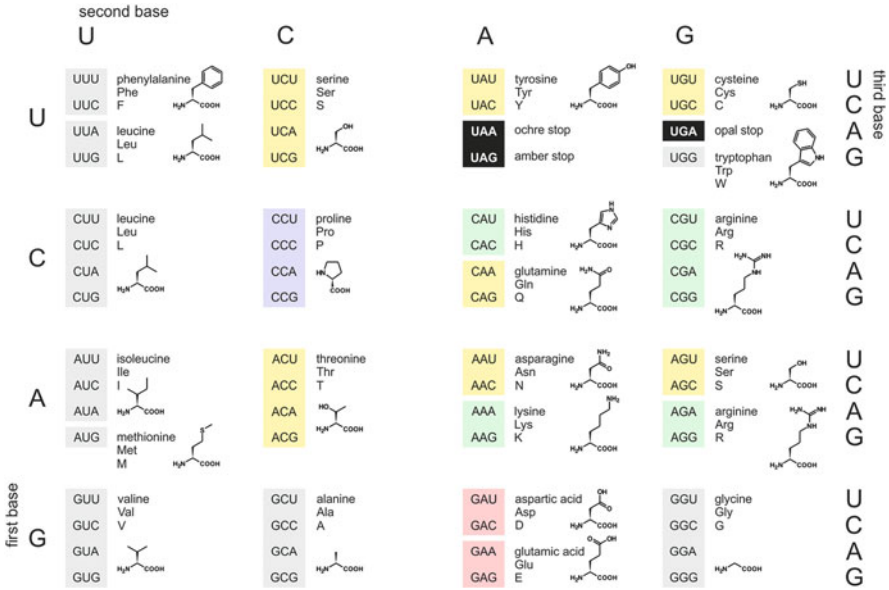


Fig. 4.1 The canonical amino acids and the genetic code. Codons that encode amino acids with hydrophobic side chains are highlighted in *gray*; *yellow*, polar side chains; *red*, acidic side chains; *green*, basic side chains; and *purple*, cyclic side chain. The translation termination codons are highlighted in *black*

used for the biosynthesis of proteins and peptides by ribosomal translation. Often, these amino acids are termed proteinogenic, standard or natural. I will rather call them canonical amino acids (cAAs) as they are the 20 protein building blocks that are prescribed by the standard genetic code (Fig. 4.1). As we shall see below, a number of amino acids, natural or synthetic, can be incorporated into proteins although they are not prescribed by the genetic code. I will refer to these as the noncanonical amino acids (ncAAs).

4.1.1.1 Physicochemical Properties and Chemical Diversity of the Canonical Amino Acids

Amino acids are composed of at least one amino and one carboxyl group along with a side chain, which is specific to each amino acid. CAAs are 2- or alpha-amino acids; the amino group and the carboxyl group are connected via a central carbon atom that carries the side chain. Two steric conformations are possible around the central C atom, but cAAs comprise exclusively the L-enantiomers. The smallest cAA, glycine, is an exemption: it carries two protons at the central C atom and is therefore achiral.

At physiological pH, the amino group is protonated (positively charged) and the carboxyl group is deprotonated (negatively charged) which is why the amino acids behave as zwitterions (Bischoff and Schlüter 2012).

The individual side chains of the 20 cAAs are unique with regard to size, shape, and polarity (Fig. 4.1). The side chains of alanine (Ala), valine (Val), leucine (Leu), isoleucine (Ile), methionine (Met), phenylalanine (Phe), and tryptophan (Trp) are hydrophobic and increase in size in the given order. In soluble proteins, they are usually buried in the protein interior. Cysteine (Cys), serine (Ser), threonine (Thr), tyrosine (Tyr), glutamine (Gln), and asparagine (Asn) carry polar side chains. The side chains of aspartic acid (Asp) and glutamic acid (Glu) can ionize to carry negative charge, while lysine (Lys), histidine (His), and arginine (Arg) can be positively charged [Bischoff and Schlüter (2012) and references therein].

However, the chemical diversity among the side chains of the cAAs is rather low (O'Donoghue et al. 2013). Linear and branched aliphatic side chains (as in Gly, Ala, Val, Ile, Leu) occur frequently. Four cAAs (Phe, Tyr, Trp, and His) are aromatic. Hydroxyl groups are found in three cAAs (Ser, Thr, and Tyr). Two amino acids carry carboxyl groups (Glu, Asp) or their amides (Gln, Asn). Sulfur is present in the side chains of Cys (thiol) and Met (thioether), which renders them prone to oxidation. Lys contains an amino group and Arg an amine-containing guanidino group. Proline (Pro) is the only cAA with a secondary amine group and a cyclic side chain. The unprotonated side chains of Arg, Lys, His, Cys, Asp, Glu, and Tyr can act as potent nucleophiles (Bischoff and Schlüter 2012).

All cAAs except Leu, Ile, Val, Ala, and Phe were observed to be chemically modified after their translation into proteins (Walsh et al. 2005). Posttranslational modifications, such as phosphorylation, glycosylation, alkylation, or acylation to name but a few, can increase the chemical diversity of proteins by at least two orders of magnitude (Walsh 2014).

4.1.2 Noncanonical Amino Acids

As pointed out in Sect. 4.1.1, I refer to cAAs as the 20 L-amino acids that are prescribed by the genetic code. For other amino acids, either of natural or synthetic origin, I will call them noncanonical to avoid terms such as "unnatural" or "nonproteinogenic." Many of these compounds can be introduced into proteins and peptides in the lab, or they even occur in nature.

4.1.2.1 Natural Occurrence of ncAAs

The number of naturally occurring L- and D-amino acids that are excluded from ribosomal translation is at least 20 times larger than the small set of cAAs (Hunt 1985; Walsh et al. 2013).

Natural ncAAs are mainly found in plants and microorganisms. Often, the microorganisms secrete them into their environment where some have antibiotic activity (Hunt 1985). Many natural ncAAs are (end) products of secondary metabolism or intermediates of metabolic pathways. Occasionally, they originate from the metabolism or detoxification of foreign compounds (Hunt 1985). Other naturally occurring ncAAs, such as ornithine, citrulline, and homoserine, are metabolic intermediates of the primary metabolism (Walsh et al. 2013).

Besides occurring as free amino acids, ncAAs are constituents of low molecular weight compounds, for instance, with glutamic acid by an isopeptide linkage, or they constitute parts of pigments (Hunt 1985). Other natural sources of ncAAs are depsipeptides (Hunt 1985), non-ribosomal peptides, or hybrid polyketide scaffolds (Walsh et al. 2013).

4.1.2.2 The Structural and Chemical Diversity of Naturally Occurring ncAAs

The structures of selected naturally occurring ncAAs are shown in Table 4.1. Like the cAAs, many natural ncAAs have aliphatic side chains. The chain lengths do not often exceed six carbon atoms although there are some rather large residues (**1**, **2**)[1]. Chemical diversity in the ncAAs is attained by *C*-, *O*-, or *N*-methylation (**3**, **4**); by hydroxyl (**5**, **6**), carboxyl (**7**), and amino (**8**, **9**) substitution of the side chains; or by cyclization (**10**). Other types of nitrogenous functions besides amines include guanidino (**11**) and cyano (**12**) groups. Numerous sulfur or selenium analogs (**13**) of Cys and Met also occur. Aldehydes and ketones (**14**, **15, 16**) are present in the side chains of ncAAs as well as chlorine (**17**). Notably, some ncAAs contain unsaturated groups, such as alkenyl (**10**, **18**) or alkynyl (**19**) functional groups. These may serve as reactive groups for bioorthogonal conjugations (see below) (Hunt 1985; Walsh et al. 2013).

Aromatic ncAAs based on a benzene ring are relatively limited in comparison to the aliphatic ncAAs. Analogs containing chlorinated phenol rings can be found in microorganisms (Hunt 1985).

Heterocycles, both aromatic and aliphatic, form a large body of ncAAs. Many of them contain nitrogen, fewer oxygen, and sulfur as heteroatoms (**20**, **21**, **22**) (Hunt 1985).

D-Amino acids also occur naturally; they are often generated from their L-isomers by amino acid racemases and epimerases (Walsh et al. 2013).

4.1.2.3 The Biosynthesis of Natural ncAAs

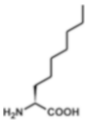
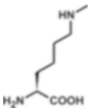
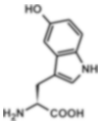
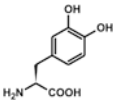
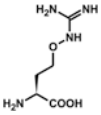
The biosynthesis routes of only a handful of compounds are known to date (Walsh et al. 2013) (see the selected examples in Table 4.1).

A probable route for many familiar-looking analogs is that of the modification of an existing cAA by similar mechanisms as those involved in posttranslational modifications (Hunt 1985). For instance, 3,4-dihydroxy-L-phenylalanine (**6**) is produced by enzymatic hydroxylation of tyrosine (Nagatsu et al. 1964). The synthesis of delta-5-acetylornithine or *O*-acetylserine is another well-documented example of such modifications. Hydroxyproline, desmosine and isodesmosine, or pyridinoline, occur in mammalian urine. They arise from posttranslational modification of cAAs in proteins and are released as free amino acids after metabolic turnover (Hunt 1985).

Some ncAAs occur as intermediates or by-products of cellular biosynthesis pathways. Their production can be upregulated under certain conditions. For instance, the methionine analog norleucine (**26**) is a by-product of the branched

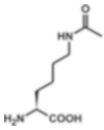
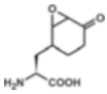
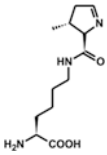
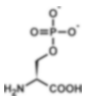
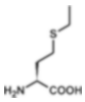
[1] NcAAs 1-24 are listed in Table 4.1 and ncAAs 25-106 can be found in Table 4.2.

Table 4.1 Selected naturally occurring noncanonical amino acids

	Name (synonym)	Structure	Analog of cAA	Natural source
1	(2*S*)-2-Aminononanoic acid		–	S-500 antibiotics (Hunt 1985)
2	(2*S*)-2-Aminobutanoic acid		Ala	*E. coli* [for biosynthesis, see Fotheringham et al. (1999)]
3	4-Methyl-L-proline		Pro	Cyanobacteria; *Nostoc* sp. GSV224 [for biosynthesis, see Luesch et al. (2002)]
4	N^6-Methyl-L-lysine		Lys	Present in human urine (Asatoor 1969) [for lysine methylation, see Kim and Paik (1965)]
5	5-Hydroxy-L-tryptophan		Trp	Trp metabolism; *Chromobacterium violaceum* [for biosynthesis, see Letendre et al. (1974)]
6	3,4-Dihydroxy-L-phenylalanine		Tyr/ Phe	Tyr metabolism [for biosynthesis, see Nagatsu et al. (1964)]
7	*trans*-4-Carboxy-L-proline		Pro	*Afzelia bella* seeds (Welter et al. 1978)
8	*cis*-3-Amino-L-proline		Pro	Fruit bodies of *Morchella esculenta* and other morel mushrooms (Hatanaka 1969)
9	L-2,4-Diaminobutyric acid		–	*Pseudomonas aeruginosa* PAO1 [for biosynthesis, see Vandenende et al. (2004)]
10	2-(Methylenecyclopropyl)-L-glycine		–	Unripe fruits of the Jamaican ackee tree (*Blighia sapida*, Sapindaceae) [for biosynthesis, see Kean and Lewis (1981)]
11	L-Canavanine		Arg	Leguminosae [for biosynthesis, see Rosenthal (1982)]
12	Beta-cyano-L-alanine		–	*Vicia* sp. (Ressler and Malodeczky 1962)

(continued)

Table 4.1 (continued)

	Name (synonym)	Structure	Analog of cAA	Natural source
13	L-Selenocysteine		Cys	Different species Natural expansion of the genetic code, SCS (Ambrogelly et al. 2007) [for biosynthesis, see Ambrogelly et al. (2007)]
14	(2*S*)-2-Amino-4-oxobutanoic acid		Asp	Key intermediate in amino acid biosynthesis (Hunt 1985)
15	(2*S*)-2-Amino-4-oxopentanoic acid (L-4-oxonorvaline)		–	*Clostridium sticklandii* (Hunt 1985)
16	N^6-Acetyl-L-lysine		Lys	*Beta vulgaris* (Fowden 1972)
17	3,4-Dichloro-L-proline		Pro	*Penicillium islandicum* (Yoshioka et al. 1973)
18	(2*S*)-2-Amino-4,5-hexadienoic acid		–	*Amanita solitaria* (Chilton and Tsou 1972)
19	L-Propargylglycine		–	Undefined streptomycete (Scannell et al. 1971)
20	2,3-Epoxy-4-oxohexahydro-L-phenylalanine		Phe	*Streptomyces griseoplanus* (Shah et al. 1970)
21	L-Piperazic acid		Pro	*Kutzneria* spp. [for biosynthesis, see Neumann et al. (2012)]
22	L-Pyrrolysine		–	*Methanosarcina* spp. *Desulfitobacterium hafniense* Natural expansion of the genetic code; SCS (Krzycki 2005) [for biosynthesis, see Ambrogelly et al. (2007)]
23	*O*-Phospho-L-serine		Ser/Cys	Cys precursor in some methanogenic archaea charged onto $tRNA^{Cys}$ by a natural SepRS (Sauerwald et al. 2005)
24	L-Ethionine		Met	Metabolic by-product found in *Escherichia coli*, *Bacillus megaterium*, *Pseudomonas aeruginosa*, and *Aerobacter aerogenes* (Fisher and Mallette 1961)

chain amino acid pathway. It accumulates when the enzymes of the leucine biosynthetic pathway are derepressed and the flux through the pathway is increased (Bogosian et al. 1989).

4.1.2.4 The Biological Functions of Natural ncAAs

The biological functions of natural (free) ncAAs are largely unknown. NcAAs primarily constitute building blocks for small bioactive peptide scaffolds (Walsh et al. 2013).

Many ncAAs are toxic. They act as antimetabolites (Fowden et al. 1967), particularly those with structural similarities to one of the 20 cAAs. Since the mechanism is mostly due to competitive inhibition of an enzyme acting on the respective cAAs, the toxic effects can be reversed by the corresponding cAA.

NcAAs arising from ancillary metabolic routes, such as L-norleucine (**26**, see Sect. 4.1.2.3), most probably do not have a function but they can interfere with cellular processes. NcAAs may act as substrates of cellular enzymes either inactivating them (Jung 1985) or leading to toxic products. For instance, ethionine (**24**), an S-ethyl analog of methionine, is a substrate for methionine adenosyltransferase (MAT, EC 2.5.1.6) in eukaryotes (Alix 1982). It can actively compete with methionine for the enzyme, which leads to the formation of *S*-adenosyl-L-ethionine (SAE) in addition to *S*-adenosyl-L-methionine (SAM). While SAM acts as an important methyl donor in the cell, SAE is extremely toxic: it causes the ethylation of cellular macromolecules, such as DNA, RNA, and nuclear proteins. Its interference with many metabolic functions of SAM and other biosynthetic and regulatory processes gives rise to numerous malfunctions (Alix 1982).

Except for selenocysteine (**13**) and pyrrolysine (**22**) (see Sects. 4.2.4.1 and 4.2.4.2), so far there is no evidence for the functional ribosomal translation of the ncAAs into proteins of their producing organism. However, another important mode of action of the toxic ncAAs is their eventual incorporation into proteins where they specifically and stoichiometrically replace the corresponding cAA. For instance, under conditions that lead to the production of high levels of leucine, the by-product norleucine (**26**) can accumulate and infiltrate translation. This accidental incorporation of norleucine in proteins is prevented by exogenous methionine (Bogosian et al. 1989).

The accidental incorporation of ncAAs can lead to variant proteins with impaired function. For instance, if an ncAA replaces a cAA that is essential for catalytic activity, enzyme function may be greatly impaired or entirely lost. Replacement outside the active center can lead to structure perturbations potentially affecting enzyme function. Notably, the accidental replacement of a cAA by a structurally analogous ncAA occurs randomly in all newly synthesized proteins of a cell, and any position has an equal chance of replacement. The aberrant proteins eventually accumulate, which results in growth arrest and finally cell death (Fowden et al. 1967).

Nature has taken some very effective measures to prevent ncAAs from wreaking havoc with protein translation, particularly in the organisms capable of their biosynthesis. NcAAs, however structurally related, are generally much less

efficiently charged onto tRNAs than the corresponding cAAs. Their aminoacylation efficiency is as low as that of near-cognate amino acids (O'Donoghue et al. 2013). In some cases, this principle is carried to the extreme: Leguminous plants produce the arginine analog L-canavanine (**11**), and the ingestion of canavanine-rich food is known to provoke various pathogenic symptoms. Canavanine is a very toxic ncAA because it is rapidly channeled into ribosomal translation in the absence or deficiency of arginine and the accumulation of canavanine-containing polypeptides eventually leads to cell death ("canavanine death"; see Rodgers and Shiozawa (2008) and references therein). The bruchid beetle *Caryedes brasiliensis* (Thunberg 1816) preys on the seeds of the legume vine, *Dioclea megacarpa*, which contain substantial amounts of the toxic ncAA. The beetle has developed an advanced strategy to cope with the high amounts of L-canavanine in its food: Its advanced arginyl-tRNA synthetase, that is, the enzyme that charges arginine onto its cognate tRNA (see Sect. 4.2.3), clearly discriminates between arginine and its close analog canavanine (Rosenthal et al. 1976).

Species that do not produce ncAAs have evolved an additional layer of defense to protect themselves from their toxic effects. Although most ncAAs enter the cell by the same amino acid transport systems as the cAAs, they often have a lower affinity for the transporter (Rodgers and Shiozawa 2008).

4.2 Assembling the Building Blocks into (Poly)peptides

In their free form, amino acids play important roles in cellular processes: They act as intermediates in the biosynthesis of a diverse array of small molecules, and their metabolism leads to the formation of important biomolecules such as cellular messengers and hormones (Barrett and Elmore 1998).

Most importantly, amino acids constitute monomers that can be linked by peptide bonds between their amino and carboxyl groups to form macromolecules, the (poly)peptides. The amino acid sequence in the (poly)peptides is either determined by the genetic information contained on messenger RNA (ribosomal translation) or by the order of the modules in a specialized non-ribosomal peptide synthetase (NRPS; non-ribosomal peptide synthesis).

The following two sections briefly recap the hallmarks of the two major routes for the production of (poly)peptides in nature: ribosomal translation and non-ribosomal peptide synthesis.

4.2.1 Non-ribosomal Peptide Synthesis

Non-ribosomal peptides (NRPs) are natural products of stunning chemical diversity. Their biological activities are extremely versatile: antibiotic, immunosuppressive, cytostatic, or toxic NRPs have been described. Some of these are in clinical use, such as the immunosuppressant drug cyclosporine, which is applied after organ transplantation to prevent transplant rejection (Caboche et al. 2010).

In comparison to polypeptides produced by ribosomal translation, NRPs are rather short. With rare exemptions, they mostly comprise between two and 20 monomers (Caboche et al. 2010). While ribosomal translation produces exclusively linear polypeptides, NRPs are often cyclic or branched. Ribosomal polypeptides contain only the 20 cAAs, whereas non-ribosomal peptide synthesis draws its chemical diversity from the incorporation of cAAs, ncAAs, as well as D- and beta-amino acids. Moreover, many amino acids carry modifications such as *N*-methyl and *N*-formyl groups or fatty acids; they may as well be glycosylated or acylated (Walsh et al. 2013; Finking and Marahiel 2004).

NRPs are biosynthesized by non-ribosomal peptide synthetases (NRPSs). These large multienzyme complexes are organized in modules (see Fig. 4.2a): Each module adds one predefined amino acid to the growing peptide chain. The number and the order of the modules present in an NRPS determine the length and the structure of the synthesized NRP. This is the reason why NRPSs are not dependent on messenger RNA as the template for peptide synthesis (see Sect. 4.2.3). However, a certain NRPS can synthesize only one type of NRP (Walsh et al. 2013; Finking and Marahiel 2004).

Recent findings somewhat blur the strict distinction between peptide structures derived by ribosomal translation and non-ribosomal peptide synthesis. Generally, D- and beta-amino acids are considered typical building blocks for NRPs. However, peptides synthesized by ribosomal translation can contain them as well by post-translational maturation (Walsh 2014).

4.2.2 The Genetic Code

The genetic information of living organisms is stored in nucleic acids. While different types of RNA, such as ribosomal and transfer RNAs, contribute to implementing it, the major part of the genetic information stored in nucleic acids is realized in the form of proteins. This presents living creatures with an information translation problem: Their DNA uses four nucleobases, adenine (A), cytosine (C), guanine (G), and thymine (T), for information storage, and the alphabet in RNA is the same except that uracil (U) substitutes for T. However, their proteins use 20 canonical amino acids and no nucleobases to execute the genetic instructions. Consequently, there must be rules that govern the translation of the genetic information from the genetic alphabet to the amino acid alphabet of the proteins.

The genetic code is exactly this cipher: It constitutes a set of rules that guarantee the correct translation of information between the two realms. Base triplets are employed as the basic information unit rather than doublets or single bases. All possible triplet combinations of the four nucleobases result in 64 independent coding units, or codons, that can be assigned to the canonical amino acids (Fig. 4.1). It is evident that there are more than threefold more coding units than amino acids; nevertheless, all coding units are assigned in the standard genetic code as depicted in Fig. 4.1. This can only be achieved if more than one codon is assigned per amino acid. Indeed, the genetic code is degenerate: Methionine and tryptophan

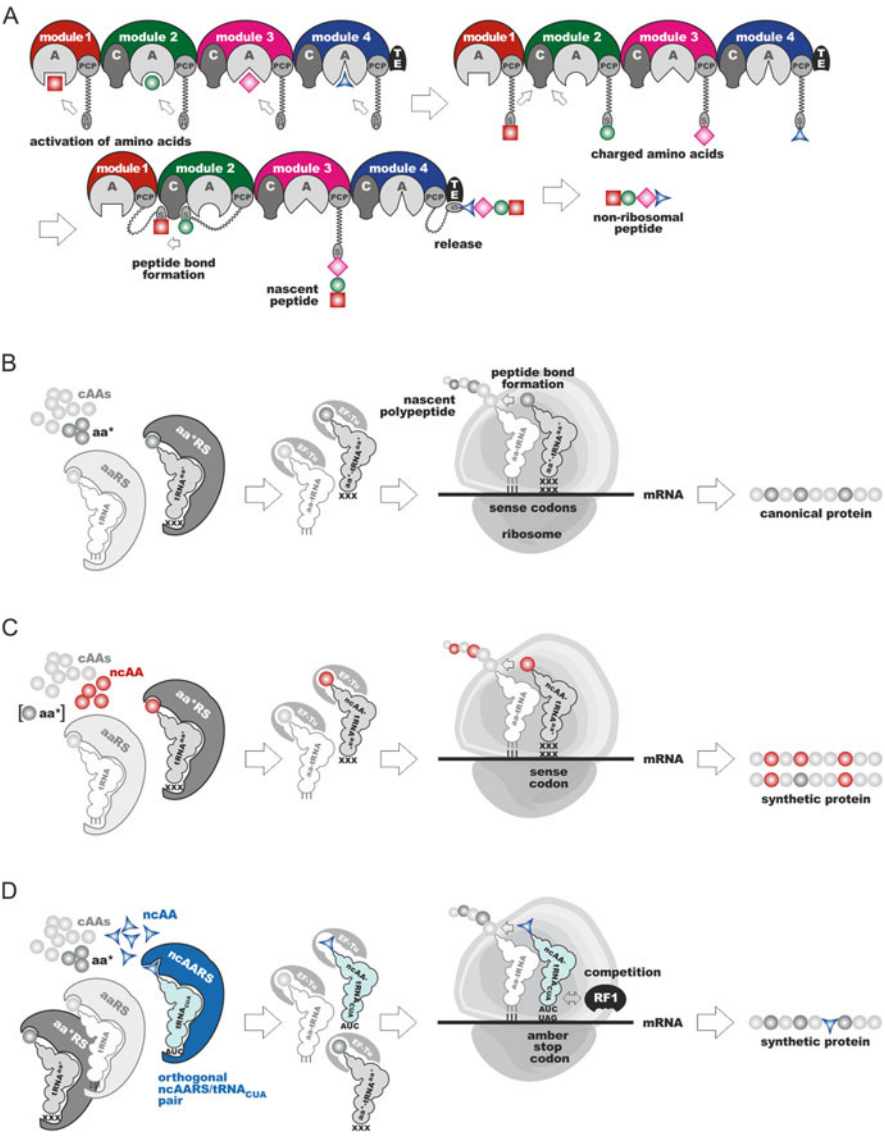


Fig. 4.2 Canonical and noncanonical amino acids as building blocks for (poly)peptides. (**a**) Non-ribosomal peptide synthesis: The amino acids are activated at the adenylation site A and charged onto the free thiol group (SH) of the flexible phosphopantetheine arm (*zigzag line*) of the peptidyl carrier protein (PCP). Condensation of adjacent residues occurs at the condensation domain C. Each module contains at least an A, a C, and a PCP domain; starter modules (module 1) lack a C domain and the last module (module 4) contains a thioesterase (TE) domain for the release of the finished non-ribosomal peptide. (**b**) Ribosomal translation: CAAs are charged onto their cognate tRNAs by the amino acid-specific aminoacyl-tRNA synthetases (aaRS); the charged aminoacyl-tRNAs (aa-tRNA) are delivered to the ribosome by the elongation factor EF-Tu; at the ribosome, decoding of the codons with the appropriate cAAs and peptide bond formation occurs; finally, a protein containing only cAAs is released. aa*, a specific cAA. (**c**) Residue-specific incorporation of ncAAs: An auxotrophic host is depleted for a specific cAA [aa*] whose

are the only amino acids encoded by only one codon, AUG and UGG, respectively. All other amino acids are assigned between two and six codons (Fig. 4.1). Three coding units, the ochre (UAA), amber (UAG), and opal (UGA) stop codons, designate the message "translation termination." With a few exemptions (see Sect. 4.2.4), the genetic code is largely the same in all extant life forms; that is why it is designated "universal" or "standard."

The assignment of the triplet codons to the amino acids in the standard genetic code is not random (Crick 1968; Lenstra 2014). Codons in which the third base is U or C always code for the same amino acid. This holds true for codons ending with A and G as well, except in two cases: The rare amino acids methionine (encoded by AUG) and tryptophan (encoded by UGG) appear to be exceptions to this rule as they are encoded by a single codon each. The codons can be arranged in 16 family boxes that are sets of four codons each that differ only in the third position. Eight out of the 16 family boxes code for a single amino acid (Ala, Arg, Gly, Leu, Pro, Ser, Thr, Val; Fig. 4.1). Codons representing a single amino acid start with the same pair of bases, e.g., the triplets coding for His start with CA and so do they that encode Gln; AU is the initiator pair of codons for Ile and Met; Ala codons start with GC; and so on (refer to Fig. 4.1). However, in the case of Leu, Ser, and Arg, this is different: These amino acids are encoded by six codons each, that is, a family box plus two additional codons. All codons in the family box have the same initial pair of bases, but the two additional codons start with another doublet: Leu, CU (family box) vs. UU (additional codons); Ser, UC vs. AG; and Arg, CG vs. AG (Fig. 4.1). All codons with C in the second position encode the same amino acid if they share the same initial doublet (UC, Ser; CC, Pro; AC, Thr; GC, Ala). This is also true for codons with G in the second position but only if the first base is C (CG, Arg) or G (GG, Gly). All codons with U at the second position encode hydrophobic amino acids (Fig. 4.1).

Mathematical coding models allow to tackle the question why just one unique genetic code evolved (Lenstra 2014). However, it is unknown why only 20 cAAs were assigned to 64 codons and why they are those that we see today but not others (O'Donoghue et al. 2013). The origin of the genetic code is still enigmatic and highly controversial (Di Giulio 2005). The early frozen accident theory of Crick postulates "that the code is universal because at the present time any change would

Fig. 4.2 (continued) analog is to be incorporated into a target protein. The structurally similar ncAA supplemented in the medium accumulates in the cells and can be charged onto the $tRNA^{aa*}$ by the aa*-specific aaRS. The misacylated ncAA-$tRNA^{aa*}$ participates in translation and a synthetic protein containing the ncAA at the positions of aa* is produced. Trace amounts of the aa* lead to inhomogeneous protein preparations. (**d**) Site-specific incorporation of ncAAs at amber stop (UAG) codons: The host strain is equipped with an orthogonal ncAARS/$tRNA_{CUA}$ pair and supplemented with the ncAA. The orthogonal pair acts independently of the cellular aaRSs and tRNAs. The ncAA-specific ncAARS charges the ncAA onto the suppressor $tRNA_{CUA}$ which incorporates it in response to an amber stop codon. At the ribosome, the ncAA-$tRNA_{CUA}$ competes with release factor RF1 for the stop codon. A synthetic protein with the ncAA inserted at the position specified by the amber stop codon is produced. See text for further details.

be lethal, or at least very strongly selected against" (Crick 1968). For a comprehensive review of the theories on the origin of the universal genetic code, see Di Giulio (2005).

4.2.3 Ribosomal Translation

Ribosomal translation (Fig. 4.2b) is the cellular process that assembles polypeptides following the instructions encoded on a messenger RNA (mRNA) template. The mRNA is a transcript of the relevant genetic information stored on the (genomic) DNA. Besides the mRNA, which acts as an information carrier, a number of other major players are involved in ribosomal translation: the transfer RNAs (tRNAs), the aminoacyl-tRNA synthetases (aaRSs), the ribosomes, and various protein factors such as initiation and elongation factors as well as release factors.

The tRNAs are the adaptor molecules that physically link the nucleotide sequence of the mRNA with the amino acid sequence of the polypeptide. One end of the tRNA molecule forms base pairs with the codons on the mRNA via a three-base sequence, i.e., the anticodon. The opposite end of the tRNA molecule carries the amino acid. Thus, the tRNA "reads" the codons on the mRNA and decodes them (Ibba and Söll 2000). Isoacceptor tRNAs decode all codons specifying the same amino acid (Fig. 4.1) and normally there are no cognate tRNAs for the three stop codons. Instead, the stop codons are specifically recognized by the release factors. The release factors are proteins that mimic tRNAs and provoke the termination of ribosomal translation (Moore and Steitz 2011).

The genetic code is only correctly interpreted if a cAA is bound to a tRNA that "reads" and decodes a corresponding codon for this amino acid. Consequently, the correct aminoacylation, or "charging," of the tRNAs with their cognate cAAs is of paramount importance for translation fidelity. The aminoacyl-tRNA synthetases catalyze the aminoacylation reaction and their accuracy prevents translational errors. Most organisms contain 20 aaRSs (Moghal et al. 2014), one for each cAA, and all catalyze the same reaction: first, a specific aaRS selects the cAA and activates it by the formation of an aminoacyl adenylate. This activation step consumes ATP. In the next catalytic step, the aaRS transfers the aminoacyl moiety to the bound tRNA (Ibba and Söll 2000).

It is important to note that the way aaRSs recognize their cognate tRNAs may differ in bacteria, archaea, and eukarya. The aaRSs recognize specific bases, the identity determinants, which can vary among the three kingdoms of life. Consequently, a bacterial aaRS does not necessarily charge an archaeal tRNA and vice versa (Giege et al. 1998). Actually, this incompatibility of the translational components is exploited for the generation of so-called orthogonal aaRS/tRNA pairs for site-specific incorporation of ncAAs (see Sect. 4.2.8).

There are several levels of quality control of the aminoacylation (Ling et al. 2009): Each synthetase charges only its cognate isoacceptor tRNAs with the appropriate cAA while non-cognate tRNAs and cAAs with the inappropriate size,

structure, charge, or hydrophobicity will be efficiently discriminated. Thus, at the first editing level, false substrates are prevented from entering the respective binding pockets of the aaRSs. If a near-cognate cAA is not excluded from the amino acid-binding pocket and is erroneously activated, some aaRSs [IleRS, PheRS, ValRS, LeuRS, MetRS, ThrRS, AlaRS, ProRS, LysRS, SerRS (Ling et al. 2009)] possess the ability to hydrolyze the misactivated intermediate before it is transferred to the tRNA. This "pre-transfer editing" can be tRNA dependent. At yet another editing level, the hydrolysis of misaminoacylated tRNAs ("post-transfer editing") may occur (Moghal et al. 2014; Ling et al. 2009).

After aminoacylation, an elongation factor binds to the aminoacyl-tRNA and delivers it to the ribosome. Evidently, this process provides another layer of quality control complementary to the aaRS because the elongation factor binds misaminoacylated tRNAs only weakly (Ling et al. 2009).

At the ribosome, the aminoacyl-tRNA base-pairs with the correct codon on the mRNA. Correct anticodon/codon pairing is a prerequisite for the transfer of the aminoacyl moiety to the nascent polypeptide chain and for peptide bond formation. The editing of non-cognate anticodon/codon interactions provides another layer of translational quality control. The ribosome that catalyzes the formation of the peptide bond is a ribonucleoprotein particle that primarily consists of rRNA. Its catalytic function actually relies primarily on the rRNA; that is why the ribosome is a ribozyme (Moore and Steitz 2011).

However, the ribosome does not rigorously edit tRNAs mischarged with near-cognate amino acids or ncAAs. If a mischarged tRNA escapes the prior layers of editing, it is very likely to be translated.

4.2.4 Natural Exceptions of the Standard Genetic Code

The genetic code was initially regarded as universal and frozen (see Sect. 4.2.2), which implies that it is the same in all extant life forms and it cannot be changed. However, naturally occurring deviations from the standard genetic code as well as its expansion with ncAAs have provided alternative perspectives.

Approximately 20 variations of the standard genetic code as depicted in Fig. 4.1 are known to date (Ambrogelly et al. 2007). The code is far more flexible than initially anticipated, and even dynamic modulations in response to environmental changes were recently reported (Prat et al. 2012). The deviations occur mainly in microorganisms and mitochondria and are confined to codon reassignments of the cAAs—yet with two important exemptions (see Sects. 4.2.4.1 and 4.2.4.2).

For instance, the codon assignment in vertebrate mitochondria is different from the standard code: In the organelles, AUA (designates Ile in the standard code, Fig. 4.1) codes for methionine and UGA (opal stop in the standard code, Fig. 4.1) for tryptophan (Ambrogelly et al. 2007; Knight et al. 2001). The AUA codon experienced the same reassignment in the mitochondria of yeasts, and *Mycoplasma* spp. also use UGA to encode Trp. Variant codes were found in yeasts, ciliates, mitochondria of diatoms, algae, and metazoa [for a comprehensive overview, see

Knight et al. (2001)]. The ochre (UAA) and amber (UAG) stop codons changed their meaning from translation stop to Gln in some diplomonads, ciliates, and in the green alga *Acetabularia acetabulum*. Some codons were frequently reassigned, even though to different alternative meanings. For instance, UAG was reassigned from stop to Leu, Ala, and Gln, and AGA and AGG changed their message from Arg to Ser, Gly, and stop (Knight et al. 2001). Interestingly, the same changes occurred independently in organisms from different lineages (Ambrogelly et al. 2007).

The described variations change the genetic code in as much as they change the codon assignments to the different cAAs. There are two ncAAs, selenocysteine (Sec, **13**) and pyrrolysine (Pyl, **22**), that are introduced into the genetic code under strictly defined conditions. For instance, the marine ciliate *Euplotes crassus* uses the codons UGA, UGU, and UGC to encode cysteine. However, if a certain genetic element, a so-called SECIS element (see Sect. 4.2.4.1), is present, the UGA can also encode Sec in selenoproteins (Turanov et al. 2009).

4.2.4.1 Selenocysteine (Sec)

Sec (**13**) is the selenium analog of Cys, i.e., in Sec, the sulfur is replaced by a selenium to form a selenol moiety. The selenolate anion is more stable than the thiolate anion because the pK_a of selenol (5.2) is nearly three units lower than that of thiol (8.5). Moreover, selenol has a lower redox potential (–488 mV) than thiol (–233 mV). Consequently, selenol is fully deprotonized at physiological pH, and Sec is substantially more active than Cys. These features of the selenol group are the basis for the unique catalytic efficiencies of selenoproteins (Ambrogelly et al. 2007).

The selenoproteins are found in organisms from all three kingdoms of life (Labunskyy et al. 2014) (see the selenoprotein database SelenoDB, www.selenodb.org). Sec occurs commensurably frequently in mammalian proteins and 25 genes encoding selenoproteins have been identified in humans (Reeves and Hoffmann 2009). They are involved in redox reactions to protect cells against oxidative damage, e.g., glutathione peroxidase; they catalyze thyroid hormone activation, e.g., iodothyronine deiodinases; and they play a role in protein folding quality control. Sec is absolutely essential for human health and mutations in the selenoproteins are linked to various disorders (Reeves and Hoffmann 2009; Schmidt and Simonovic 2012).

Selenoproteins are relatively scarce in bacteria; nevertheless, the microbial selenoproteome of the Sargasso Sea apparently includes quite a number of prokaryotic proteins with Sec (Zhang et al. 2005a). With a few exceptions such as the algae *Chlamydomonas* spp. and *Emiliania huxleyi*, Sec is absent from higher plants and fungi (Ambrogelly et al. 2007; Labunskyy et al. 2014).

Many selenoproteins possess only a single Sec, though there are examples with more, e.g., the human Sel P contains 10 Sec residues (Burk and Hill 2005).

Sec is encoded by an opal (UGA) stop codon. As outlined in Sect. 4.2.2, all 64 codons are occupied by cAAs in the standard genetic code. Therefore, the message of the UGA codon must be transiently changed from translation

termination to Sec for the expression of selenoproteins. This transient switch of codon meaning calls for a number of protagonists: In the first place, a tRNA that can read the UGA stop codon is indispensable. As already mentioned, normally there are no tRNAs for the stop codons; instead, release factors terminate translation. A tRNA carrying an anticodon for a stop codon is called a suppressor tRNA (as it suppresses the meaning of the stop codon); for Sec incorporation, this is the suppressor $tRNA_{UCA}^{[Ser]Sec}$ (SelC in *E. coli*) (Commans and Böck 1999). Furthermore, a specialized translation elongation factor SelB and a structural element in the mRNA transcript are required. The stem–loop structure in the mRNA triggers the recoding of the UGA codon and is designated the selenocysteine insertion sequence (SECIS) element (Labunskyy et al. 2014). Evidently, the mRNA context determines an alternative interpretation of the genetic code. Note that Sec does not have its own aaRS.

In prokaryotes such as *E. coli*, the incorporation of Sec in a selenoprotein starts with the aminoacylation of $tRNA_{UCA}^{[Ser]Sec}$ with serine by the seryl-tRNA synthetase (SerRS). The Ser-$tRNA_{UCA}^{[Ser]Sec}$ is converted into Sec-$tRNA_{UCA}^{[Ser]Sec}$ by the action of the selenocysteine synthase (SelA) and the selenophosphate synthetase (SelD). Selenide (Se^{2-}) and ATP are used as cofactors in the reaction, and the transformation of the Ser molecule takes place while it is bound to $tRNA_{UCA}^{[Ser]Sec}$ (Böck et al. 1991).

Sec is not incorporated at any opal stop codon but only at specified in-frame UGA codons in genes encoding selenoproteins. The SECIS element flags the relevant in-frame UGA codon(s) as Sec incorporation message rather than translation termination signal. In bacteria, the SECIS elements are located inside the coding region of the selenoproteins, immediately downstream of the recoded UGA codon. SelB binds specifically to Sec-$tRNA_{UCA}^{[Ser]Sec}$ as well as to the SECIS element. Binding to the SECIS element induces a conformational change that enables SelB to functionally interact with the ribosome. In this way, it delivers the charged suppressor tRNA to the ribosome where it outcompetes the release factor for decoding of UGA (Labunskyy et al. 2014). SelB and the general elongation factor of *E. coli*, EF-Tu, share homologous sequences at their N-termini. In spite of that, EF-Tu does not recognize Sec-$tRNA_{UCA}^{[Ser]Sec}$ neither does SelB interact with any other aminoacyl-tRNA. Moreover, it efficiently discriminates against Ser-$tRNA_{UCA}^{[Ser]Sec}$ (Böck et al. 1991).

In eukarya and archaea, Sec is incorporated into selenoproteins in a similar process yet with a few subtle differences (Labunskyy et al. 2014; Cobucci-Ponzano et al. 2012). As in prokaryotes, a SecRS is not found in eukaryotes and archaea. The biosynthesis of Sec-$tRNA_{UCA}^{[Ser]Sec}$ in these lineages involves a tRNA-dependent modification of serine as well, yet the pathway includes an additional catalytic step. Again, SerRS charges $tRNA_{UCA}^{[Ser]Sec}$ with serine in the first step but then a phosphoseryl-$tRNA_{UCA}^{[Ser]Sec}$ kinase (PSTK) phosphorylates the Ser bound to the suppressor tRNA to form an *O*-phosphoseryl-$tRNA_{UCA}^{[Ser]Sec}$ intermediate (Sep-$tRNA_{UCA}^{[Ser]Sec}$). Sep-$tRNA_{UCA}^{[Ser]Sec}$ finally serves as a substrate for the eukaryotic and archaeal Sec synthase (Labunskyy et al. 2014; Cobucci-Ponzano et al. 2012). Similar to bacteria, eukarya and archaea use a *cis*-acting SECIS

element for Sec incorporation. However, the stem–loop is not located within the coding region but rather in the 3′-untranslated region of the mRNA.

In eukaryotes, two proteins are necessary to mediate the incorporation of Sec at the UGA codon: The SECIS-binding protein 2 (SBP2) binds the SECIS element and is stably associated with the ribosome. SBP2 also interacts with the Sec-specific translation elongation factor eEFSec, which recruits Sec-$tRNA_{UCA}^{[Ser]Sec}$. eEFSec is highly specific for Sec-$tRNA_{UCA}^{[Ser]Sec}$ and does neither interact with Sep-$tRNA_{UCA}^{[Ser]Sec}$ nor with Ser-$tRNA_{UCA}^{[Ser]Sec}$ or any other aminoacylated tRNA (Labunskyy et al. 2014). In archaea, SECIS-binding factors have not yet been identified (Cobucci-Ponzano et al. 2012).

Proteins that do not natively contain Sec can be equipped with this natural ncAA by a recombinant approach. Hilvert and co-workers used this strategy to probe the interaction of a catalytically important Cys residue of a cytochrome P450 with the heme cofactor (Aldag et al. 2009). To achieve this, they engineered the P450 gene to mimic that of a selenoprotein: they changed the relevant Cys codon to UGA and engineered a simplified yet functional SECIS element into the coding region (Aldag et al. 2009). However, in bacterial expression hosts such as *E. coli*, this approach is severely hampered by the sequence constraint imposed on the target gene by the SECIS element.

To overcome the difficulties of recombinant Sec (**100**) incorporation in *E. coli*, the group of Söll devised an alternative approach. It is independent of the SECIS element and facilitates site-specific insertion of Sec at a desired position in the target protein. Central to this strategy is a synthetic suppressor tRNA that is a substrate for SerRS and SelA/SelD as well as EF-Tu. The synthetic suppressor $tRNA^{UTu}$ is charged with Ser by SerRS and SelA/SelD converts Ser to Sec as described previously. EF-Tu binds to the Sec-$tRNA^{UTu}$ generated in this way and delivers it to the ribosome as any other cellular aminoacyl-tRNA. The design of two different suppressors, $tRNA^{UTu}_{am}$ and $tRNA^{UTu}_{op}$, facilitates the incorporation of Sec at amber UAG or opal UGA codons (Aldag et al. 2013).

4.2.4.2 Pyrrolysine (Pyl)

Pyrrolysine (Pyl, **22**) is the second natural ncAA that is included in the genetic code of selected organisms. However, the incorporation mechanism of Pyl is entirely different to that of Sec.

Pyl is found mainly in proteins from the archaeal *Methanosarcina* species, in the gram-positive bacterium *Desulfitobacterium hafniense*, in the human intestinal bacterium *Bilophila wadsworthia*, as well as in selected species of *Clostridium* and *Deltaproteobacteria* (Gaston et al. 2011).

Pyl is incorporated into polypeptides in response to an in-frame amber UAG codon. The most important difference to Sec is that a pyrrolysyl-tRNA synthetase (PylRS, encoded by *pylS*) specifically charges Pyl onto its cognate $tRNA_{CUA}^{Pyl}$ (encoded by *pylT*) (Krzycki 2005). Thus, the decoding and translation mechanism of Pyl resembles that of the cAAs. The Pyl biosynthesis cluster *pylBCD* was identified and it was shown that Pyl is biosynthesized from two molecules of lysine.

Also, heterologous Pyl biosynthesis and its incorporation at an amber codon were demonstrated in recombinant *E. coli* (Gaston et al. 2011).

Initially, a PYLIS element similar to the SECIS element was hypothesized as the trigger for Pyl incorporation. However, amber codons can be recombinantly suppressed in *E. coli* and other expression hosts using the PylRS/$tRNA_{CUA}^{Pyl}$ pair (see Sect. 4.2.8) and no PYLIS element is required. How the natural hosts of the PylRS/$tRNA_{CUA}^{Pyl}$ pair regulate the incorporation of Pyl is still largely unknown (Zhang et al. 2005b).

4.2.4.3 Phosphoserine (Sep)

Some methanogenic archaea lack cysteinyl-tRNA synthetase (CysRS) (Sauerwald et al. 2005); nevertheless, Cys is found in their proteins. Consequently, an alternative pathway for the formation of Cys-$tRNA^{Cys}$ must exist. In fact, a similar tRNA-dependent process as for the biosynthesis of Sec-$tRNA_{UCA}^{[Ser]Sec}$ is active in these organisms.

First, a specific *O*-phosphoseryl-tRNA synthetase (SepRS) aminoacylates $tRNA^{Cys}$ with *O*-phospho-L-serine (Sep, **23**). Then, Sep-tRNA:Cys-tRNA synthase (SepCysS) converts Sep to Cys while still bound to the $tRNA^{Cys}$ (Sauerwald et al. 2005). SepRS, like PylRS, is a new aaRS that evolved to activate and charge a natural ncAA onto a cognate tRNA. In this case, however, Sep is not incorporated into proteins.

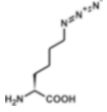
4.2.5 Changing the Genetic Code

Changing the genetic code in a deliberate way is not an easy task. In fact, it represents a highly challenging scientific problem: "...The problem with attempting to change protein biosynthesis is that nature is not especially forgiving. So, when manipulating protein biosynthesis in vivo or in vitro, investigators run up against a system that has been exposed (for billions of years) to all sorts of selections and perturbations" (Tamura and Schimmel 2002). Changing the genetic code is like a struggle with an extremely disciplined and resourceful opponent.

Changing the genetic code involves freeing one of the 64 triplet codons to obtain a "blank" codon and reassigning it—a truly herculean task. Nevertheless, genetic code engineering and the expansion of the genetic code are emerging technologies that are ever more adopted by researchers from outside the laboratories that have been developing the methods. The field has made tremendous progress in the recent years. Table 4.2 lists a selection of ncAAs that have been used for genetic code manipulations.

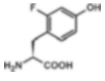
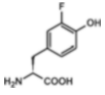
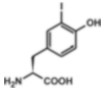
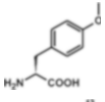
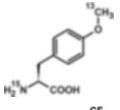
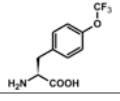
Changing the genetic code means adding new chemistries to life. Now this concept is at the heart of synthetic biology and a compelling promise to all kinds of disciplines. Prepare to be lured in!

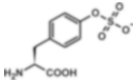
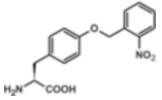
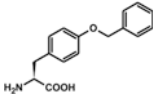
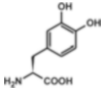
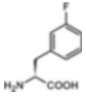
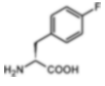
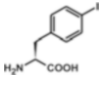
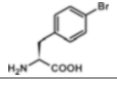
Table 4.2 Selected noncanonical amino acids for protein engineering

	Name	Structure	Analog of cAA	Incorporation method	Application	References
25	L-Selenomethionine		Met	SPI	Structure elucidation[a]	Hendrickson et al. (1990)
26	L-Norleucine		Met	SPI	Enzyme activity ↑	Cirino et al. (2003), Hoesl et al. (2011)
					Non-oxidizable Met analog	Wolschner et al. (2009)
27	L-Methoxinine		Met	SPI	Oxidized Met mimic	Wolschner et al. (2009)
28	L-Azidohomoalanine		Met	SPI	BOC with alkynes, BONCAT	Dieterich et al. (2006), Link and Tirrell (2003)
					Biosynthesis and residue-specific incorporation	Ma et al. (2014)
29	L-Azidonorleucine		Met	SPI[b]	BONCAT with alkynes	Ngo et al. (2009)
30	L-Homopropargylglycine		Met	SPI	BOC with azides	Teeuwen et al. (2009)
					BONCAT	Beatty et al. (2006)
31	(2*S*,4*S*)-4-Fluoroproline		Pro	SPI	Folding and stability	Steiner et al. (2008)
					Protective effect[c]	Acevedo-Rocha et al. (2013)
32	(2*S*,4*R*)-4-Fluoroproline		Pro	SPI	Folding and stability	Crespo and Rubini (2011), Edwardraja et al. (2011)
					Protective effect[c]	Acevedo-Rocha et al. (2013)

(continued)

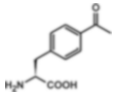
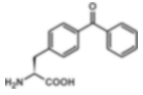
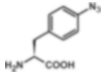
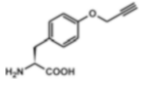
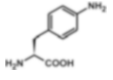
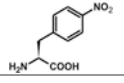
Table 4.2 (continued)

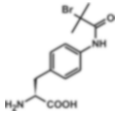
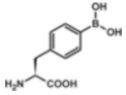
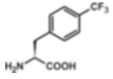
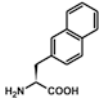
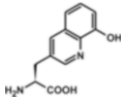
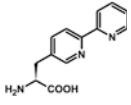
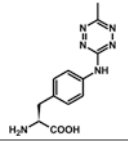
	Name	Structure	Analog of cAA	Incorporation method	Application	References
33	(2*S*,4*S*)-4-Hydroxyproline		Pro	SPI	Protective effect[c]	Acevedo-Rocha et al. (2013)
34	(2*S*,4*R*)-4-Hydroxyproline		Pro	SPI	Protective effect[c]	Acevedo-Rocha et al. (2013)
35	*ortho*-Fluoro-L-tyrosine		Tyr	SPI	Protective effect[c]	Acevedo-Rocha et al. (2013)
36	*meta*-Fluoro-L-tyrosine		Tyr	SPI	Stability Prolonged shelf life Protective effect[c]	Deepankumar et al. (2014) Budisa et al. (2010) Acevedo-Rocha et al. (2013)
37	*meta*-Iodo-L-tyrosine		Tyr	SCS	Structure elucidation[a]	Sakamoto et al. (2009)
38	*O*-Methyl-L-tyrosine		Tyr	SCS	Spectral properties	Wang et al. (2003b)
39	(2*S*)-2-(^{15}N)Amino-3-{4-[(^{13}C)methyloxy]phenyl} propanoic acid		Tyr	SCS	Isotope-labeled NMR probe	Jones et al. (2010)
40	*O*-Trifluoromethyl-L-tyrosine		Tyr	SCS	Polyfluorinated ^{19}F-NMR probe	Jones et al. (2010)

41	L-*O*-Sulfotyrosine		Tyr	SCS	Phage display[d]	Liu et al. (2008)
42	L-*O*-(2-Nitrobenzyl) tyrosine		Tyr	SCS	Photocaged Tyr	Jones et al. (2010)
43	L-*O*-Benzyltyrosine		Tyr	SCS	Enzyme activity ↑	Kolev et al. (2014)
44	3,4-Dihydroxy-L-phenylalanine		Tyr	SCS	Reactive amino acid	Deepankumar et al. (2015)
				SPI	Redox sensor	Alfonta et al. (2003)
45	*ortho*-Fluoro-L-phenylalanine		Phe	SPI	Phe "fingers"	Minks et al. (2000)
46	*meta*-Fluoro-L-phenylalanine		Phe	SPI	Phe "fingers"	Minks et al. (2000)
					Enzyme activity ↑	Hoesl et al. (2011)
					Protective effect[c]	Acevedo-Rocha et al. (2013)
47	*para*-Fluoro-L-phenylalanine		Phe	SPI	Phe "fingers"	Minks et al. (2000)
					Prolonged shelf life	Budisa et al. (2010)
					Protective effect[c]	Acevedo-Rocha et al. (2013)
48	*para*-Iodo-L-phenylalanine		Phe	SCS	Spectral properties	Wang et al. (2003b)
					Structure elucidation[a]	Kodama et al. (2010), Xie et al. (2004)
49	*para*-Bromo-L-phenylalanine		Phe	SCS	Spectral properties	Wang et al. (2003b)

(continued)

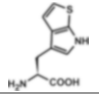
Table 4.2 (continued)

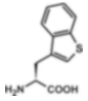
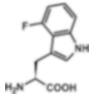
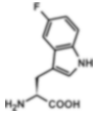
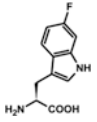
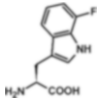
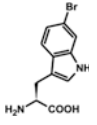
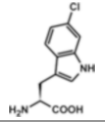
	Name	Structure	Analog of cAA	Incorporation method	Application	References
50	*para*-Acetyl-L-phenylalanine		Phe	SCS	Enzyme activity ↑	Kolev et al. (2014)
					Phage display[d]	Liu et al. (2008)
				SPI[e]	BOC with hydrazines and alkoxyamines	Wang et al. (2003a)
						Datta et al. (2002)
51	*para*-Benzoyl-L-phenylalanine		Phe	SCS	Photo-cross-linker	Chin et al. (2002a)
52	*para*-Azido-L-phenylalanine		Phe	SPI[f]	BOC with alkynes	Carrico et al. (2007)
				SCS	BOC with alkynes	Chin et al. (2002b, 2003a), Young et al. (2009), Liu et al. (2007), Deiters et al. (2003)
				SCS	Photo-cross-linker	Chin et al. (2002b)
53	*para*-Propargyloxy-L-phenylalanine		Phe	SCS	BOC with azides	Young et al. (2009), Liu et al. (2007), Deiters et al. (2003), Deiters and Schultz (2005)
54	*para*-Amino-L-phenylalanine		Phe	SCS	Spectral properties	Wang et al. (2003b)
					Enzyme activity ↑	Kolev et al. (2014), Jackson et al. (2006)
					BOC (oxidative coupling to *ortho*-aminophenols)	Behrens et al. (2011)
					Biosynthesis and site-specific incorporation	Mehl et al. (2003)
55	*para*-Nitro-L-phenylalanine		Phe	SCS	Enzyme activity ↑	Jackson et al. (2006)

56	L-4-(2′-Bromoisobutyramido) phenylalanine		Phe	SCS	BOC (radical polymerization)	Peeler et al. (2010)
57	*para*-Borono-L-phenylalanine		Phe	SCS	Phage display[d] BOC by transition metal-catalyzed cross-coupling	Liu et al. (2008) Brustad et al. (2008)
58	*para*-Trifluoromethyl-L-phenylalanine		Phe	SCS	Polyfluorinated ^{19}F-NMR probe	Jones et al. (2010)
59	L-3-(2-Naphthyl)alanine		Phe	SCS	Spectral properties Enzyme activity ↑ SCR	Wang et al. (2003b) Kolev et al. (2014) Kwon et al. (2003)
60	L-3-(8-Hydroxyquinolin-3-yl)alanine		Phe	SCS	Metal chelating, probe for structure–function studies	Jones et al. (2010)
61	L-(2,2′-Bipyridin-5-yl) alanine		Phe	SCS	Phage display[d] Metal chelating, probe for structure–function studies	Liu et al. (2008) Jones et al. (2010)
62	L-4-(6-Methyl-*s*-tetrazin-3-yl)aminophenylalanine		Phe	SCS	BOC with *trans*-cyclooctenes	Seitchik et al. (2012)

(continued)

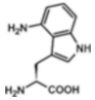
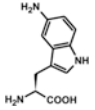
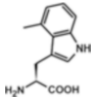
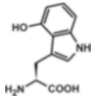
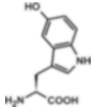
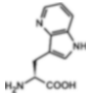
Table 4.2 (continued)

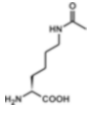
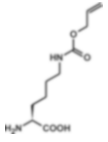
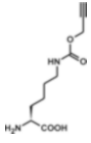
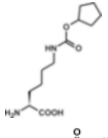
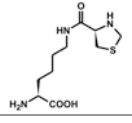
	Name	Structure	Analog of cAA	Incorporation method	Application	References
63	5′,5′,5′-Trifluoro-L-leucine		Leu	SPI	PFAA, thermo- and chemostability	Tang et al. (2001)
64	5,5,5,5′,5′,5′-Hexafluoro-L-leucine		Leu	SPI	PFAA, thermo- and chemostability	Tang and Tirrell (2001)
65	L-4-Oxonorvaline		Leu	SPI[g]	BOC with hydrazines and alkoxyamines	Tang et al. (2009)
66	L-3-(1,2,4-Triazol-3-yl) alanine		His	SPI	Enzyme activity ↑	Beiboer et al. (1996)
67	L-2-Methylhistidine		His	SPI	Enzyme activity ↓	Schlesinger and Schlesinger (1967, 1969)
68	Beta-(selenolo[3,2-*b*] pyrrol-3-yl)-L-alanine		Trp	SPI	Spectral properties Structure elucidation[a]	Budisa et al. (2004) Bae et al. (2001)
69	Beta-(selenolo[2,3-*b*] pyrrol-3-yl)-L-alanine		Trp	SPI	Spectral properties	Budisa et al. (2004)
70	Beta-(thieno[3,2-*b*]pyrrol-3-yl)-L-alanine		Trp	SPI	Spectral properties	Budisa et al. (2004)
71	Beta-(thieno[2,3-*b*]pyrrol-3-yl)-L-alanine		Trp	SPI	Spectral properties	Budisa et al. (2004)

72	L-3-(Benzothien-3-yl) alanine		Trp	SCS	Spectral properties	Kwon and Tirrell (2007)
73	4-Fluoro-L-tryptophan		Trp	SPI	Protective effect[c] Photochemistry Spectral properties	Acevedo-Rocha et al. (2013) Staudt et al. (2013) Budisa et al. (2004)
74	5-Fluoro-L-tryptophan		Trp	SPI	Spectral properties Prolonged shelf life Enzyme activity ↑	Budisa et al. (2004) Budisa et al. (2010) Parsons et al. (1998)
75	6-Fluoro-L-tryptophan		Trp	SPI	Spectral properties	Budisa et al. (2004)
76	7-Fluoro-L-tryptophan		Trp	SPI	Spectral properties	Budisa et al. (2004)
77	6-Bromo-L-tryptophan		Trp	SCS	Spectral properties	Kwon and Tirrell (2007)
78	6-Chloro-L-tryptophan		Trp	SCS	Spectral properties	Kwon and Tirrell (2007)

(continued)

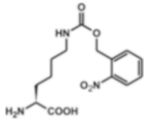
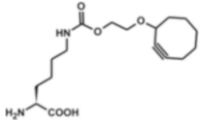
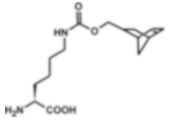
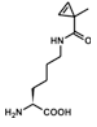
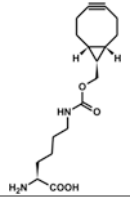
Table 4.2 (continued)

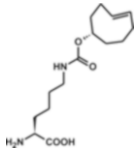
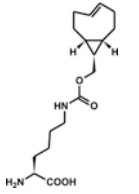
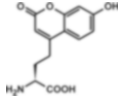
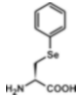
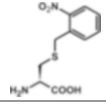
	Name	Structure	Analog of cAA	Incorporation method	Application	References
79	4-Amino-L-tryptophan		Trp	SPI	Spectral properties	Bae et al. (2003)
					Fluorescence ~ pH[h]	Budisa and Pal (2004)
					Protective effect[c]	Acevedo-Rocha et al. (2013)
					Photochemistry	Staudt et al. (2013)
80	5-Amino-L-tryptophan		Trp	SPI	Fluorescence ~ pH[h]	Budisa and Pal (2004)
81	4-Methyl-L-tryptophan		Trp	SPI	Spectral properties	Budisa et al. (2004)
82	4-Hydroxy-L-tryptophan		Trp	SPI	Spectral properties	Budisa and Pal (2004)
83	5-Hydroxy-L-tryptophan		Trp	SPI	Spectral properties	Budisa and Pal (2004)
84	4-Aza-L-tryptophan		Trp	SPI	Photochemistry	Staudt et al. (2013)
					Blue fluorescence; biosynthesis and residue-specific incorporation	Lepthien et al. (2008)

85	5-Aza-L-tryptophan		Trp	SPI	Blue fluorescence; biosynthesis and residue-specific incorporation	Lepthien et al. (2008)
86	7-Aza-L-tryptophan		Trp	SPI	Protective effect[c]	Acevedo-Rocha et al. (2013)
87	L-N^6-Acetyllysine		Pyl/Lys	SCS	PTM mimic	Neumann et al. (2009)
88	L-N^6-[(Allyloxy)carbonyl] lysine		Pyl/Lys	SCS	Substrate of wt PylRS BOC	Wan et al. (2014) Kaya et al. (2009)
89	N^6-[(Propargyloxy) carbonyl]-L-lysine		Pyl/Lys	SCS	Substrate of wt PylRS BOC with azides	Wan et al. (2014) Kaya et al. (2009)
90	N^6-Cyclopentyloxy-carbonyl-L-lysine		Pyl/Lys	SCS	Substrate of wt PylRS SCR	Wan et al. (2014) Krishnakumar et al. (2013)
91	N^6-L-Thiaprolyllysine		Pyl/Lys	SCS	BOC with 2-cyanobenzothiazole	Nguyen et al. (2011)

(continued)

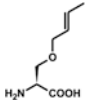
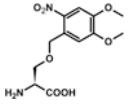
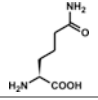
Table 4.2 (continued)

	Name	Structure	Analog of cAA	Incorporation method	Application	References
92	N^6-[(2-Nitrobenzyl)oxy] carbonyl-L-lysine		Pyl/Lys	SCS	Photocaged Lys	Jones et al. (2010)
93	N^6-[2-(Cyclooct-2-yn-1-yloxy)ethyloxy]carbonyl-L-lysine		Pyl/Lys	SCS	BOC by metal-free "click chemistry"	Plass et al. (2011)
94	N^6-(*endo*-norborn-2-en-5-methyloxy)carbonyl-L-lysine		Pyl/Lys	SCS	BOC with tetrazines	Kaya et al. (2012)
95	N^6-(1-Methylcycloprop-2-ene-1-carbonyl)-L-lysine		Pyl/Lys	SCS	BOC with tetrazoles ("photoclick" reaction)	Yu et al. (2012)
96	N^6-[(Bicyclo[6.1.0]non-4-yn-9-ylmethoxy) carbonyl]-L-lysine		Pyl/Lys	SCS	BOC with tetrazines	Lang et al. (2012a)

97	N^6-{[(*4E*)-Cyclooct-4-en-1-yloxy]carbonyl}-L-lysine		Pyl/Lys	SCS	BOC with tetrazines	Lang et al. (2012a)
98	N^6-{[(*4E*)-Bicyclo[6.1.0]non-4-en- 9-ylmethoxy]carbonyl}-L-lysine		Pyl/Lys	SCS	BOC with tetrazines	Lang et al. (2012a)
99	L-(7-Hydroxycoumarin-4-yl)ethylglycine		–	SCS	Enzyme activity ↑	Ugwumba et al. (2010)
100	L-Selenocysteine		–	SCS	Direct incorporation with synthetic suppressor tRNA and elongation factor	Aldag et al. (2013)
101	L-Phenylselenocysteine		Sec	SCS	Stable precursor of **102**; BOC of **102** with thiols	Wang et al. (2007b)
102	Dehydroalanine		Ala	SCS[i]	By oxidative elimination of **101** BOC with thiols	Wang et al. (2007b)
103	L-*S*-(2-Nitrobenzyl) cysteine		Cys	SCS	Photocaged Cys	Jones et al. (2010)

(continued)

Table 4.2 (continued)

	Name	Structure	Analog of cAA	Incorporation method	Application	References
104	L-*O*-Crotylserine		Ser	SCS	BOC by olefin metathesis	Ai et al. (2010)
105	L-*O*-(4,5-Dimethoxy-2-nitrobenzyl)serine		Ser	SCS	Photocaged Ser	Jones et al. (2010)
106	L-Homoglutamine		Gln	SCS	Incorporation at quadruplet codon	Anderson et al. (2004)

Abbreviations: *cAA* canonical amino acid; *BOC* bioorthogonal conjugation; *BONCAT* bioorthogonal noncanonical amino acid tagging, includes fluorescence labeling; *PFAA* polyfluorinated amino acid; *PTM* posttranslational modification; *SCR* sense codon recoding; *SPI* supplementation-based incorporation, residue specific; *SCS* stop codon suppression, site specific; *wt PylRS* wild-type pyrrolysyl-tRNA synthetase; ↑ enhanced; ↓ impaired

[a]Phase determination in X-ray crystallography

[b]Co-expression of *Ec*MetRS L13N Y260L H301L

[c]Protective effects against harsh conditions such as protein-denaturing agents; alkylating, denaturing, or inhibitory substances; tolerance toward organic solvents or surfactants

[d]Phage display of synthetic proteins with an expanded genetic code (Liu et al. 2008)

[e]Co-expression of *Ec*PheRS T251G A294G

[f]Co-expression of *Ec*PheRS A294G

[g]Co-expression of *Ec*LeuRS T252Y

[h]pH-sensitive fluorescence

[i]Dehydroalanine *per se* is not stable in aqueous milieu and can, therefore, not be incorporated by SCS. Instead, the stable precursor L-phenylselenocysteine (**101**) is incorporated by amber SCS first and then converted to dehydroalanine by treatment with H_2O_2 (Wang et al. 2007b)

4.2.6 Incorporation of ncAAs in In Vitro Expression Systems

Changing the genetic code in in vitro systems is a vast field and I will only briefly touch upon it by highlighting several selected studies. The interested reader is referred to the excellent review by Hong et al. (2014) for an overview.

Cell-free protein synthesis (CFPS) is the biosynthesis of proteins in a test tube without using intact, living cells (Hong et al. 2014). Typically, *E. coli* lysates are used that contain the intact cellular machinery for transcription and translation. The manipulation of the genetic code in cell-free protein expression systems basically relies on the same principles as in the in vivo systems described in detail in the Sects. 4.2.7 and 4.2.8. As the cell-free system resembles a chemical reaction that is catalyzed by multiple enzymes, the components in the mixture are easier to manipulate than living cells. Certain components of the translation machinery such as aaRSs or tRNAs can be added in defined amounts. Cell-free expression systems entirely composed of purified components were also described (Swartz 2001). Such systems are perfect alternatives for cultured cells if the chosen ncAAs are toxic and/or not taken up by the cells or when the protein product impairs cell viability (Zheng and Kwon 2012). The drawback is the usually lower overall protein yield because the reaction scales are limited and the cell-free synthesis systems are not easy to scale (Hong et al. 2014).

Wild-type and mutant aaRSs can either be expressed before the cell-free lysate is prepared or may be added in purified form. Together with suitable suppressor tRNAs, they were successfully used to incorporate ncAAs at stop codons (Goerke and Swartz 2009).Very conveniently, it is possible to chemically aminoacylate the tRNAs with ncAAs (Noren et al. 1989) before adding them to the reaction or to charge them using ribozymes (Bessho et al. 2002).

The group of Dougherty successfully introduced chemically aminoacylated tRNAs into *Xenopus* oocytes to site-specifically modify G-coupled receptors. In this way, they were able to insert a diverse set of ncAAs into the receptor proteins using stop or quadruplet codons. It allowed them to study the electrophysiology of the chemically modified receptors. The chemical aminoacylation of tRNAs is difficult and the modified receptor protein was produced in low yet sufficient amounts to serve as a mechanistic probe (Dougherty and Van Arnam 2014).

Using CFPS, Ugwumba et al. incorporated L-(7-hydroxycoumarin-4-yl) ethylglycine (Hco, **99**) at the catalytically active position Tyr309 of the phosphotriesterase arPTE of *Agrobacterium radiobacter* (Ugwumba et al. 2010). The enzyme catalyzes the hydrolysis of toxic organophosphate phosphotriester pesticides such as paraoxon. Improving the native activity of the enzyme was difficult: The authors screened a large number of cAA mutants for increased activity. However, none of these mutants showed a greater increase in activity than the synthetic Tyr309Hco variant that originated from the single exchange of a cAA for an ncAA (Ugwumba et al. 2010). Obviously, the functional scope of the cAAs was too narrow to improve arPTE activity. Only after expanding it to ncAAs an improvement was possible. Kolev et al. reported a similar observation along these lines for an engineered cytochrome P450 (Kolev et al. 2014).

The group of Yokoyama used CPFS for the site-specific incorporation of *para*-iodo-L-phenylalanine (**48**) for the X-ray crystal structure determination of proteins by single-wavelength anomalous diffraction (SAD) (Kodama et al. 2010).

The site-specific introduction of *para*-propargyloxy-L-phenylalanine (**53**) into T4 lysozyme facilitated the oriented immobilization of the enzyme on super-paramagnetic beads. The stability of the enzyme was better preserved by the oriented immobilization than if the enzyme was coupled nonspecifically via Arg or Lys residues or via the N-terminus (Wu et al. 2015).

The incorporation of *para*-azido-L-phenylalanine (**52**) and *para*-propargyloxy-L-phenylalanine (**53**) into superfolder green fluorescent protein enabled the directed polymerization of the protein using copper(I)-catalyzed azide–alkyne cycloaddition [CuAAC; "click chemistry" (Meldal and Tornøe 2008)] (Albayrak and Swartz 2014).

4.2.7 Engineering the Genetic Code: In Vivo Residue-Specific Incorporation of ncAAs

4.2.7.1 Principle

When it comes to engineering the genetic code in vivo, chemically or otherwise aminoacylated tRNAs cannot be used as they do not enter the cell. Consequently, the adaptor molecules delivering the ncAAs to the ribosome must be charged within the cell.

For the residue-specific incorporation of ncAAs, the host aaRSs are exploited. Put simply, the host translational machinery is hijacked to infiltrate ribosomal translation with the amino acid analog. In this way, sense codons can be transiently reassigned (Fig. 4.2c).

AaRSs are very accurate enzymes. They aminoacylate their cognate tRNAs with an error rate of ~10^{-4} (Ling et al. 2009). However, they can tolerate analogs that resemble their cognate cAAs in structure, chemistry, or shape. Naturally, the analogs are far worse substrates for the aaRSs than their cognate cAAs. If their concentration is reduced, however, and the concentration of the ncAAs is at the same time increased, the specificity of the aaRSs can be overcome (Ngo and Tirrell 2011). It is possible to control the intracellular availability of the cAAs and ncAAs by using amino acid auxotrophic host strains. These organisms are incapable of biosynthesizing one or several of the cAAs and rely on their supplementation in the culture medium for growth.

The residue-specific incorporation of ncAAs exploits exactly the natural substrate tolerance of the aaRSs. It is a very simple and straightforward approach and the successful, deliberate translation of an ncAA was first reported in the 1950s (Cohen and Cowie 1957; Cowie and Cohen 1957). Since then, the approach has been shown for a palette of chemically and structurally diverse ncAAs in *E. coli* (Ngo and Tirrell 2011; Wiltschi 2012; Budisa 2004), yeast (Budisa et al. 2010; Wiltschi et al. 2008), and mammalian cells (Beatty and Tirrell 2008; Dieterich et al. 2006).

Experimentally, the incorporation of an ncAA into a target protein resembles very closely the expression of recombinant proteins (Wiltschi 2012). Most importantly, a host strain is required that is auxotrophic for the cAA whose analog is to be incorporated. A collection of auxotrophic *E. coli* strains is available via the Coli Genetic Stock Center (CGSC) in Yale (http://cgsc.biology.yale.edu/). Auxotrophic *Saccharomyces cerevisiae* strains can be obtained from the EUROSCARF collection (www.euroscarf.de). Another important prerequisite for residue-specific incorporation of ncAAs is an inducible expression construct of the target protein.

The cells are first grown to the desired cell density before the onset of protein expression. To supplement the auxotrophy of the host, the appropriate amounts of the cAA must be supplied in the medium. Upon induction of the expression of the target gene, the cells are supplemented with the ncAA in the medium instead of the cAA. The cells are depleted for the cAA while the ncAA can accumulate (Wiltschi 2012) and exactly the condition is reached which allows to overcome the substrate specificity of the ncAAs (Ngo and Tirrell 2011). In the virtual absence of the cAA, the ncAA can be recognized by the appropriate aaRS and it is charged onto the cognate tRNA(s) despite being a much worse substrate.

Since the ncAA is used to supplement the amino acid auxotrophy of the host, this incorporation technique is also called supplementation-based incorporation (SPI). SPI is not confined to a single ncAA. Using a multi-auxotrophic host, the group of Budisa demonstrated the simultaneous incorporation of several different ncAAs in the same target protein (Lepthien et al. 2010; Merkel et al. 2010).

The ncAAs can only infiltrate translation if they pass the different layers of translation quality control (see Sect. 4.2.3). As evidenced by the many successful residue-specific incorporations reported in the literature, ncAAs undermine the control mechanisms quite efficiently. Presumably, they can do so because they were very early excluded from the standard genetic code and evolution forced the aaRSs to prevent their misincorporation merely in those organisms that biosynthesize them (see the example in Sect. 4.1.2.4).

Amino acid analogs not closely resembling their canonical counterparts are not incorporated into target proteins at times. Elevating the cellular expression level of the corresponding aaRS by its co- or overexpression can overcome the incorporation failure (Kim et al. 2004; Link et al. 2004; Kiick et al. 2000, 2001; Tang and Tirrell 2001). For even more dissimilar analogs, aaRS mutants with relaxed substrate specificity must be used (see Sect. 4.2.7.2) (Tang et al. 2009; Carrico et al. 2007; Datta et al. 2002; Kirshenbaum et al. 2002; Kast and Hennecke 1991).

As outlined in Sect. 4.2.3, the ribosome does not edit incorrectly aminoacylated tRNAs. That is why the ncAA-tRNAs can participate in translation once they have been formed. The ncAA will be stochastically incorporated into the target protein at any position that is normally occupied by the cAA it is analogous to. However, the ncAA is incorporated in the same way into any cellular protein that happens to be expressed in its presence. In this respect, the residue-specific incorporation of ncAAs resembles the accidental incorporation of ncAAs described in Sect. 4.1.2.4. The accidental incorporation of the ncAAs into the host proteome

with its potential toxic effects is an acceptable side effect as long as it does not impair the translation of the target protein. Obviously, the host cells will not grow during the expression of the target protein with the ncAA as they lack one of the 20 cAAs. Rather, they represent protein factories that run on the translation machinery that was produced during the growth phase in the presence of the cAAs. The incorporation of the ncAA into the target protein will be the more efficient the sooner it is finished. Eventually, the machinery might cease to work properly because it becomes too intoxicated by newly synthesized components themselves containing the ncAA.

Evidently, tightly controlled expression conditions are crucial for the homogeneous labeling of the target protein with an ncAA. This means that the host cells must be properly depleted for the cAA before the analog is added and the target protein expression is turned on. Experimentally, this can be achieved by either growing the cells on limiting amounts of the cAA which will eventually be consumed as indicated by the growth arrest of the cultured cells (Wiltschi 2012). Alternatively, the cells can be first grown with excess cAA and must then be thoroughly washed before they are shifted to fresh medium containing the ncAA (van Hest et al. 2000). The ncAA competes with its canonical counterpart for the aminoacylation and even minute amounts of free cAA can spoil the labeling efficiency. Due to the substrate specificities of the host aaRSs, the cAAs are always preferentially charged onto their cognate tRNA(s).

The mass analysis of the intact synthetic proteins containing ncAAs, for instance, by electrospray ionization mass analysis coupled to liquid chromatography (ESI LC-MS), represents a routine method to assess the homogeneity of the labeling.

Residue-specific labeling of proteins with ncAAs is a quite simple technique. Basically, neither the target genes nor the host translation machinery need any genetic manipulation. It can be performed in any environment equipped for protein expression, which makes it ever more popular for protein engineering (Zheng and Kwon 2012).

4.2.7.2 Examples for Genetic Code Engineering by Residue-Specific ncAA Incorporation

Numerous reports in the literature describe the effects of residue-specific ncAA incorporation into structural proteins as well as enzymes. Very recently, Zheng et al. comprehensively reviewed the manipulation of enzyme properties using ncAAs (Zheng and Kwon 2012). For the sake of brevity, I will focus here on selected examples to illustrate the potential of the approach.

Residue-Specific Incorporation of Seleno-ncAAs for Protein Structure Determination

Already in very early works, Cohen and Cowie demonstrated the tolerance of the *E. coli* translational apparatus for L-selenomethionine (**25**) (Cohen and Cowie 1957; Cowie and Cohen 1957). Today, the global labeling of proteins with ncAAs containing selenium, such as **25** and beta-(selenolo[3,2-*b*]pyrrol-3-yl)-L-alanine

(**68**), is a widespread method for the 3D structure analysis of proteins. Selenono-ncAAs are particularly well suited for the phase determination in X-ray crystallography (Bae et al. 2001; Hendrickson et al. 1990).

Manipulation of Enzyme Activity by ncAAs

The incorporation of the His analog L-3-(1,2,4-triazol-3-yl)alanine (**66**) into porcine pancreas phospholipase A2 produced a variant that was highly active at low pH while the parent protein containing His was inactive under the same conditions (Beiboer et al. 1996). However, the incorporation of **66** and another His analog, L-2-methylhistidine (**67**), into the alkaline phosphatase of *E. coli* inactivated the enzyme (Schlesinger and Schlesinger 1967, 1969).

NcAAs can improve the activity of an enzyme toward a specific substrate. For instance, the turnover number k_{cat} of glutathione-*S*-transferase was increased nearly fourfold toward 1-chloro-2,4-dinitrobenzene as the substrate by the incorporation of 5-fluoro-L-tryptophan (**74**). In contrast, the k_{cat} for phenanthrene 9,10-oxide and 4-phenyl-3-buten-2-one remained unchanged (Parsons et al. 1998).

The replacement of all 13 Met residues with L-norleucine (**26**) in the heme domain of cytochrome P450 BM-3, a single peptide cytochrome P450:NADPH-P450 reductase from *Bacillus megaterium*, increased peroxygenase activity nearly twofold; however, the thermostability of the variant enzyme was significantly reduced (Cirino et al. 2003).

Holzberger et al. achieved almost quantitative replacement, i.e., a substitution level of approximately 92 %, of the 32 Pro residues of the large fragment of DNA polymerase I from *Thermus aquaticus* with (2*S*,4*R*)-4-fluoroproline (**32**) (Holzberger and Marx 2010). They showed that numerous residues in a highly dynamic protein of high molecular weight (540 amino acids, 63 kDa) can be replaced by a noncanonical analog without effecting crucial enzyme properties. The multi-labeled KlenTaq DNA polymerase was still fully enzymatically active and it retained its fidelity and sensitivity. The multiple replacement of Pro by **32** caused some loss in thermostability (Holzberger and Marx 2010). However, the expression level of KlenTaq DNA polymerase with **32** was noticeably lower (0.2–0.5 mg/L) than that of the parent protein with Pro (8 mg/L). Attempts to incorporate the stereoisomer (2*S*,4*S*)-4-fluoroproline (**31**) failed (Holzberger and Marx 2010).

Antranikian and co-workers performed a comprehensive study on the effects of the incorporation of Met, Pro, Phe, and Tyr analogs in the lipase TTL of the thermophilic bacterium *Thermoanaerobacter thermohydrosulfuricus* (Hoesl et al. 2011). Lipases are conventionally used in synthetic organic chemistry and in pharmaceutical chemistry as biocatalysts with a broad substrate tolerance (Reetz 2002). TTL must be heated to gain full activity. Without thermal activation, a TTL variant containing L-norleucine (**26**) in place of Met exhibited a more than tenfold increase in activity as compared to the parent enzyme. (Hoesl et al. 2011). The study revealed the remarkable positional effects of fluorine substituents in aromatic ncAAs. While the substitution of Phe by *meta*-fluoro-L-phenylalanine (**46**) produced a variant that exceeded the parent's activity by 25 % after thermal activation, the variant enzyme containing *para*-fluoro-L-phenylalanine (**47**) was only 40 % as

active. A TTL variant with all Tyr residues exchanged for *ortho*-fluoro-L-tyrosine (**35**) was fully active after heat activation. In contrast, fluorination of the Tyr residues in the *meta*-position (*meta*-fluoro-L-tyrosine, **36**) inactivated the enzyme (Hoesl et al. 2011).

Mehta et al. reported similar observations along these lines. They were able to modulate the specificity of the histone acetyltransferase p300/CBP-associated factor, PCAF, from *Tetrahymena thermophila* by substituting Phe with *ortho*- (**45**), *meta*- (**46**), or *para*-fluoro-L-phenylalanine (**47**) (Mehta et al. 2011). PCAF promiscuously acetylates a target histone but also a nonhistone substrate. Fluorination of the Phe residues in the *meta*-position conferred selectivity for the histone substrate. In contrast, the variant containing the fluorine in the *para*-position of the Phe rings was inactive with the histone substrate, and *ortho*-fluorination abolished the enzyme activity altogether (Mehta et al. 2011). However, incorporation of the same three monofluorinated Phe regioisomers into the histone acetyltransferase, tGCN5, from *Tetrahymena thermophila* reduced the catalytic activity of the enzyme. The *meta*-fluorinated Phe analog had the least disruptive effect (Voloshchuk et al. 2009). In comparison to the Phe-containing parent protein, the fluorinated tGCN5 variants were less resistant toward proteolytic digestion with chymotrypsin, which cleaves after phenylalanine residues.

The group of Budisa studied the enzymatic activity of various different variants of the thermophilic lipase, TTL, in hostile environments (Acevedo-Rocha et al. 2013). They incorporated L-norleucine (**26**) and L-azidohomoalanine (**28**); 4-amino-L-tryptophan (**79**), 4-fluoro-L-tryptophan (**73**), and 7-aza-L-tryptophan (**86**); (2*S*,4*S*)-4-fluoro-L-proline (**31**), (2*S*,4*R*)-4-fluoro-L-proline (**32**), (2*S*,4*S*)-4-hydroxy-L-proline (**33**), and (2*S*,4*R*)-4-hydroxy-L-proline (**34**); *ortho*-fluoro-L-tyrosine (**35**) and *meta*-fluoro-L-tyrosine (**36**); and *meta*-fluoro-L-phenylalanine (**46**) and *para*-fluoro-L-phenylalanine (**47**). They found variants whose activity in organic solvents and surfactants was noticeably improved in comparison to the parent lipase. Selected analogs protected the enzyme against reducing, alkylating, denaturing, and inhibitory substances (Acevedo-Rocha et al. 2013).

Manipulation of Protein Folding and Stability by ncAAs

If any prediction is to be made about the effects that can be elicited by the incorporation of ncAAs into proteins, then it is that fluorine appears to have a great impact on the folding and stability. A couple of crystal structures of fluorinated proteins have begun to reveal the underlying principles.

Proline plays a central role in protein folding and stability (Milner-White et al. 1992), and its fluorinated analogs can be used to modulate protein structure and folding. Global labeling with (2*S*,4*R*)-4-fluoro-L-proline (**32**) enhanced the stability of ubiquitin (Crespo and Rubini 2011). The same Pro analog improved the thermostability of a single-chain antibody (Edwardraja et al. 2011). In contrast, substitution of the ten Pro residues of green fluorescent protein (GFP) by **32** drove the protein into insolubility. An exchange of the Pro residues for the isomer with the fluorine atom in a *cis*-configuration at C4 of the proline ring, (2*S*,4*S*)-4-fluoro-L-proline (**31**), furnished the GFP with excellent folding and stability properties that

exceeded those of the parent protein (Steiner et al. 2008). The incorporation of **31** and **32** often provokes opposing effects (Holzberger and Marx 2010; Crespo and Rubini 2011; Edwardraja et al. 2011; Steiner et al. 2008). The electronic properties of the fluorine substituent at the C4 position of the pyrrolidine ring of Pro can affect the *cis*/*trans* isomerization of peptide bonds and consequently the folding of proteins (for a more detailed recent review, see Odar et al. (2015) and the references therein).

Besides fluorinated Pro analogs, the fluoro derivatives of the aromatic amino acids have been used to modulate protein folding and stability. The global incorporation of *meta*-fluoro-L-tyrosine (**36**) into the omega-transaminase (ω-TA) from *Vibrio fluvialis* improved the thermostability and DMSO tolerance of the enzyme (Deepankumar et al. 2014). Baker et al. labeled the S5 phosphotriesterase from *Pseudomonas diminuta* with *para*-fluoro-L-phenylalanine (**47**) and found the refoldability and thermo-activity of the enzyme enhanced (Baker and Montclare 2011). The fluorination of the Tyr (by **36**), Phe (by **47**), or Trp residues (by 5-fluoro-L-tryptophan, **74**) prolonged the shelf life of *Candida antarctica* lipase B (Budisa et al. 2010).

The group of Tirrell pushed the limits of protein folding and stability by using polyfluorinated ncAAs. They enhanced the thermo- as well as the chemostability of coiled-coil peptides by substituting the Leu residues with 5′,5′,5′-trifluoro-L-leucine (**63**) and 5,5,5,5′,5′,5′-hexafluoroleucine (**64**) (Tang and Tirrell 2001; Tang et al. 2001). However, the concept was not directly applicable to a larger protein such as the chloramphenicol acetyltransferase (CAT). The global replacement of the 13 Leu residues of CAT with **63** resulted in the greatly reduced thermostability of the protein. Obviously, the trifluorinated leucines did not properly fit into the matrix of the CAT protein and provoked its swift thermal inactivation. To recover the loss in thermostability, Montclare et al. combined protein engineering by random mutagenesis of cAAs with protein engineering using ncAAs. Only two rounds of directed evolution in the presence of **63** adapted the CAT protein matrix such that it could accommodate the polyfluorinated Leu analog and regained its thermostability (Montclare and Tirrell 2006). In a similar approach, eleven rounds of random mutagenesis were necessary to accommodate **63** in GFP. Folding, as well as the physical and spectroscopic behavior of the sequence optimized GFP variant with **63** incorporated, was nearly identical to that of wild-type GFP (Yoo et al. 2007).

In a completely different approach using Met analogs, Wolschner et al. assessed the aggregation propensity of recombinant human prion protein with respect to the oxidation of Met residues. They observed that Met oxidation caused misfolding and aggregation while L-norleucine (**26**) prevented the human prion protein from aggregating. **26** is a non-oxidizable Met analog because the sulfur of Met (Fig. 4.1) is replaced by a CH_2 group. In contrast, L-methoxinine (**27**), an oxygen analog of Met that mimics oxidized Met, promoted the prion protein aggregation (Wolschner et al. 2009).

Currently, the effects of the substitution of a cAA with its noncanonical analog can rather be guessed than predicted—even though occasionally it is an educated

guess as, for instance, a body of evidence suggests that fluorination can improve protein stability. Evidently, the impact of an ncAA on the activity and/or stability of an enzyme is higher, (1) the more its polarity, chemistry, or shape deviates from the canonical counterpart and (2) the more abundantly it occurs in the target protein. To enhance stability and activity at once still remains a challenge though. As outlined above, even subtle positional changes of a single fluorine atom in an aromatic side chain can severely affect the catalytic activity or substrate binding affinity of a target enzyme (Zheng and Kwon 2012).

Manipulation of the Fluorescence Properties of Proteins by ncAAs

Trp and to a lesser extent Phe and Tyr (Fig. 4.1) are naturally fluorescent cAAs. Trp analogs have proved to be particularly valuable tools for the manipulation of the spectral properties of proteins. For the sake of brevity, I will list only a few highlight studies here. A comprehensive review by Budisa and Pal details the spectral properties of fluorescent and nonfluorescent proteins with a Trp-expanded genetic code that includes methyl- (**81**), amino- (**79**, **80**), hydroxy- (**82**), fluoro- (**73**–**76**), and aza-tryptophans (**84**–**86**) as well as Trp analogs containing sulfur or selenium (**68**–**71**) (Budisa and Pal 2004).

The replacement of the two Trp residues of enhanced cyan fluorescent protein by 4-amino-L-tryptophan (**79**) causes a very pronounced red shift of its fluorescence. In this way, Budisa and co-workers generated a gold fluorescent protein (Bae et al. 2003). The same group used a number of different Trp analogs to analyze the role of the Trp residues in enhanced green fluorescent protein and enhanced cyan fluorescent protein (Budisa et al. 2004).

Incorporation of 4- (**84**) or 5-aza-L-tryptophan (**85**) at the single Trp residue of the naturally colorless human annexin A5 resulted in an intrinsically blue fluorescing protein (Lepthien et al. 2008).

The group of Grieninger used **79**, **84**, and 4-fluoro-L-trypthophan (**73**), to manipulate the photocycle of a flavoprotein (Staudt et al. 2013).

The monofluorinated Phe analogs *ortho*- (**45**), *meta*- (**46**), and *para*-fluoro-L-phenylalanine (**47**) show characteristic and prominent shoulders ("fingers") in the range from 260 to 270 nm of the UV absorbance spectrum. They do not overlap with the contributions of Trp and Tyr at about 280 nm in the UV absorbance spectra of proteins. Therefore, the "phenylalanine fingers" represent spectroscopic probes to identify labeled target proteins among unlabeled cellular proteins by simple UV spectroscopy (Minks et al. 2000).

Reactive ncAAs and Bioorthogonal Conjugation

A chemical conjugation reaction is called bioorthogonal if it selectively links two compatible nonnatural, ideally non-perturbing, chemical groups. Importantly, the reaction occurs under physiological conditions, ideally within living cells. Only a limited number of chemical reactions fulfill these criteria and a handful have been adapted for protein engineering (Kim et al. 2013). NcAAs carrying a bioorthogonal functionality can be installed in proteins, and they can react with ligands bearing a

complementary reactive group. In this way, artificial posttranslational modifications with a ligand of choice are feasible.

For instance, terminal azido groups react selectively with terminal alkynes in the presence of copper (I) as the catalyst [Cu(I)-catalyzed azide–alkyne conjugation, CuAAC, also known as "click chemistry" (Meldal and Tornøe 2008)]. In addition, the azide moiety can undergo a chemoselective Staudinger ligation with phosphine ligands (Kiick et al. 2002). Azido groups have not yet been found in naturally occurring ncAAs and alkyne moieties are rare (Hunt 1985); thus both represent bioorthogonal groups. The Met analogs L-azidohomoalanine (**28**) and L-homopropargylglycine (**30**) bear an azido and a terminal alkyne group in their side chains, respectively, so that they can be used for CuAAC with proteins. Importantly, both ncAAs can be incorporated residue-specifically using the SPI technique.

Preferably, ncAAs with bioorthogonal functionalities are incorporated site-specifically into a target protein to, e.g., facilitate the directed attachment of a ligand at only one predefined position. Nevertheless, the residue-specific incorporation of reactive ncAAs was described as well.

Strable et al. (2008) demonstrated the successful incorporation of **28** and **30** into viruslike particles and subsequent fluorescence labeling. Prasuhn et al. incorporated **28** into phage particles. The phages were then conjugated to a lanthanide-metal chelate to determine their plasma circulation lifetime in mice by inductively coupled plasma optical emission spectroscopy (ICP-OES) (Prasuhn et al. 2008).

Link et al. inserted **28** into the outer membrane protein C (OmpC) for cell surface labeling by "click chemistry" (Link and Tirrell 2003). Teeuwen et al. (2009) replaced Met with **28** and **30** in elastin-like proteins and modified the functionalized proteins with fluorophores, polyethylene glycol, and an enzyme engineered to carry the complementary reactive group.

Süssmuth and co-workers reported on the chemical diversification of lantibiotics by the incorporation of ncAAs and artificial posttranslational modification. In addition to other ncAAs without reactive side chains, they incorporated **28** and **30** into the peptide antibiotics and then conjugated them with sugar molecules and fluorescent dyes (Oldach et al. 2012).

In addition to **28** and **30**, *para*-azido-L-phenylalanine (**52**) is a photoreactive ncAA that has been successfully used for CuAAC. However, the *E. coli* PheRS does not accept **52** as a substrate. Carrico et al. used the *E. coli* PheRS A294G mutant to label an artificial extracellular matrix protein with **52**. The *E. coli* PheRS carrying mutations at position A294 shows particularly relaxed substrate specificity. It accepts a number of Phe and Tyr analogs due to the increased volume of its amino acid-binding pocket (Kirshenbaum et al. 2002; Kast and Hennecke 1991). Carrico and co-workers exploited the photoreactivity of the azido-Phe-labeled artificial extracellular matrix protein for protein patterning by photolithography (Carrico et al. 2007).

The reaction of keto groups with hydrazines or alkoxyamines forming stable hydrazone or oxime bonds is bioorthogonal (Kim et al. 2013). Datta et al. replaced the Phe residues of murine dehydrofolate reductase with *para*-acetyl-L-

phenylalanine (**50**) for subsequent bioorthogonal conjugation via hydrazine reagents (Datta et al. 2002). Similar to **52**, the keto-functionalized ncAA **50** is not recognized by the *E. coli* TyrRS. Thus, for its incorporation at Phe codons, another PheRS mutant with relaxed substrate specificity, *Ec*PheRS T251G A294G, was used.

Tang et al. (2009) employed the editing-defective *Ec*LeuRS T252Y mutant for the incorporation of the reactive ketoamino acid L-4-oxonorvaline (**65**).

Proteome-Wide Residue-Specific Incorporation of ncAAs

As outlined in the previous section, the residue-specific incorporation of reactive ncAAs into single proteins demonstrated the feasibility of bioorthogonal conjugation reactions at proteins. Recently, the technique has experienced its expansion to the labeling of whole proteomes with ncAAs carrying bioorthogonal groups. The group of Tirrell has pioneered the proteome-wide residue-specific incorporation of reactive ncAAs.

The cells are treated with the reactive ncAA at the same time when a stimulus is applied. Consequently, all proteins newly synthesized in response to the stimulus will incorporate the ncAA with the functional side chain. The bioorthogonal groups can be conjugated to an affinity tag such as biotin for the isolation of the proteins. Subsequently, the isolated proteins are identified by mass analysis (bioorthogonal noncanonical amino acid tagging, BONCAT). Alternatively, the stimulus-induced proteins may be visualized by attaching fluorophores to their functional moieties (fluorescent noncanonical amino acid tagging, FUNCAT) (Ngo and Tirrell 2011; Landgraf et al. 2015; Hinz et al. 2013).

BONCAT and FUNCAT experiments usually employ L-azidohomoalanine (**28**) or less frequently L-homopropargylglycine (**30**), which is conjugated by CuAAC to biotin or fluorescent dyes carrying the compatible functional moiety. Commercial Click-iT AHA (**28**) and HPG (**30**) kits for the detection of nascent protein synthesis in a variety of cell types and tissues are available (http://www.lifetechnologies.com).

BONCAT was reported for *E. coli* (Ngo et al. 2009; Nessen et al. 2009), but it was primarily applied to analyze the proteomes of mammalian cell lines such as HEK293 (Dieterich et al. 2006), HeLa (Bagert et al. 2014), and primary human CD4 T lymphocytes (Howden et al. 2013). Eichelbaum et al. (2012) used the labeling technique for the quantitative analysis of the secretome, i.e., the subset of secreted proteins, in mammalian cells. The method proved particularly useful to study nascent protein biosynthesis in cultured neurons (Dieterich et al. 2006; Landgraf et al. 2015; Yoon Byung et al. 2012) as well as in intact hippocampal neuropil sections after dopaminergic stimulation (Hodas et al. 2012). BONCAT was successfully applied to study newly synthesized proteins in intact zebrafish larvae (Hinz et al. 2012).

FUNCAT, unlike BONCAT, was first developed in *E. coli* (Beatty et al. 2005) and was then successfully applied in mammalian cells (Beatty et al. 2006; Ngo et al. 2013), cultured neurons (Yoon Byung et al. 2012; Dieterich et al. 2010), as well as animals (Hinz et al. 2012). The reactive ncAAs **28** and **30** can be used with

two different stimuli to track spatial and temporal changes in protein biosynthesis (Beatty and Tirrell 2008). Applying FUNCAT, Ouellette et al. showed that pathogenic *Chlamydia* co-opt the functions of lysosomes to acquire essential amino acids (Ouellette et al. 2011). Hang and co-workers visualized the proteome of *Salmonella typhimurium* during infection of mammalian cells (Grammel et al. 2010). Ngo et al. demonstrated the specificity of the method by selectively labeling a specific cell population in mixed cultures of *E. coli* and mammalian cells. They used the reactive Met analog L-azidonorleucine (**29**), which is only translationally active upon co-expression of the *E. coli* MetRS L13N Y260L H301L mutant ("NLL-MetRS") (Ngo et al. 2009).

To avoid the cytotoxic copper catalyst that is necessary for CuAAC, Tirrell and co-workers devised fluorophore-functionalized strain-promoted cycloalkynes for copper-free labeling of azido groups. They demonstrated the utility of their approach for FUNCAT with cultured rat fibroblasts (Beatty et al. 2010).

BONCAT in combination with stable-isotope labeling of amino acids in cell culture (SILAC) and nanoscale liquid chromatography–tandem mass spectrometry allows to quantitate relative differences in the abundance of cellular proteins (Howden et al. 2013).

4.2.8 Expansion of the Genetic Code: Site-Specific Incorporation of ncAAs

4.2.8.1 Principle

Residue-specific incorporation of ncAAs is an all-or-none process: all codons specifying one of the cAAs are reassigned to a noncanonical analog and it is not possible to deliberately introduce the ncAA at selected positions while its canonical counterpart occupies the remaining codons. If the ncAA is to be inserted at a specific position, for instance, in the catalytic site of an enzyme, the coding sequence must contain the appropriate codon at the desired position and all other codons specifying the same cAA must be mutated. As a direct consequence, particularly if the concerned cAA occurs frequently, the folding and stability as well as the activity of the protein can be severely compromised. The residue-specific incorporation of ncAAs is a valued method to modulate the physicochemical properties of proteins, such as folding and stability, because they originate from the synergy of all amino acid residues in the protein matrix. To manipulate a protein at a single or only a few positions, however, a site-specific alternative is indispensable.

Most importantly, the site-specific incorporation of an ncAA (Fig. 4.2d) warrants a codon to specify it. Since all codons of the genetic code are occupied by cAAs or translation termination signals (Fig. 4.1), there is no "blank" codon that could be assigned to the ncAA. Rather, a sense or a stop codon must be *reassigned*. Considering how the message of the codons is decoded, namely, by a single cAA-specific aaRS that charges all isoacceptor tRNAs (see Sect. 4.2.3), clearly

reveals the dilemma: How can the meaning of a single codon be changed without compromising the meaning of the other codons encoding the same message?

In theory, this means to extract one codon from the genetic code, to change its meaning entirely, and to append it again as an extension to the existing code. At the same time, the loss of this codon must not derogate the running system, which is the ribosomal translation relying on the standard genetic code.

There are two examples from nature that solved exactly this problem in a stunningly elegant way: selenocysteine and pyrrolysine. Apparently, both naturally occurring ncAAs entered the genetic code *after* it had been established in the present form (Ambrogelly et al. 2007). Their different incorporation mechanisms at stop codons (see Sects. 4.2.4.1 and 4.2.4.2) are a source of inspiration for the expansion of the genetic code with other ncAAs (Aldag et al. 2009, 2013; Wan et al. 2014).

Stop codons have been the main targets for reassignment to the site-specific incorporation of ncAAs. Occasionally, they are called nonsense codons, which implies that they make no sense and are therefore "blank" codons. It is true that the stop codons do not encode a cAA but they specify where translation must stop. Consequently, if a stop codon is reassigned, the genetic code loses a coding unit for the translation termination message.

However, the mechanism by which stop codons are decoded does indeed make a difference for their reassignment. In contrast to the sense codons, the meaning of the stop codons is not decoded by charged tRNAs but rather by the proteinaceous release factors (see Sect. 4.2.3). The meaning of a stop codon, for instance, the amber UAG stop, can be easily changed if the anticodon of an existing tRNA is mutated to the appropriate anticodon. Only a single base change in the anticodon of the tRNAs decoding the CAG (Gln), AAG (Lys), and GAG (Glu); UUG (Leu), UCG (Ser), and UGG (Trp); and UAU and UAC (both Tyr) codons suffices (the potential target bases are underlined) to generate an amber suppressor tRNA. This will expand the pool of existing tRNAs by a species that is able to decode the amber codon. If the anticodon switch does not impair the aminoacylation of the amber suppressor tRNA, it can still be charged with its cognate cAA, which could be Gln, Lys, Glu, Leu, Ser, Trp, or Tyr. At the ribosome, the mischarged amber suppressor tRNA can compete with the release factor for the amber stop codon. In each case, where it outcompetes the release factor, it delivers its charged cAA to the nascent polypeptide in response to the amber codon and translation progresses to the next downstream stop codon. In this way, the meaning of the amber codon as specified by the genetic code, namely, "translation termination," is suppressed in favor of a new meaning, namely, "insert a cAA." Such suppressor tRNAs do exist in nature and the suppression of stop codons aided in deciphering the genetic code (Brenner et al. 1965).

Being a natural phenomenon, stop codon suppression (SCS) and suppressor tRNAs can be exploited to incorporate ncAAs in a site-specific manner if the substrate specificity of the aaRS is changed in favor of the ncAA (Liu and Schultz 2010).

The amber codon has received by far the most attention for reassignment to ncAAs, but the UAA ochre codon (Xiao et al. 2013; Wan et al. 2010) and the UGA opal codon (Kodama et al. 2010) were used as well. Anderson et al. (2004) and the group of Chin (see Sect. 4.2.8.3) explored four-base codons, so-called quadruplet codons for the site-specific incorporation of ncAAs.

The preference of the amber codon is due to its low frequency in the genomes of many host organisms. For instance, in *E. coli*, the amber codon is the rarest of the three stop codons (http://www.kazusa.or.jp/codon/) (Lajoie et al. 2013; Nakamura et al. 2000). It is anticipated that the accidental suppression of genomic amber codons leads to less severe effects than the suppression of a more frequent stop codon. In analogy to the accidental residue-specific incorporation of ncAAs into the host proteome, the accidental suppression of genomic stop codons can occur. However, due to the fact that stop codons generally occur less frequently than sense codons, the effects are considered to be minor.

In addition to the stop codons as site-specific insertion signals for ncAAs, the recoding of sense codons was assessed but this approach is still in its infancy (see Sect. 4.2.8.5).

Experimentally, the site-specific incorporation of an ncAA, e.g., at an in-frame stop or quadruplet codon, requires the co-expression of the suppressor tRNA and an aaRS that specifically charges it with the ncAA (Fig. 4.2d). To truly expand the genetic code, the aaRS must be specific for its ncAA as well as for the suppressor tRNA and it must not interact with any of the host tRNAs. Likewise, the suppressor tRNA must not be charged by the cAA-specific host aaRSs. As such, the aaRS and the suppressor tRNA form an orthogonal pair (o-pair) that acts independently of and does not interfere with the host aaRSs and their cognate tRNAs.

An ncAA-charged suppressor tRNA (ncAA-$tRNA_{sup}$) competes with the release factor for its cognate stop codon. To outcompete the release factor, the ncAA-$tRNA_{sup}$ must implicitly be present in sufficient amounts. Thus, the read-through efficiency at the stop codon depends fundamentally on the specificity and efficacy of the orthogonal pair. The translation of ncAAs, however, is generally much less efficient than that of cAAs (O'Donoghue et al. 2013). In addition, the local sequence context dictates how efficiently a specific stop codon can be suppressed (Wu et al. 2013). The position of the in-frame stop codon does indeed matter and the ncAA is not incorporated equally efficiently at any position. In *E. coli*, proteins containing ncAAs at amber positions are produced at levels of 25–75 % of the wild-type protein, which equals to titers in the range of several milligrams to tens of milligrams per liter of cell culture (Wang et al. 2009). The translational fidelity of the ncAAs is >99 %. The latest amber-specific ncAA incorporation systems for yeasts deliver similar titers and translational fidelities as in *E. coli*. The suppression efficiencies in mammalian cells are between 13 % and 41 % with a high fidelity of >95 % (Wang et al. 2009).

The read-through of more than one amber stop codon is usually very inefficient but it can be improved by deleting the release factor 1 (RF1). RF1 specifically recognizes amber codons in *E. coli* (Scolnick et al. 1968). However, the deletion of RF1 in *E. coli* wild-type strains has been considered lethal until recently (Johnson

et al. 2012). Mukai et al. constructed a conditional RF1 mutant strain in which the expression of RF1 could be suppressed under controlled conditions (Mukai et al. 2010). Using an appropriate o-pair, they were able to suppress six in-frame amber codons in their model protein glutathione-*S*-transferase with *meta*-iodo-L-tyrosine (**37**) in this strain. Johnson et al. reverted the lethality of the RF1 deletion by curing the +1-frameshift and the A246T mutation of release factor RF2 (Johnson et al. 2011), which specifically decodes ochre (UAA) and opal (UGA) codons. RF2 A246T has a fivefold reduced ability to recognize UAA codons, and the T246A back mutation fully restored the UAA recognition (Johnson et al. 2011). The resulting strain facilitated the efficient suppression of ten amber codons in enhanced green fluorescent protein by *para*-acetyl-L-phenylalanine (**50**) using a specific o-pair (Johnson et al. 2011). Wu et al. achieved incorporation of different ncAAs at multiple sites in an elastin-mimetic protein polymer by using a strain that expresses a mutant RF1 with an attenuated translation termination activity (Wu et al. 2013).

Proteins in which one or several in-frame stop codons are suppressed usually carry a purification tag at their carboxy-termini. This ensures the selective purification of only those proteins where the read-through with the ncAA occurred. The protein preparations are almost uniform since the incorporation fidelity, that is, the identity of the amino acid at the position of the stop codon, is very high (Wang et al. 2009).

In contrast, it is normally not possible to discriminate labeled from partially or unlabeled proteins when the ncAA is incorporated in a residue-specific manner. The substitution efficiency can range from very inhomogeneous to quantitative, depending on the ncAA and the protein. In these cases, the incorporation of certain ncAAs can be improved by the co-(over)expression of the host aaRS (Kiick et al. 2000, 2001; Tang and Tirrell 2001; Kirshenbaum et al. 2002) or of mutant aaRSs with relaxed substrate specificity (Sharma et al. 2000; Petrović et al. 2013; Kwon and Lim 2015; Kwon et al. 2003) (see also the sections 4.2.7.1. and "Reactive ncAAs and Bioorthogonal Conjugation").

To avoid accidental incorporation by the cellular translation machinery, ncAAs that are incorporated site-specifically by SCS should differ as much as possible from their canonical counterparts. This can be a difficult premise for the monofluorinated ncAAs because they often resemble their cAAs very closely (Odar et al. 2015). Different approaches to cope with this issue have been reported. Furter first showed the incorporation of *para*-fluoro-L-phenylalanine (**47**) at an amber codon by SCS with a PheRS/$tRNA_{CUA}^{Phe}$ pair from yeast. He used this pair in an analog-resistant *E. coli* strain that expressed a mutant PheRS A294S that excluded **47** from its active site (Furter 1998). Wilkins et al. avoided the discrimination of monofluorinated Tyr from Tyr altogether by using *O*-nitrobenzyl derivatives of fluorinated Tyr analogs. These were not recognized by the host TyrRS but required the co-expression of the appropriate o-pair for incorporation into the target protein. The protective *O*-nitrobenzyl group could be removed by UV irradiation after incorporation into the target protein, which liberated the fluorinated analog (Wilkins et al. 2010). Minnihan et al. (2011) managed to evolve

an orthogonal mutant aaRS that is specific for monofluorinated Tyr but does not activate Tyr.

4.2.8.2 Orthogonal Pairs

The site-specific incorporation of ncAAs relies on orthogonal pairs consisting of an aaRS and a suppressor tRNA for the activation, charging, and delivery of the analogs to the ribosome where they are used to decode in-frame stop or quadruplet codons. The orthogonality of the aaRS/$tRNA_{sup}$ pairs is crucial for this concept to work (see the Sect. 4.2.8.1).

The generation of o-pairs draws on the fact that the tRNA identity elements evolved independently in archaea, prokaryotes, and eukarya (Giege et al. 1998). Consequently, aaRSs and their cognate tRNAs from an evolutionarily distant organism are often orthogonal in common expression hosts. For instance, aaRS/tRNA pairs from archaea were shown to be orthogonal in *E. coli* (Wan et al. 2014; Wang et al. 2001) and those from archaea or *E. coli* turned out to be orthogonal in yeast or mammalian cells (Mukai et al. 2008; Chin et al. 2003a).

The existing aaRSs recognize either one of the cAAs or Pyl and the cognate tRNA(s) as their substrates. The natural aaRSs must be evolved to specifically accept ncAAs and suppressor tRNAs, and the selection of efficient o-pairs demands a suitable screening assay. Basically a two-step procedure is followed (Liu and Schultz 2010): During the positive selection step, cells transformed with a library of o-pairs are screened against a drug or antibiotic in medium containing the cAAs as well as a desired ncAA. In addition to the o-pair library, the cells express a drug resistance gene that contains an amber codon at a permissive position. This means that (theoretically) any amino acid can be incorporated at that position without impairing the function of the gene product. Successful suppression of the amber codon by o-pairs using either the ncAA or a cAA allows the cells to survive the positive selection step while cells carrying nonfunctional o-pairs will die. The second, negative selection step discriminates against o-pairs that use cAAs. It is performed with the surviving cells from the positive screen. They are co-transformed with a toxic gene carrying an amber codon at a permissive site. When grown without the ncAA, only those cells will survive that express o-pairs that do *not use* a cAA. Usually, two to three rounds of positive/negative screening are performed and it facilitates the selection of mutant aaRSs as well as mutant $tRNA_{sup}$ (Liu and Schultz 2010).

Several variations of the above-described procedure have been published. In an early approach, Wang et al. used chloramphenicol together with a permissive amber mutant of chloramphenicol acetyl transferase (CAT) for the positive and negative screening of an o-pair library in *E. coli*. They grew the survivors of the positive screen on a replica plate and then challenged them on chloramphenicol medium without the ncAA. They discarded the surviving cells, which obviously expressed o-pairs that were able to use cAAs, and retrieved the chloramphenicol-sensitive cells expressing the ncAA-specific o-pairs from the replica plate (Wang et al. 2001). The same group refined the assay and used ampicillin together with an amber mutant of beta-lactamase for the positive selection and the toxic barnase inhibitor

barstar for the negative selection (Wang and Schultz 2001). Later, barnase was still used to counter-select cAA-specific o-pairs but CAT together with chloramphenicol prevailed in the positive selection (Zhang et al. 2002). Chin et al. used a barnase gene with three permissive in-frame amber codons to improve the specificity of the selected o-pairs. In addition, they added another component to the positive screen: They included a permissive amber mutant of T7 RNA polymerase that drove the expression of GFP from a T7 promoter. This enabled them to select fluorescent cells in the presence of the ncAA (Chin et al. 2002a). Santoro et al. used a similar amber T7 RNA polymerase/pT7-GFP construct and applied fluorescence-activated cell sorting (FACS) to select for nonfluorescent cells in the negative selection (Santoro et al. 2002). Recently, Skerra and co-workers devised a screen that employs FACS during the positive and the negative selection step. They introduced the amber mutation directly upstream of the GFP fluorophore such that its suppression with the ncAA or a cAA during the positive selection leads to a fluorescent signal that can be sorted by FACS. The fluorescence readout in the absence of the ncAA provides a direct measure for the specificity of the selected o-pairs (Kuhn et al. 2010).

O-pairs are generally evolved for a desired ncAA but some pairs accept structurally and chemically similar analogs (Stokes et al. 2009; Young et al. 2011). This polysubstrate specificity resembles the substrate tolerance of the natural aaRSs.

The positive/negative selection scheme outlined above was first developed for *E. coli* (Wang et al. 2001; Wang and Schultz 2001; Zhang et al. 2002) and later adapted for *S. cerevisiae* (Chin et al. 2003b). O-pairs that had been evolved in these hosts could be successfully transferred to other prokaryotic and eukaryotic cells (Mukai et al. 2008; Wang et al. 2010; Young et al. 2009; Liu et al. 2007), which spared the tedious adaption of the selection procedure. Moreover, large aaRS and tRNA libraries can be screened in *E. coli* and *S. cerevisiae* (Liu and Schultz 2010).

Usually, the specificity of an o-pair for the desired ncAA is evolved first using the established procedure for amber suppression. Afterwards, the o-pair may be adapted to decode other stop codons (Wan et al. 2010) or quadruplet codons (Anderson et al. 2004).

O-Pairs Derived from TyrRS

Tyrosyl-tRNA synthetase (TyrRS) was one of the first natural aaRSs that was engineered for the site-specific incorporation of Tyr and Phe analogs (Liu and Schultz 2010). The Schultz group chose to evolve the TyrRS from *Methanocaldococcus jannaschii* together with its cognate $tRNA_{AUA/GUA}^{Tyr}$ (Wang et al. 2001) because the enzyme appeared to be an ideal candidate: The *Mj*TyrRS has a minimalist anticodon loop-binding domain, which is why the anticodon of its cognate tRNAs represents only a minor identity element. This is an important prerequisite for the efficient recognition of an amber suppressor $tRNA_{CUA}$. Moreover, *Mj*TyrRS lacks an editing mechanism (Wang et al. 2001), and an ncAA would most probably not be proofread. The $tRNA^{Tyr}$ could be easily turned into an amber suppressor *Mj*$tRNA_{CUA}^{Tyr}$ by introducing a single mutation into the anticodon (see the genetic code in Fig. 4.1). While the suppressor tRNA was a poor substrate for

the *E. coli* TyrRS, it was efficiently aminoacylated by the *Mj*TyrRS, and it functioned well in the ribosomal translation of the bacterial host (Wang et al. 2001).

To evolve the specificity of the orthogonal *Mj*TyrRS/*Mj*tRNA$_{CUA}^{Tyr}$ pair for the Tyr analog *O*-methyl-L-tyrosine (**38**), Wang et al. mutated five active site residues within 6.5 Å of the *para*-position of the aryl ring of the bound tyrosine (Wang et al. 2001). Applying the positive/negative selection scheme outlined in the previous section, they selected a mutant *Mj*TyrRS, which was approximately 13-fold more affine for **38** than for Tyr, and the turnover of the analog occurred approximately eightfold faster than that of Tyr (Wang et al. 2001). In this way, they had generated a catalytically highly effective o-pair for the incorporation of **38** at amber codons. Following these lines, the orthogonal *Mj*TyrRS/*Mj*tRNA$_{CUA}^{Tyr}$ pair was evolved for a vast variety of other Tyr and Phe analogs (see Liu and Schultz (2010) for an overview).

The *Ec*TyrRS/*Ec*tRNA^{Tyr} pair from *E. coli*, which is orthogonal in *S. cerevisiae* and *P. pastoris*, was evolved for the incorporation of Phe and Tyr analogs in the yeasts (Liu and Schultz 2010). The *Ec*TyrRS/*Bacillus stearothermophilus* tRNA^{Tyr} was used to evolve o-pairs for Tyr and Phe analog incorporation in mammalian cells (Liu and Schultz 2010).

O-Pairs Derived from PylRS

The pyrrolysyl-tRNA synthetase (PylRS) is an outstanding tool for genetic code engineering (Wan et al. 2014). Together with its cognate amber suppressor tRNA$_{CUA}^{Pyl}$, it represents a naturally occurring o-pair for the decoding of amber codons (see Sect. 4.2.4.2 for the incorporation mechanism of Pyl).

Natural PylRS, for instance, from the methanogenic archaeon *Methanosarcina mazei* (*Mm*PylRS), accepts more than 20 structurally and chemically diverse substrates, which are mainly Lys derivatives (Wan et al. 2014). Analogs bearing alkene, alkyne, azido, norbornene, and aminothiol groups are activated by wild-type *Mm*PylRS. Once charged onto *Mm*tRNA$_{CUA}^{Pyl}$, they are incorporated into proteins in response to amber codons (Wan et al. 2014). The enzyme uses relatively unspecific hydrophobic interactions to recognize the side chain of its canonical substrate Pyl (**22**), which itself is a derivative of Lys (refer to Fig. 4.1). Moreover, it does not have an editing function. The relaxed substrate binding combined with the missing editing function obviously affords the extraordinary substrate tolerance of the PylRS (Wan et al. 2014).

The *Mm*PylRS/*Mm*tRNA$_{CUA}^{Pyl}$ pair was further evolved for the site-specific incorporation of substrates that are only weakly accepted by the wild-type pair or of Lys analogs with highly complex side chains, e.g., see compounds **95** (Yu et al. 2012) or **96**–**98** (Lang et al. 2012a). Notably, PylRS can be engineered to accept not only aliphatic Pyl or Lys derivatives but also aromatic Phe derivatives (Wan et al. 2014; Wang et al. 2011). Using an evolved PylRS/tRNA$_{CUA}^{Pyl}$ pair, alpha-hydroxy acids could be selectively incorporated into proteins. Refer to Wan et al. (2014) and references therein for a comprehensive overview of the mutant PylRSs and their mutations relative to the wild-type synthetase.

The suppressor $tRNA_{CUA}^{Pyl}$ has a unique structure, which is specifically recognized by the PylRS. This is why the PylRS/$tRNA_{CUA}^{Pyl}$ pair is orthogonal in the commonly used prokaryotic and eukaryotic expression hosts (Wan et al. 2014). Wild-type and mutant PylRS/$tRNA_{CUA}^{Pyl}$ pairs have been used to site-specifically incorporate ncAAs into proteins in *E. coli* (Blight et al. 2004), in *S. cerevisiae* and mammalian cells (Mukai et al. 2008), or even in animals such as *Drosophila melanogaster* (Bianco et al. 2012) and *Caenorhabditis elegans* (Greiss and Chin 2011).

PylRS shows a high flexibility toward the anticodon of its cognate $tRNA^{Pyl}$. The $tRNA_{UCA}^{Pyl}$ opal and $tRNA_{UUA}^{Pyl}$ ochre suppressors derived from the amber suppressor *Mm*$tRNA_{CUA}^{Pyl}$ were able to decode opal UGA and ochre UAA stop codons, respectively, yet less efficiently than $tRNA_{CUA}^{Pyl}$, which suppresses amber codons. The mutant suppressor tRNAs for the rare Arg codon AGG (Odoi et al. 2013) as well as for quadruplet codons (O'Donoghue et al. 2012) could not completely outcompete the competing host tRNAs.

O-Pairs Derived from Other aaRSs

Wu et al. evolved the *Ec*LeuRS/*Ec*$tRNA^{Leu}$ pair from *E. coli* for the incorporation of cysteine, tyrosine, and alanine analogs at amber codons in *S. cerevisiae* (Wu et al. 2004). A mutant o-pair derived from the TrpRS/$tRNA_{CCA}^{Trp}$ of *Bacillus subtilis* facilitated the incorporation of 5-hydroxy-L-tryptophan (**83**) at opal codons in mammalian cells (Zhang et al. 2004). Anderson et al. used the LysRS of *Pyrococcus horikoshii* and a designed mutant suppressor tRNA to decode a quadruplet AGGA codon with L-homoglutamine (**106**) in *E. coli* (Anderson et al. 2004). Chatterjee et al. used a mixed archaeal orthogonal pair consisting of the ProRS from *Pyrococcus horikoshii* and *Archaeoglobus fulgidus* $tRNA^{Pro}$ to incorporate Pro at amber codons as well as quadruplet CCCU and CUAG codons in *E. coli* (Chatterjee et al. 2012).

4.2.8.3 Orthogonal Ribosomes

Chin and co-workers devised orthogonal ribosomes to enhance the efficiency of amber decoding (Wang et al. 2007a). They intended to separate the translation of ncAA-containing synthetic proteins from the translation of cellular proteins. To this end, they equipped their mRNAs with a Shine–Dalgarno sequence that directed their translation by the orthogonal ribosome, which carried the appropriate anti-Shine–Dalgarno sequence at the 3′-terminal region of the 16S rRNA, but precluded translation by host ribosomes (Rackham and Chin 2005). In addition, the orthogonal ribosome carried mutations in the region of the 16S rRNA, which forms the entry site for the aminoacyl-tRNA, to improve the competition of the ncAA-charged amber suppressor $tRNA_{CUA}$ with the release factor RF1. They were able to elevate the read-through efficiency from 24 % to 62 % with *para*-benzoyl-L-phenylalanine (**51**) and the orthogonal BpaRS/*Mj*$tRNA_{CUA}$ pair at a single in-frame amber codon and from below 1 % to 22 % read-through efficiency at two in-frame amber codons (Wang et al. 2007a).

Drawing on the orthogonal ribosome with improved amber decoding, Neumann et al. used the same approach to produce an orthogonal ribosome for the improved decoding of quadruplet AGGA codons (Neumann et al. 2010). The resulting new orthogonal ribosome facilitated the efficient read-through not only at quadruplet AGGA codons but also at amber stop codons: They successfully incorporated *para*-azido-L-phenylalanine (**52**) in response to an AGGA codon and N^6-[(propargyloxy) carbonyl]-L-lysine (**89**) in response to an amber UAG codon in the same target protein.

For a recent review of orthogonal ribosomes, see Hoesl and Budisa (2011a).

4.2.8.4 Strains with Synthetic Genetic Codes

The competition of the suppressor $tRNA_{CUA}$ with the release factor RF1 hampers the efficient recoding of amber codons with ncAAs in *E. coli*. Deletion of RF1 eliminates the competition (see Sect. 4.2.8.1) and in an RF1-free strain multiple in-frame amber codons can be efficiently suppressed by an ncAA (Wu et al. 2013; Mukai et al. 2010; Johnson et al. 2011). However, the accidental incorporation of the ncAA into cellular proteins regularly terminating on amber codons can occur more frequently as well. As a consequence, proteins with aberrant sequence lengths accumulate which can eventually poison the cells.

There is a solution to this problem: The amber codon must disappear entirely from the host genome. If all amber codons in the *E. coli* genome are replaced by another stop codon, e.g., the ochre UAA codon, the amber codon is freed from the genetic code. At that point, it represents a real blank codon and can be reintroduced into the amber-less strain with a new meaning.

The groups of George Church and Farren Isaacs successfully mastered this herculean task (Lajoie et al. 2013; Isaacs et al. 2011). The amber stop is the least frequent codon in *E. coli* occurring at the end of only 321 genes (http://www.kazusa.or.jp/codon/) (Lajoie et al. 2013; Nakamura et al. 2000). Recently developed powerful in vivo genome editing technologies such as multiplex automated genome engineering (MAGE) and conjugative assembly genome engineering (CAGE) facilitated the exchange of all 321 known UAG stops by ochre codons (Isaacs et al. 2011). Then, the gene encoding the amber-specific release factor RF1 was deleted in the amber-less strain (Lajoie et al. 2013). Different ncAAs, such as *para*-acetyl-L-phenylalanine (**50**), *para*-azido-L-phenylalanine (**52**), and *O*-phospho-L-serine (**23**), were efficiently incorporated into GFP carrying several in-frame amber codons by using the appropriate o-pairs (Lajoie et al. 2013). The fitness of the new amber-less strain was only slightly impaired in comparison to the wild type. It showed increased resistance toward an infection with bacteriophage T7. The altered genetic code of the amber-less host obviously caused the mistranslation of the phage proteins (Lajoie et al. 2013).

Recently, the attempt to remove the amber codon from the genetic code was extended to a eukaryotic organism. Dymond et al. (2011) described the first partially synthetic yeast chromosomes on which the amber codons had been swapped to opal codons.

4.2.8.5 Sense Codon Recoding

There have been initial attempts to recode sense codons selectively instead of reassigning them in a global fashion as outlined in Sect. 4.2.7. In principle, the reassignment of sense codons should be possible in the cases where more than one codon specifies a cAA. In this way, the degeneracy of the genetic code could be exploited to expand it with ncAAs. However, changing the meaning of a sense codon appears even more challenging than of a stop codon.

At amber stop codons, the release factor RF1 competes with the ncAA-charged suppressor $tRNA_{CUA}$. As outlined above, this competition can be alleviated by the deletion of RF1 (Sect. 4.2.8.1; Mukai et al. 2010; Johnson et al. 2011), and the deletion of RF1 in an amber stop-free strain background truly liberates the amber codon for reassignment (Sect. 4.2.8.4; Lajoie et al. 2013). At a sense codon, however, a tRNA carrying an ncAA must efficiently compete with the cognate aminoacylated tRNA. The deletion of the cognate tRNA would lead to the unintentional incorporation of the ncAA at the respective codon throughout the genome. As was demonstrated for the amber codon, it should be technically feasible to substitute a sense codon by (a) synonymous codon(s) in the genome to "free" it for reassignment. However, the concomitant change in the frequency of the codons in the genome might adversely affect protein expression (Novoa and Ribas De Pouplana 2012).

The group of Tirrell showed that the degeneracy of the genetic code can indeed be exploited to expand the genetic code with an ncAA. Phe is encoded by two codons, UUU and UUC, in the standard genetic code (Fig. 4.1). In *E. coli*, both codons are read by a single $tRNA_{GAA}^{Phe}$. $tRNA_{GAA}^{Phe}$ reads the UUC codon by standard Watson–Crick base pairing between the codon and the anticodon while it decodes the UUU codon by a wobble G–U base pair at the first position of the anticodon, which is the third position of the codon. The wobble base pair is weaker than the regular G–C Watson–Crick base pair. Kwon et al. reasoned that a mutant $tRNA_{AAA}^{Phe}$ would be able to read the UUU codon by Watson–Crick base pairing and would, therefore, decode it faster than the wild-type $tRNA_{GAA}^{Phe}$. They performed the anticodon switch and used the mutant $tRNA_{AAA}^{Phe}$ together with a mutant PheRS with relaxed substrate specificity. They showed that the mutant PheRS/$tRNA_{AAA}^{Phe}$ pair was able to insert L-3-(2-naphthyl)alanine (**59**) specifically at UUU codons (Kwon et al. 2003).

Recently, Krishnakumar et al. took a similar approach to incorporate N^6-cyclopentyloxycarbonyl-L-lysine (**90**) at the Arg CGG codon (Krishnakumar et al. 2013). They switched the anticodon of *Mm*$tRNA_{CUA}^{Pyl}$ to *Mm*$tRNA_{CCG}^{Pyl}$ and used the *Mm*PylRS/*Mm*$tRNA_{CCG}^{Pyl}$ pair with **90** in *Mycoplasma capricolum*. *M. capricolum* uses only a single isoacceptor $tRNA_{ACG}^{Arg}$ to read the four Arg codons, CGU, CGC, CGA, and CGG (Fig. 4.1). However, they observed predominantly Arg incorporation at the CGG codons instead of **90** (Krishnakumar et al. 2013). Obviously, a simple anticodon switch is not (in any case) sufficient to facilitate sense codon recoding. The thorough analysis of the translation machinery with respect to cognate and non-cognate tRNA pools, RNA

modifications, and codon usage will be vital for the further development of the approach (Krishnakumar and Ling 2014).

4.2.8.6 Coupling the Incorporation of ncAAs with Their Biosynthesis

Some ncAAs are quite expensive which can be a severe drawback for their application in protein engineering at larger scales. The biosynthesis of ncAAs from simple precursors could alleviate this flaw. As outlined in Sect. 4.1.2.1, many ncAAs occur naturally, but in the most cases, the biosynthesis pathway is currently unknown. A couple of studies have tackled this issue.

In *E. coli*, Mehl et al. biosynthesized *para*-amino-L-phenylalanine (**54**) from chorismate, an intermediate in the biosynthesis of aromatic amino acids. They introduced three genes, *papABC* from *Streptomyces venezuelae* whose gene products catalyzed the transformation of chorismate to *para*-aminophenylpyruvic acid. The transamination of *para*-aminophenylpyruvic acid to **54** was catalyzed by a host enzyme, most probably by a nonspecific tyrosine aminotransferase from *E. coli* (Mehl et al. 2003). The biosynthesized **54** was then incorporated into a target protein in response to an amber stop codon using a specific orthogonal pair (Mehl et al. 2003). They obtained roughly 2 mg/L of the synthetic protein.

The group of Budisa exploited the endogenous tryptophan synthase of *E. coli* for the biotransformation of indole derivatives to the corresponding tryptophan analogs (Lepthien et al. 2008). They supplemented a Trp-auxotrophic *E. coli* host with 4-azaindole and 5-azaindole and subsequently the biosynthesized 4-aza-L-tryptophan (**84**) and 5-aza-L-tryptophan (**85**), respectively, were incorporated into a target protein in the residue-specific manner (Lepthien et al. 2008).

Recently, the same group achieved the biosynthesis of the reactive amino acid L-azidohomoalanine (**28**) from *O*-acetyl-L-homoserine with azide as the substrate nucleophile. They expressed *O*-acetylhomoserine sulfhydrylase (OAHSS) from *Corynebacterium glutamicum* (Ma et al. 2014) in *E. coli*. The enzyme accepts a number of substrate nucleophiles such as thiols, selenols, and cyanide as well as azide. Since *E. coli* is not capable of biosynthesizing the precursor *O*-acetyl-L-homoserine, it was supplemented in the medium together with sodium azide. The biosynthesized **28** was residue-specifically incorporated at the single Met position of the model protein barstar (Ma et al. 2014). Well over 5 mg/L of labeled protein were produced.

4.2.8.7 Application Examples

Since the initial attempts to site-specifically incorporate *para*-fluoro-L-phenylalanine (**47**) and *O*-methyl-L-tyrosine (**38**) into target proteins by Furter (1998) and the Schultz group (Wang et al. 2001), respectively, ~100 ncAAs have been successfully incorporated using o-pairs (Wan et al. 2014; Liu and Schultz 2010). With some occasional exceptions such as Cys, Met, and Ala analogs (Wu et al. 2004; Ai et al. 2010), derivatives of phenylalanine or tyrosine and lysine were mainly used together with o-pairs that were evolved from TyrRS/$tRNA_{sup}$ or PylRS/$tRNA_{sup}$ pairs. To list all the successful o-pairs and their application would go far beyond the scope of this study. Again, I will highlight prominent examples and the interested

reader is referred to the literature for a more detailed review (Wan et al. 2014; Liu and Schultz 2010).

Site-Specific Incorporation of Reactive ncAAs

Amino acids possessing a reactive functionality provide a chemical handle to conjugate a protein with a ligand or to immobilize it to a solid support. As described in Sect. 4.2.7.2, some reactive ncAAs, mainly Met analogs, can be incorporated into proteins in a residue-specific manner. This approach is particularly valuable to study newly synthesized proteins in whole proteomes. To attach an artificial posttranslational modification at a predefined position, it is more convenient to install the reactive handle site-specifically (Kim et al. 2013). The site-specific incorporation of reactive ncAAs was exploited for the PEGylation of protein therapeutics, to synthesize antibody–drug conjugates and bispecific antibodies, to conjugate proteins with DNA or peptide nucleic acids, and to decorate proteins with fluorescent dyes or spectroscopic probes (reviewed in Kim et al. 2013).

Residue-specifically incorporated L-azidohomoalanine (**28**) or L-homopropargylglycine (**30**) moieties are usually modified using "click chemistry" (CuAAC, Sect. 4.2.7.2). The palette of orthogonal chemical conjugation reactions can be vastly expanded, though, if the appropriate o-pairs are used.

NcAAs bearing a carbonyl group in their side chain react with hydrazines or alkoxyamines to form stable hydrazones or oximes, respectively (Dirksen and Dawson 2008). Terminal azido groups react with terminal alkynes in the presence of Cu(I) as the catalyst (CuAAC). Strain-promoted Huisgen 1,3-dipolar cycloadditions between azides and cyclooctynes (SpAAC) and, most recently, inverse electron-demand Diels–Alder (IEDDA) reactions between electron-deficient tetrazines and strained alkenes have been established (Jewett and Bertozzi 2010). The site-specific incorporation of other reactive ncAAs into target proteins facilitates the selective modification by Michael reactions (Wang et al. 2007b), olefin metathesis (Ai et al. 2010), transition metal-catalyzed cross-couplings (Brustad et al. 2008), radical polymerization (Peeler et al. 2010), oxidative couplings (Behrens et al. 2011), acyl-transfer reactions (Nguyen et al. 2011), and photoclick reactions (Yu et al. 2012).

Para-acetyl-L-phenylalanine (**50**) has been extensively used for the conjugation to hydrazine ligands (Kim et al. 2013; Wang et al. 2003a). However, the acidic reaction conditions at a $pH < 5$ are unsuitable for acid-labile proteins and living cells (Kim et al. 2013). Cho et al. functionalized a human growth hormone with **50** and, subsequently, conjugated the carbonyl groups to aminooxy-polyethylene glycol (PEG). To produce sufficient amounts of the PEGylated hormone for clinical trials, they performed the labeling with **50** in a 1000 L bioreactor (Cho et al. 2011). This report clearly indicates that the site-specific incorporation of ncAAs using o-pairs is scalable despite the moderate product titers (see Sect. 4.2.8.1).

The genetic code of *E. coli* was expanded with the photoreactive Phe derivative *para*-benzoyl-L-phenylalanine (**51**) to map protein–protein interactions (Chin et al. 2002a). Another photoreactive ncAA, *para*-azido-L-phenylalanine (**52**), was added to the genetic code of bacteria (Chin et al. 2002b), yeasts (Chin et al. 2003a;

Young et al. 2009; Deiters et al. 2003), and mammalian cells (Liu et al. 2007). **52** is not only a reactive photo-cross-linker but it can be conjugated to alkyne ligands by "click chemistry" (CuAAC) or to phosphine ligands by the Staudinger ligation (Tsao et al. 2005). The alkyne pendant of **52**, *para*-propargyloxy-L-phenylalanine (**53**), was also introduced into proteins by amber stop suppression for the bioorthogonal conjugation, e.g., with fluorophores (Young et al. 2009; Liu et al. 2007; Deiters et al. 2003; Deiters and Schultz 2005).

The group of Carell incorporated the unsaturated Pyl derivatives N^6-[(allyloxy) carbonyl]lysine (**88**, alkene derivative of Lys/Pyl) and N^6-[(propargyloxy)carbonyl] lysine (**89**, alkyne derivative of Lys/Pyl) at multiple positions in yellow fluorescent protein using the wild-type PylRS/$tRNA_{CUA}^{Pyl}$ pair from *Methanosarcina mazei*. The alkyne-functionalized yellow fluorescent protein was artificially glycosylated with different azido sugars (Kaya et al. 2009).

The Cu(I)-catalyzed click reaction of azido groups with alkyne moieties generates a stable triazole linkage. However, the Cu(I) catalyst can cause undesirable side reactions and it is toxic to living cells (Kim et al. 2013). To alleviate this drawback, Plass et al. used a mutant PylRS/$tRNA_{CUA}$ o-pair to genetically incorporate Pyl derivatives carrying side chains with ring-constrained cyclooctynes, such as compound **93.** SpAAC of azides with cyclooctynes allows for metal-free "click chemistry" (Plass et al. 2011). However, the bulky structure of the cyclooctyne ncAAs and their hydrophobic side chains can affect protein structure and folding, an additional disadvantage to the complex synthesis and short shelf life of these compounds at ambient temperature (Kim et al. 2013).

Electron-deficient tetrazines react selectively with strained alkenes, e.g., norbornene-, *trans*-cyclooctene and cyclopropene groups, or ring-constrained alkynes. During the IEDDA reaction, a stable bicyclic linkage is formed between the reactants. The reactions are particularly attractive for in vivo conjugations since they are completed within minutes under physiological conditions and they do not demand a catalyst (Kim et al. 2013). Norbornene derivatives of Pyl, such as compound **94**, were successfully used for bioorthogonal conjugation with electron-deficient tetrazines to label proteins site-specifically in vitro, in *E. coli*, and in live mammalian cells (Kaya et al. 2012; Lang et al. 2012b). Lang et al. performed amber stop codon suppression with a *trans*-cyclooctene-bearing ncAA (**97**) and bicyclic ncAAs (**96**, **98**). They successfully conjugated these dienophiles with tetrazine derivatives in vitro, in *E. coli*, or in living mammalian cells (Lang et al. 2012a).

Mehl and co-workers established the same conjugation reaction in a reversed configuration: They incorporated a tetrazine derivative of Phe (**62**) into a target protein and selectively conjugated the residue with *trans*-cyclooctenes (Seitchik et al. 2012).

Yu et al. devised a mutant PylRS/$tRNA_{CUA}$ pair that is specific for a Pyl derivative with a ring-constrained cyclopropene side chain (compound **95**). They demonstrated the site-specific incorporation of **95** into target proteins in *E. coli* and mammalian cells for the subsequent "photoclick" reaction with tetrazoles (Yu et al. 2012).

Behrens and co-workers inserted *para*-amino-L-phenylalanine (**54**) into viral capsids by amber stop codon suppression and conjugated it to *ortho*-aminophenol derivatives, e.g., of PEG by oxidative coupling. The conjugation reaction proceeded fast at near-physiological pH in aqueous solution (Behrens et al. 2011).

Wang et al. demonstrated the Michael addition of thiol derivatives to genetically incorporated dehydroalanine (**102**). However, dehydroalanine *per se* is not stable in aqueous milieu. Therefore, they incorporated a stable precursor, L-phenylselenocysteine (**101**), by amber SCS. Treatment with H_2O_2 converted the precursor to the reactive dehydroalanine (Wang et al. 2007b).

The insertion of L-*O*-crotylserine (**104**) at two spatially adjacent positions by amber suppression facilitated the formation of an olefin bridge *via* an intramolecular metathesis reaction (Ai et al. 2010). Brustad et al. (2008) employed a genetically incorporated boronate amino acid (**57**) for transition metal-mediated cross-couplings, and 1,2-aminothiol derivatives of Pyl, such as compound **91**, were used for acyl transfer by cyanobenzothiazole condensation (Nguyen et al. 2011). The group of Mehl encoded L-4-(2′-bromoisobutyramido)phenylalanine (**56**) by amber codons. They showed that **56** is a suitable initiator of radical polymerization from which polymers can be grown to produce protein–polymer conjugates (Peeler et al. 2010).

Site-Specific Incorporation of ncAAs to Probe Protein Structure, Dynamics, and Function

NcAAs can serve as probes for structure–function studies. Site-specifically inserted ncAAs have been used to probe protein structure, dynamics, and folding, as well as protein–protein interactions and ligand binding [reviewed in Jones et al. (2010)]. Isotope-labeled *O*-methyl-L-tyrosine (**39**) and polyfluorinated ncAAs, e.g., *O*-trifluoromethyl-L-tyrosine (**40**) and *para*-trifluoromethyl-L-phenylalanine (**58**), were used for NMR studies. Photocaged ncAAs, such as L-*O*-(2-nitrobenzyl)tyrosine (**42**), L-*S*-(2-nitrobenzyl)cysteine (**103**), or a nitrobenzyl derivative of Lys (**92**) as well as L-*O*-(4,5-dimethoxy-2-nitrobenzyl)serine (**105**), facilitate the spatiotemporal control of biological activity by light. The photocaging 2-nitrobenzyl- or 4,5-dimethoxy-2-nitrobenzyl groups can be cleaved by the irradiation with UV or blue light, respectively, converting the amino acids from an inactive (ncAA) to an active (cAA) form (Deiters et al. 2006). Metal-chelating ncAAs, e.g., L-(2,2′-bipyridin-5-yl)alanine (**61**) and L-3-(8-hydroxyquinolin-3-yl)alanine (**60**), can be used as biophysical probes and for the structure determination of proteins (see Jones et al. (2010) and references therein). As well, genetically incorporated *para*-iodo-L-phenylalanine (**48**) (Xie et al. 2004) and *meta*-iodo-L-tyrosine (**37**) (Sakamoto et al. 2009) were successfully employed for protein structure elucidation.

A number of Trp analogs were incorporated into fluorescent proteins in a residue-specific manner to engineer their spectral properties (see Sect. 4.2.7.2). Kwon et al. selectively exchanged the Trp residue at position 66 in the fluorophore of cyan fluorescent protein (CFP) with 6-bromo-L-tryptophan (**77**), 6-chloro-L-tryptophan (**78**), and L-3-(benzothien-3-yl)alanine (**72**). These Trp analogs are not

recognized by the *E. coli* TrpRS. Therefore, they used the yPheRS(T415G)/ytRNA$_{CUA_UG}^{Phe}$ o-pair from *S. cerevisiae* to suppress an in-frame amber stop codon at the indicated fluorophore position with the different Trp analogs. The resulting CFP variants showed blue-shifted fluorescence emission and absorption maxima (Kwon and Tirrell 2007).

Wang and co-workers generated spectral variants of GFP by the site-specific incorporation of the following tyrosine analogs at position Tyr66 in the fluorophore (Wang et al. 2003b): *para*-amino-L-phenylalanine (**54**), *O*-methyl-L-tyrosine (**38**), *para*-iodo-L-phenylalanine (**48**), *para*-bromo-L-phenylalanine (**49**), and L-3-(2-naphthyl)alanine (**59**).

Manipulation of Enzyme Activity by the Site-Specific Incorporation of ncAAs

So far, only a few studies have addressed the manipulation of enzyme activity by site-specific incorporation of ncAAs. For instance, the selective insertion of *para*-amino-L-phenylalanine (**54**) and *para*-nitro-L-phenylalanine (**55**) at Phe124 in the active site of the nitroreductase from *E. coli* enhanced the catalytic activity of the enzyme (Jackson et al. 2006).

The regioselectivity and catalytic efficiency of the cytochrome P450 enzyme, long-chain fatty acid monooxygenase CYP102A1 (P450 BM-3) from *Bacillus megaterium*, were improved by the incorporation of the aromatic ncAAs, **54**, *para*-acetyl-L-phenylalanine (**50**), L-*O*-benzyltyrosine (**43**), and L-3-(2-naphthyl) alanine (**59**), at different positions of the active site. The functional improvements introduced by the ncAAs could not be reproduced by mutagenesis with the 20 cAAs (Kolev et al. 2014), similar to the observations of Ugwumba et al. (2010).

Wu et al. introduced the photocaged Cys analog L-*S*-(2-nitrobenzyl)cysteine (**103**) at the catalytic Cys164 residue of caspase 3, which allowed the photoregulation of the enzyme activity (Wu et al. 2004).

Other Applications of the Site-Specific Incorporation of ncAAs

Alfonta et al. demonstrated the successful site-specific incorporation of 3,4-dihydroxy-L-phenylalanine (**44**) at amber codons (Alfonta et al. 2003). **44** is a redox-active amino acid and can be used to engineer synthetic redox proteins.

Liu and co-workers devised a phage display system that includes ncAAs in addition to cAAs for library diversification. They successfully used this system with L-*O*-sulfotyrosine (**41**), *para*-acetyl-L-phenylalanine (**50**), L-(2,2′-bipyridin-5-yl) alanine (**61**), or *para*-borono-L-phenylalanine (**57**) for the directed evolution of antibodies (Liu et al. 2008).

Neumann et al. employed L-N^6-acetyllysine (**87**) to mimic the acetylation of lysine residues in histones (Neumann et al. 2009). They evolved a PylRS/tRNA$_{CUA}$ o-pair for the incorporation of **87** at amber stop codons.

Söll and co-workers adapted the naturally occurring SepRS (see Sect. 4.2.4.3) for the incorporation of *O*-phospho-L-serine (Sep, **23**) by amber SCS in *E. coli* (Park et al. 2011). They mutated the anticodon of the *Methanocaldococcus jannaschii* tRNA^{Cys} to generate an amber suppressor *Mj*tRNA$_{CUA}^{Sep}$. To charge *Mj*tRNA$_{CUA}^{Sep}$ with **23**, they employed the SepRS from *Methanococcus*

maripaludis (*Mmp*SepRS). A mutant elongation factor EF-Tu, which effectively bound the *Mj*tRNA$_{CUA}^{Sep}$ aminoacylated with **23**, complemented the *Mmp*SepRS/*Mj*tRNA$_{CUA}^{Sep}$ o-pair (Park et al. 2011).

Combining Residue- and Site-Specific Incorporation of ncAAs

Budisa and co-workers first showed that the site-specific insertion of an ncAA into a target protein can be combined with the residue-specific incorporation of another (Hoesl and Budisa 2011b). They inserted the photo-cross-linker *para*-benzoyl-L-phenylalanine (**51**) at a permissive site in their model proteins, GFP, and the lipase TTL from *Thermoanaerobacter thermohydrosulfuricus*, using the appropriate o-pair for amber SCS. Simultaneously, they exchanged all Pro residues of GFP for (2*S*,4*S*)-4-fluoro-L-proline (**31**) and all Met residues of the lipase for norleucine (**26**) (Hoesl and Budisa 2011b).

Deepankumar et al. applied the combined residue- and site-specific incorporation technique to immobilize the omega-transaminase from *Sphaerobacter thermophilus* on a solid support and, at the same time, to enhance the stability of the enzyme. They used a specific o-pair to incorporate 3,4-dihydroxy-L-phenylalanine (**44**) at an in-frame amber codon and (2*S*,4*R*)-4-fluoro-L-proline (**32**) was incorporated residue-specifically. Subsequently, they coupled the functionalized enzyme to carrier beads via the reactive **44** residues. The immobilized enzyme showed enhanced thermostability and excellent reusability (Deepankumar et al. 2015).

Acknowledgments This work has been supported by the Federal Ministry of Science, Research and Economy (BMWFW); the Federal Ministry of Traffic, Innovation and Technology (bmvit); the Styrian Business Promotion Agency SFG; the Standortagentur Tirol; the government of Lower Austria; and ZIT—Technology Agency of the City of Vienna through the COMET Funding Program managed by the Austrian Research Promotion Agency FFG.

I am grateful for support by CHEM21 in the frame of the Innovative Medicines Initiative Joint Undertaking under grant agreement n°115360, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution.

References

Acevedo-Rocha CG, Hoesl MG, Nehring S et al (2013) Non-canonical amino acids as a useful synthetic biological tool for lipase-catalysed reactions in hostile environments. Catal Sci Technol 3(5):1198–1201

Ai H-W, Shen W, Brustad E, Schultz PG (2010) Genetically encoded alkenes in yeast. Angew Chem Int Ed Engl 49(5):935–937

Albayrak C, Swartz JR (2014) Direct polymerization of proteins. ACS Synth Biol 3(6):353–362

Aldag C, Gromov IA, García-Rubio I et al (2009) Probing the role of the proximal heme ligand in cytochrome P450cam by recombinant incorporation of selenocysteine. Proc Natl Acad Sci USA 106(14):5481–5486

Aldag C, Bröcker MJ, Hohn MJ et al (2013) Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. Angew Chem Int Ed Engl 52(5):1441–1445

Alfonta L, Zhang Z, Uryu S, Loo JA, Schultz PG (2003) Site-specific incorporation of a redox-active amino acid into proteins. J Am Chem Soc 125(48):14662–14663
Alix JH (1982) Molecular aspects of the in vivo and in vitro effects of ethionine, an analog of methionine. Microbiol Rev 46(3):281–295
Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. Nat Chem Biol 3(1):29–35
Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG (2004) An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci USA 101(20):7566–7571
Asatoor AM (1969) The occurrence of εN-methyllysine in human urine. Clin Chim Acta 26 (1):147–154
Bae JH, Alefelder S, Kaiser JT et al (2001) Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. J Mol Biol 309 (4):925–936
Bae JH, Rubini M, Jung G et al (2003) Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. J Mol Biol 328(5):1071–1081
Bagert JD, Xie YJ, Sweredoski MJ et al (2014) Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. Mol Cell Proteomics 13(5):1352–1358
Baker PJ, Montclare JK (2011) Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase. ChemBioChem 12(12):1845–1848
Barrett GC, Elmore DT (1998) Biological roles of amino acids and peptides. In: Amino acids and peptides. Cambridge University Press, Cambridge, pp 174–199
Beatty KE, Tirrell DA (2008) Two-color labeling of temporally defined protein populations in mammalian cells. Bioorg Med Chem Lett 18(22):5995–5999
Beatty KE, Xie F, Wang Q, Tirrell DA (2005) Selective dye-labeling of newly synthesized proteins in bacterial cells. J Am Chem Soc 127(41):14150–14151
Beatty KE, Liu JC, Xie F et al (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells. Angew Chem Int Ed Engl 45(44):7364–7367
Beatty KE, Fisk JD, Smart BP et al (2010) Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. ChemBioChem 11(15):2092–2095
Behrens CR, Hooker JM, Obermeyer AC, Romanini DW, Katz EM, Francis MB (2011) Rapid chemoselective bioconjugation through oxidative coupling of anilines and aminophenols. J Am Chem Soc 133(41):16398–16401
Beiboer SHW, Berg BVD, Dekker N, Cox RC, Verheij HM (1996) Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A2. Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH. Protein Eng 9 (4):345–352
Bessho Y, Hodgson DRW, Suga H (2002) A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. Nat Biotechnol 20(7):723–728
Bianco A, Townsley FM, Greiss S, Lang K, Chin JW (2012) Expanding the genetic code of Drosophila melanogaster. Nat Chem Biol 8(9):748–750
Bischoff R, Schlüter H (2012) Amino acids: chemistry, functionality and selected non-enzymatic post-translational modifications. J Proteom 75(8):2275–2296
Blight SK, Larue RC, Mahapatra A et al (2004) Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. Nature 431(7006):333–335
Böck A, Forchhammer K, Heider J, Baron C (1991) Selenoprotein synthesis: an expansion of the genetic code. Trends Biochem Sci 16:463–467
Bogosian G, Violand BN, Dorward-King EJ, Workman WE, Jung PE, Kane JF (1989) Biosynthesis and incorporation into protein of norleucine by Escherichia coli. J Biol Chem 264(1):531–539
Brenner S, Stretton AOW, Kaplan S (1965) Genetic code: the 'nonsense' triplets for chain termination and their suppression. Nature 206(4988):994–998

Brustad E, Bushey ML, Lee JW, Groff D, Liu W, Schultz PG (2008) A genetically encoded boronate-containing amino acid. Angew Chem Int Ed Engl 47(43):8220–8223
Budisa N (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. Angew Chem Int Ed Engl 43(47):6426–6463
Budisa N, Pal PP (2004) Designing novel spectral classes of proteins with a tryptophan-expanded genetic code. Biol Chem 385(10):893–904
Budisa N, Pal PP, Alefelder S et al (2004) Probing the role of tryptophans in Aequorea victoria green fluorescent proteins with an expanded genetic code. Biol Chem 385(2):191–202
Budisa N, Wenger W, Wiltschi B (2010) Residue-specific global fluorination of Candida antarctica lipase B in Pichia pastoris. Mol Biosyst 6(9):1630–1639
Burk RF, Hill KE (2005) Selenoprotein P: an extracellular protein with unique physical characteristics and a role in selenium homeostasis. Annu Rev Nutr 25(1):215–235
Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J Bacteriol 192(19):5143–5150
Carrico IS, Maskarinec SA, Heilshorn SC et al (2007) Lithographic patterning of photoreactive cell-adhesive proteins. J Am Chem Soc 129(16):4874–4875
Chatterjee A, Xiao H, Schultz PG (2012) Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. Proc Natl Acad Sci USA 109(37):14841–14846
Chilton WS, Tsou G (1972) A chloro amino acid from Amanita solitaria. Phytochemistry 11 (9):2853–2857
Chin JW, Martin AB, King DS, Wang L, Schultz PG (2002a) Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proc Natl Acad Sci USA 99 (17):11020–11024
Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG (2002b) Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc 124(31):9026–9027
Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG (2003a) An expanded eukaryotic genetic code. Science 301(5635):964–967
Chin JW, Cropp TA, Chu S, Meggers E, Schultz PG (2003b) Progress toward an expanded eukaryotic genetic code. Chem Biol 10(6):511–519
Cho H, Daniel T, Buechler YJ et al (2011) Optimized clinical performance of growth hormone with an expanded genetic code. Proc Natl Acad Sci USA 108(22):9060–9065
Cirino PC, Tang Y, Takahashi K, Tirrell DA, Arnold FH (2003) Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. Biotechnol Bioeng 83(6):729–734
Cobucci-Ponzano B, Rossi M, Moracci M (2012) Translational recoding in archaea. Extremophiles 16(6):793–803
Cohen GN, Cowie DB (1957) Total replacement of methionine by selenomethionine in the proteins of Escherichia coli. C R Hebd Seances Acad Sci 244(5):680–683
Commans S, Böck A (1999) Selenocysteine inserting tRNAs: an overview. FEMS Microbiol Rev 23(3):335–351
Cowie DB, Cohen GN (1957) Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. Biochim Biophys Acta 26(2):252–261
Crespo MD, Rubini M (2011) Rational design of protein stability: effect of (2S,4R)-4-fluoroproline on the stability and folding pathway of ubiquitin. PLoS One 6(5):e19425
Crick FHC (1968) The origin of the genetic code. J Mol Biol 38(3):367–379
Datta D, Wang P, Carrico IS, Mayo SL, Tirrell DA (2002) A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. J Am Chem Soc 124(20):5652–5653
Deepankumar K, Shon M, Nadarajan SP et al (2014) Enhancing thermostability and organic solvent tolerance of ω-transaminase through global incorporation of fluorotyrosine. Adv Synth Catal 356(5):993–998

Deepankumar K, Nadarajan SP, Mathew S et al (2015) Engineering transaminase for stability enhancement and site-specific immobilization through multiple noncanonical amino acids incorporation. ChemCatChem 7(3):417–421
Deiters A, Schultz PG (2005) In vivo incorporation of an alkyne into proteins in Escherichia coli. Bioorg Med Chem Lett 15(5):1521–1524
Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, Schultz PG (2003) Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. J Am Chem Soc 125(39):11782–11783
Deiters A, Groff D, Ryu Y, Xie J, Schultz PG (2006) A genetically encoded photocaged tyrosine. Angew Chem Int Ed Engl 45(17):2728–2731
Di Giulio M (2005) The origin of the genetic code: theories and their relationships, a review. Biosystems 80(2):175–184
Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci USA 103(25):9482–9487
Dieterich DC, Hodas JJL, Gouzer G et al (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. Nat Neurosci 13(7):897–905
Dirksen A, Dawson PE (2008) Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. Bioconjug Chem 19(12):2543–2548
Dougherty DA, Van Arnam EB (2014) In vivo incorporation of non-canonical amino acids by using the chemical aminoacylation strategy: a broadly applicable mechanistic tool. ChemBioChem 15(12):1710–1720
Dymond JS, Richardson SM, Coombes CE et al (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477(7365):471–476
Edwardraja S, Sriram S, Govindan R, Budisa N, Lee S-G (2011) Enhancing the thermal stability of a single-chain Fv fragment by in vivo global fluorination of the proline residues. Mol Biosyst 7(1):258–265
Eichelbaum K, Winter M, Diaz MB, Herzig S, Krijgsveld J (2012) Selective enrichment of newly synthesized proteins for quantitative secretome analysis. Nat Biotechnol 30(10):984–990
Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. Annu Rev Microbiol 58(1):453–488
Fisher JF, Mallette MF (1961) The natural occurrence of ethionine in bacteria. J Gen Physiol 45(1):1–13
Fotheringham IG, Grinter N, Pantaleone DP, Senkpeil RF, Taylor PP (1999) Engineering of a novel biochemical pathway for the biosynthesis of L-2-aminobutyric acid in Escherichia coli K12. Bioorg Med Chem 7(10):2209–2213
Fowden L (1972) Amino acid complement of plants. Phytochemistry 11(7):2271–2276
Fowden L, Lewis D, Tristram H (1967) Toxic amino acids: their action as antimetabolites. In: Nord FF (ed) Advances in enzymology and related areas of molecular biology. Wiley, New York, pp 89–163
Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. Protein Sci 7(2):419–426
Gaston MA, Jiang R, Krzycki JA (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr Opin Microbiol 14(3):342–349
Giege R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 26(22):5017–5035
Goerke AR, Swartz JR (2009) High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. Biotechnol Bioeng 102(2):400–416
Grammel M, Zhang MM, Hang HC (2010) Orthogonal alkynyl amino acid reporter for selective labeling of bacterial proteomes during infection. Angew Chem Int Ed Engl 49(34):5970–5974
Greiss S, Chin JW (2011) Expanding the genetic code of an animal. J Am Chem Soc 133 (36):14196–14199

Hatanaka S-I (1969) A new amino acid isolated from Morchella esculenta and related species. Phytochemistry 8(7):1305–1308
Hendrickson WA, Horton JR, LeMaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J 9(5):1665–1672
Hinz FI, Dieterich DC, Tirrell DA, Schuman EM (2012) Non-canonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. ACS Chem Neurosci 3(1):40–49
Hinz FI, Dieterich DC, Schuman EM (2013) Teaching old NCATs new tricks: using non-canonical amino acid tagging to study neuronal plasticity. Curr Opin Chem Biol 17(5):738–746
Hodas JJL, Nehring A, Höche N et al (2012) Dopaminergic modulation of the hippocampal neuropil proteome identified by bio-orthogonal non-canonical amino-acid tagging (BONCAT). Proteomics 12(15-16):2464–2476
Hoesl MG, Budisa N (2011a) In vivo incorporation of multiple noncanonical amino acids into proteins. Angew Chem Int Ed Engl 50(13):2896–2902
Hoesl MG, Budisa N (2011b) Expanding and engineering the genetic code in a single expression experiment. ChemBioChem 12(4):552–555
Hoesl MG, Acevedo-Rocha CG, Nehring S et al (2011) Lipase congeners designed by genetic code engineering. ChemCatChem 3(1):213–221
Holzberger B, Marx A (2010) Replacing 32 proline residues by a noncanonical amino acid results in a highly active DNA polymerase. J Am Chem Soc 132(44):15708–15713
Hong SH, Kwon Y-C, Jewett MC (2014) Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. Front Chem 2 (eCollection 2014)
Howden AJM, Geoghegan V, Katsch K et al (2013) QuaNCAT: quantitating proteome dynamics in primary cells. Nat Methods 10(4):343–346
Hunt S (1985) The non-protein amino acids. In: Barrett G (ed) Chemistry and biochemistry of the amino acids. Chapman and Hall, London, pp 55–137
Ibba M, Söll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69(1):617–650
Isaacs FJ, Carr PA, Wang HH et al (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333(6040):348–353
Jackson JC, Duffy SP, Hess KR, Mehl RA (2006) Improving nature's enzyme active site with genetically encoded unnatural amino acids. J Am Chem Soc 128(34):11124–11127
Jewett JC, Bertozzi CR (2010) Cu-free click cycloaddition reactions in chemical biology. Chem Soc Rev 39(4):1272–1279
Johnson DBF, Xu J, Shen Z et al (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7(11):779–786
Johnson DBF, Wang C, Xu J et al (2012) Release factor one is nonessential in Escherichia coli. ACS Chem Biol 7(8):1337–1344
Jones D, Cellitti S, Hao X et al (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. J Biomol NMR 46(1):89–100
Jung MJ (1985) Enzyme inhibition by amino acids and their derivatives. In: Barrett GC (ed) Chemistry and biochemistry of the amino acids. Chapman and Hall, London, pp 227–245
Kast P, Hennecke H (1991) Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. J Mol Biol 222(1):99–124
Kaya E, Gutsmiedl K, Vrabel M, Müller M, Thumbs P, Carell T (2009) Synthesis of threefold glycosylated proteins using click chemistry and genetically encoded unnatural amino acids. ChemBioChem 10(18):2858–2861
Kaya E, Vrabel M, Deiml C, Prill S, Fluxa VS, Carell T (2012) A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. Angew Chem Int Ed Engl 51(18):4466–4469
Kean EA, Lewis CE (1981) Biosynthesis of L-β-(methylenecyclopropyl)-alanine (hypoglycin) in Blighia sapida. Phytochemistry 20(9):2161–2164

Kiick KL, van Hest JCM, Tirrell DA (2000) Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. Angew Chem Int Ed Engl 39(12):2148–2152

Kiick KL, Weberskirch R, Tirrell DA (2001) Identification of an expanded set of translationally active methionine analogues in Escherichia coli. FEBS Lett 502(1-2):25–30

Kiick KL, Saxon E, Tirrell DA, Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc Natl Acad Sci USA 99(1):19–24

Kim S, Paik WK (1965) Studies on the origin of epsilon-N-methyl-L-lysine in protein. J Biol Chem 240(12):4629–4634

Kim W, George A, Evans M, Conticello VP (2004) Cotranslational incorporation of a structurally diverse series of proline analogues in an Escherichia coli expression system. ChemBioChem 5(7):928–936

Kim CH, Axup JY, Schultz PG (2013) Protein conjugation with genetically encoded unnatural amino acids. Curr Opin Chem Biol 17(3):412–419

Kirshenbaum K, Carrico IS, Tirrell DA (2002) Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. ChemBioChem 3(2-3):235–237

Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2(1):49–58

Kodama K, Nakayama H, Sakamoto K et al (2010) Site-specific incorporation of 4-iodo-L-phenylalanine through opal suppression. J Biochem 148(2):179–187

Kolev JN, Zaengle JM, Ravikumar R, Fasan R (2014) Enhancing the efficiency and regioselectivity of P450 oxidation catalysts by unnatural amino acid mutagenesis. ChemBioChem 15(7):1001–1010

Krishnakumar R, Ling J (2014) Experimental challenges of sense codon reassignment: an innovative approach to genetic code expansion. FEBS Lett 588(3):383–388

Krishnakumar R, Prat L, Aerni H-R et al (2013) Transfer RNA misidentification scrambles sense codon recoding. ChemBioChem 14(15):1967–1972

Krzycki JA (2005) The direct genetic encoding of pyrrolysine. Curr Opin Microbiol 8(6):706–712

Kuhn SM, Rubini M, Fuhrmann M, Theobald I, Skerra A (2010) Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting. J Mol Biol 404(1):70–87

Kwon I, Lim SI (2015) Tailoring the substrate specificity of yeast phenylalanyl-tRNA synthetase toward a phenylalanine analog using multiple-site-specific incorporation. ACS Synth Biol 4(5):634–643

Kwon I, Tirrell DA (2007) Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. J Am Chem Soc 129(34):10431–10437

Kwon I, Kirshenbaum K, Tirrell DA (2003) Breaking the degeneracy of the genetic code. J Am Chem Soc 125(25):7512–7513

Labunskyy VM, Hatfield DL, Gladyshev VN (2014) Selenoproteins: molecular pathways and physiological roles. Physiol Rev 94(3):739–777

Lajoie MJ, Rovner AJ, Goodman DB et al (2013) Genomically recoded organisms expand biological functions. Science 342(6156):357–360

Landgraf P, Antileo E, Schuman E, Dieterich D (2015) BONCAT: metabolic labeling, click chemistry, and affinity purification of newly synthesized proteomes. Methods Mol Biol 1266:199–215

Lang K, Davis L, Wallace S et al (2012a) Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels–Alder reactions. J Am Chem Soc 134(25):10317–10320

Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW (2012b) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nat Chem 4(4):298–304

Lenstra R (2014) Evolution of the genetic code through progressive symmetry breaking. J Theor Biol 347:95–108
Lepthien S, Hoesl MG, Merkel L, Budisa N (2008) Azatryptophans endow proteins with intrinsic blue fluorescence. Proc Natl Acad Sci USA 105(42):16095–16100
Lepthien S, Merkel L, Budisa N (2010) In vivo double and triple labeling of proteins using synthetic amino acids. Angew Chem Int Ed Engl 49(32):5446–5450
Letendre CH, Dickens G, Guroff G (1974) The tryptophan hydroxylase of Chromobacterium violaceum. J Biol Chem 249(22):7186–7191
Ling J, Reynolds N, Ibba M (2009) Aminoacyl-tRNA synthesis and translational quality control. Annu Rev Microbiol 63(1):61–78
Link AJ, Tirrell DA (2003) Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. J Am Chem Soc 125(37):11164–11165
Link AJ, Vink MKS, Tirrell DA (2004) Presentation and detection of azide functionality in bacterial cell surface proteins. J Am Chem Soc 126(34):10598–10602
Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79 (1):413–444
Liu W, Brock A, Chen S, Chen S, Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Methods 4(3):239–244
Liu CC, Mack AV, Tsao M-L et al (2008) Protein evolution with an expanded genetic code. Proc Natl Acad Sci USA 105(46):17688–17693
Luesch H, Hoffmann D, Hevel JM, Becker JE, Golakoti T, Moore RE (2002) Biosynthesis of 4-methylproline in cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities. J Org Chem 68(1):83–91
Ma Y, Biava H, Contestabile R, Budisa N, Di Salvo M (2014) Coupling bioorthogonal chemistries with artificial metabolism: intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. Molecules 19(1):1004–1022
Mehl RA, Anderson JC, Santoro SW et al (2003) Generation of a bacterium with a 21 amino acid genetic code. J Am Chem Soc 125(4):935–939
Mehta KR, Yang CY, Montclare JK (2011) Modulating substrate specificity of histone acetyltransferase with unnatural amino acids. Mol Biosyst 7(11):3050–3055
Meldal M, Tornøe CW (2008) Cu-catalyzed azide-alkyne cycloaddition. Chem Rev 108 (8):2952–3015
Merkel L, Schauer M, Antranikian G, Budisa N (2010) Parallel incorporation of different fluorinated amino acids: on the way to "teflon" proteins. ChemBioChem 11(11):1505–1507
Milner-White EJ, Bell LH, Maccallum PH (1992) Pyrrolidine ring puckering in cis and trans-proline residues in proteins and polypeptides: different puckers are favoured in certain situations. J Mol Biol 228(3):725–734
Minks C, Huber R, Moroder L, Budisa N (2000) Noninvasive tracing of recombinant proteins with "fluorophenylalanine-fingers". Anal Biochem 284(1):29–34
Minnihan EC, Young DD, Schultz PG, Stubbe J (2011) Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. J Am Chem Soc 133(40):15942–15945
Moghal A, Mohler K, Ibba M (2014) Mistranslation of the genetic code. FEBS Lett 588 (23):4305–4310
Montclare JK, Tirrell DA (2006) Evolving proteins of novel composition. Angew Chem Int Ed Engl 45(27):4518–4521
Moore PB, Steitz TA (2011) The roles of RNA in the synthesis of protein. Cold Spring Harb Perspect Biol 3(11):a003780
Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun 371(4):818–822
Mukai T, Hayashi A, Iraha F et al (2010) Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res 38(22):8188–8195

Nagatsu T, Levitt M, Udenfriend S (1964) Tyrosine hydroxylase: the initial step in norepinephrine biosynthesis. J Biol Chem 239(9):2910–2917
Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28(1):292
Nessen MA, Kramer G, Back J et al (2009) Selective enrichment of azide-containing peptides from complex mixtures. J Proteome Res 8(7):3702–3711
Neumann H, Hancock SM, Buning R et al (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell 36 (1):153–163
Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464(7287):441–444
Neumann CS, Jiang W, Heemstra JR, Gontang EA, Kolter R, Walsh CT (2012) Biosynthesis of piperazic acid via N5-hydroxy-ornithine in Kutzneria spp. 744. ChemBioChem 13(7):972–976
Ngo JT, Tirrell DA (2011) Noncanonical amino acids in the interrogation of cellular protein synthesis. Acc Chem Res 44(9):677–685
Ngo JT, Champion JA, Mahdavi A et al (2009) Cell-selective metabolic labeling of proteins. Nat Chem Biol 5(10):715–717
Ngo JT, Schuman EM, Tirrell DA (2013) Mutant methionyl-tRNA synthetase from bacteria enables site-selective N-terminal labeling of proteins expressed in mammalian cells. Proc Natl Acad Sci USA 110(13):4992–4997
Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. J Am Chem Soc 133(30):11418–11421
Noren C, Anthony-Cahill S, Griffith M, Schultz P (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. Science 244(4901):182–188
Novoa EM, Ribas De Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet 28(11):574–581
O'Donoghue P, Prat L, Heinemann IU et al (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. FEBS Lett 586(21):3931–3937
O'Donoghue P, Ling J, Wang Y-S, Soll D (2013) Upgrading protein synthesis for synthetic biology. Nat Chem Biol 9(10):594–598
Odar C, Winkler M, Wiltschi B (2015) Fluoro amino acids: a rarity in nature, yet a prospect for protein engineering. Biotechnol J 10(3):427–446
Odoi KA, Huang Y, Rezenom YH, Liu WR (2013) Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the Escherichia coli BL21(DE3) cell strain. PLoS One 8(3):e57035
Oldach F, Al Toma R, Kuthning A et al (2012) Congeneric lantibiotics from ribosomal in vivo peptide synthesis with noncanonical amino acids. Angew Chem Int Ed Engl 51(2):415–418
Ouellette SP, Dorsey FC, Moshiach S, Cleveland JL, Carabeo RA (2011) Chlamydia species-dependent differences in the growth requirement for lysosomes. PLoS One 6(3):e16783
Park H-S, Hohn MJ, Umehara T et al (2011) Expanding the genetic code of Escherichia coli with phosphoserine. Science 333(6046):1151–1154
Parsons JF, Xiao G, Gilliland GL, Armstrong RN (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. Biochemistry 37(18):6286–6294
Peeler JC, Woodman BF, Averick S et al (2010) Genetically encoded initiator for polymer growth from proteins. J Am Chem Soc 132(39):13575–13577
Petrović DM, Leenhouts K, van Roosmalen ML, Broos J (2013) An expression system for the efficient incorporation of an expanded set of tryptophan analogues. Amino Acids 44 (5):1329–1336
Plass T, Milles S, Koehler C, Schultz C, Lemke EA (2011) Genetically encoded copper-free click chemistry. Angew Chem Int Ed Engl 50(17):3878–3881

Prasuhn DE, Singh P, Strable E, Brown S, Manchester M, Finn MG (2008) Plasma clearance of bacteriophage Qbeta particles as a function of surface charge. J Am Chem Soc 130 (4):1328–1334
Prat L, Heinemann IU, Aerni HR, Rinehart J, O'donoghue P, Söll D (2012) Carbon source-dependent expansion of the genetic code in bacteria. Proc Natl Acad Sci USA 109 (51):21070–21075
Rackham O, Chin JW (2005) A network of orthogonal ribosome x mRNA pairs. Nat Chem Biol 1(3):159–166
Reetz MT (2002) Lipases as practical biocatalysts. Curr Opin Chem Biol 6(2):145–150
Reeves MA, Hoffmann PR (2009) The human selenoproteome: recent insights into functions and regulation. Cell Mol Life Sci 66(15):2457–2478
Ressler C, Malodeczky H (1962) Isolation and identification from common vetch of the neurotoxin beta-cyano-L-alanine, a possible factor in neurolathyrism. J Biol Chem 237(3):733–735
Rodgers KJ, Shiozawa N (2008) Misincorporation of amino acid analogues into proteins by biosynthesis. Int J Biochem Cell Biol 40(8):1452–1466
Rosenthal GA (1982) L-Canavanine metabolism in Jack Bean, Canavalia ensiformis (L.) DC. (Leguminosae). Plant Physiol 69(5):1066–1069
Rosenthal G, Dahlman D, Janzen D (1976) A novel means for dealing with L-canavanine, a toxic metabolite. Science 192(4236):256–258
Sakamoto K, Murayama K, Oki K et al (2009) Genetic encoding of 3-iodo-L-tyrosine in Escherichia coli for single-wavelength anomalous dispersion phasing in protein crystallography. Structure 17(3):335–344
Santoro SW, Wang L, Herberich B, King DS, Schultz PG (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. Nat Biotechnol 20(10):1044–1048
Sauerwald A, Zhu W, Major TA et al (2005) RNA-dependent cysteine biosynthesis in archaea. Science 307(5717):1969–1972
Scannell JP, Pruess DL, Demny TC, Weiss F, Williams T, Stempel A (1971) Antimetabolites produced by microorganisms. II. L-2-amino-4-pentynoic acid. J Antibiot (Tokyo) 24 (4):239–244
Schlesinger S, Schlesinger MJ (1967) The effect of amino acid analogues on alkaline phosphatase formation in Escherichia coli K-12. I. Substitution of triazolealanine for histidine. J Biol Chem 242(14):3369–3372
Schlesinger S, Schlesinger MJ (1969) The effect of amino acid analogues on alkaline phosphatase formation in Escherichia coli K-12. 3. Substitution of 2-methylhistidine for histidine. J Biol Chem 244(14):3803–3809
Schmidt RL, Simonovic M (2012) Synthesis and decoding of selenocysteine and human health. Croat Med J 53(6):535–550
Scolnick E, Tompkins R, Caskey T, Nirenberg M (1968) Release factors differing in specificity for terminator codons. Proc Natl Acad Sci USA 61(2):768–774
Seitchik JL, Peeler JC, Taylor MT et al (2012) Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. J Am Chem Soc 134(6):2898–2901
Shah R, Neuss N, Gorman M, Boeck LD (1970) Isolation, purification, and characterization of anticapsin. J Antibiot (Tokyo) 23(12):613–619
Sharma N, Furter R, Kast P, Tirrell DA (2000) Efficient introduction of aryl bromide functionality into proteins in vivo. FEBS Lett 467(1):37–40
Staudt H, Hoesl MG, Dreuw A et al (2013) Directed manipulation of a flavoprotein photocycle. Angew Chem Int Ed Engl 52(32):8463–8466
Steiner T, Hess P, Bae JH, Wiltschi B, Moroder L, Budisa N (2008) Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. PLoS One 3(2):e1680
Stokes AL, Miyake-Stoner SJ, Peeler JC, Nguyen DP, Hammer RP, Mehl RA (2009) Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. Mol Biosyst 5(9):1032–1038

Strable E, Prasuhn DE, Udit AK et al (2008) Unnatural amino acid incorporation into virus-like particles. Bioconjug Chem 19(4):866–875
Swartz J (2001) A PURE approach to constructive biology. Nat Biotechnol 19(8):732–733
Tamura K, Schimmel P (2002) Ribozyme programming extends the protein code. Nat Biotechnol 20(7):669–670
Tang Y, Tirrell DA (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. J Am Chem Soc 123(44):11089–11090
Tang Y, Ghirlanda G, Petka WA, Nakajima T, Degrado WF, Tirrell DA (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. Angew Chem Int Ed Engl 40(8):1494–1496
Tang Y, Wang P, Van Deventer JA, Link AJ, Tirrell DA (2009) Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase. ChemBioChem 10(13):2188–2190
Teeuwen RLM, van Berkel SS, van Dulmen THH et al (2009) "Clickable" elastins: elastin-like polypeptides functionalized with azide or alkyne groups. Chem Commun (Camb) 27:4022–4024
Tsao M-L, Tian F, Schultz PG (2005) Selective Staudinger modification of proteins containing p-azidophenylalanine. ChemBioChem 6(12):2147–2149
Turanov AA, Lobanov AV, Fomenko DE et al (2009) Genetic code supports targeted insertion of two amino acids by one codon. Science 323(5911):259–261
Ugwumba IN, Ozawa K, Xu Z-Q et al (2010) Improving a natural enzyme activity through incorporation of unnatural amino acids. J Am Chem Soc 133(2):326–333
van Hest JC, Kiick KL, Tirrell DA (2000) Efficient incorporation of unsaturated methionine analogues into proteins in vivo. J Am Chem Soc 122(7):1282–1288
Vandenende CS, Vlasschaert M, Seah SYK (2004) Functional characterization of an aminotransferase required for pyoverdine siderophore biosynthesis in Pseudomonas aeruginosa PAO1. J Bacteriol 186(17):5596–5602
Voloshchuk N, Zhu AY, Snydacker D, Montclare JK (2009) Positional effects of monofluorinated phenylalanines on histone acetyltransferase stability and activity. Bioorg Med Chem Lett 19(18):5449–5451
Wagner I, Musso H (1983) New naturally occurring amino acids. Angew Chem Int Ed Engl 22(11):816–828
Walsh CT (2014) Blurring the lines between ribosomal and nonribosomal peptide scaffolds. ACS Chem Biol 9(8):1653–1661
Walsh CT, Garneau-Tsodikova S, Gatto GJJ (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl 44(45):7342–7372
Walsh CT, O'Brien RV, Khosla C (2013) Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. Angew Chem Int Ed Engl 52 (28):7098–7124
Wan W, Huang Y, Wang Z et al (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. Angew Chem Int Ed Engl 49(18):3211–3214
Wan W, Tharp JM, Liu WR (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta 1844(6):1059–1070
Wang L, Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. Chem Biol 8(9):883–890
Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of Escherichia coli. Science 292(5516):498–500
Wang L, Zhang Z, Brock A, Schultz PG (2003a) Addition of the keto functional group to the genetic code of Escherichia coli. Proc Natl Acad Sci USA 100(1):56–61
Wang L, Xie J, Deniz AA, Schultz PG (2003b) Unnatural amino acid mutagenesis of green fluorescent protein. J Org Chem 68(1):174–176

Wang K, Neumann H, Peak-Chew SY, Chin JW (2007a) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 25(7):770–777
Wang J, Schiller SM, Schultz PG (2007b) A biosynthetic route to dehydroalanine-containing proteins. Angew Chem Int Ed Engl 46(36):6849–6851
Wang Q, Parrish AR, Wang L (2009) Expanding the genetic code for biological studies. Chem Biol 16(3):323–336
Wang F, Robbins S, Guo J, Shen W, Schultz PG (2010) Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. PLoS One 5(2):e9354
Wang Y-S, Russell WK, Wang Z et al (2011) The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. Mol Biosyst 7 (3):714–717
Welter A, Marliert M, Dardenne G (1978) Nouveaux acides amines libres de Afzelia bella: trans-hydroxy-4-L-proline et trans-carboxy-4-L-proline. Phytochemistry 17(1):131–134
Wilkins BJ, Marionni S, Young DD et al (2010) Site-specific incorporation of fluorotyrosines into proteins in Escherichia coli by photochemical disguise. Biochemistry 49(8):1557–1559
Wiltschi B (2012) Expressed protein modifications: making synthetic proteins. Methods Mol Biol 813:211–225
Wiltschi B, Wenger W, Nehring S, Budisa N (2008) Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues. Yeast 25(11):775–786
Wolschner C, Giese A, Kretzschmar HA, Huber R, Moroder L, Budisa N (2009) Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. Proc Natl Acad Sci USA 106(19):7756–7761
Wu N, Deiters A, Cropp TA, King D, Schultz PG (2004) A genetically encoded photocaged amino acid. J Am Chem Soc 126(44):14306–14307
Wu IL, Patterson MA, Carpenter Desai HE, Mehl RA, Giorgi G, Conticello VP (2013) Multiple site-selective insertions of noncanonical amino acids into sequence-repetitive polypeptides. ChemBioChem 14(8):968–978
Wu JCY, Hutchings CH, Lindsay MJ, Werner CJ, Bundy BC (2015) Enhanced enzyme stability through site-directed covalent immobilization. J Biotechnol 193:83–90
Xiao H, Chatterjee A, Choi S-H, Bajjuri KM, Sinha SC, Schultz PG (2013) Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. Angew Chem Int Ed Engl 52(52):14080–14083
Xie J, Wang L, Wu N, Brock A, Spraggon G, Schultz PG (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. Nat Biotechnol 22 (10):1297–1301
Yoo TH, Link AJ, Tirrell DA (2007) Evolution of a fluorinated green fluorescent protein. Proc Natl Acad Sci USA 104(35):13887–13890
Yoon Byung C, Jung H, Dwivedy A, O'hare Catherine M, Zivraj Krishna H, Holt Christine E (2012) Local translation of extranuclear lamin B promotes axon maintenance. Cell 148 (4):752–764
Yoshioka H, Nakatsu K, Sato M, Tatsuno T (1973) The molecular structure of cyclochlorotine, a toxic chlorine-containing cyclic pentapeptide. Chem Lett 2(12):1319–1322
Young TS, Ahmad I, Brock A, Schultz PG (2009) Expanding the genetic repertoire of the methylotrophic yeast Pichia pastoris. Biochemistry 48(12):2643–2653
Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50(11):1894–1900
Yu Z, Pan Y, Wang Z, Wang J, Lin Q (2012) Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells. Angew Chem Int Ed Engl 51(42):10600–10604
Zhang Z, Wang L, Brock A, Schultz PG (2002) The selective incorporation of alkenes into proteins in Escherichia coli. Angew Chem Int Ed Engl 41(15):2840–2842
Zhang Z, Alfonta L, Tian F et al (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proc Natl Acad Sci USA 101(24):8882–8887

Zhang Y, Fomenko D, Gladyshev V (2005a) The microbial selenoproteome of the Sargasso Sea. Genome Biol 6(4):R37

Zhang Y, Baranov PV, Atkins JF, Gladyshev VN (2005b) Pyrrolysine and selenocysteine use dissimilar decoding strategies. J Biol Chem 280(21):20740–20751

Zheng S, Kwon I (2012) Manipulation of enzyme properties by noncanonical amino acid incorporation. Biotechnol J 7(1):47–60

Synthetic Biology for Cellular Remodelling to Elicit Industrially Relevant Microbial Phenotypes

5

Paola Branduardi

Contents

5.1 Introduction 212
5.2 Rewiring the Cellular Set-Up at Genomic Level 213
5.2.1 Towards Minimal Cell, If Ever 214
5.2.2 Genome Engineering and Combinatorial Libraries for Tuned (Recombinant) Expression 214
5.3 The Long Road from mRNA to Proteins 216
5.3.1 Post-translational Regulation of Levels and Local Concentrations of Proteins 217
5.3.2 Rewiring Cellular Differentiation by Modulating Post-translational Modifications 217
5.4 Metabolic Controls: How to Turn Them to Our Purposes and Not Against? 218
5.4.1 Mimicking the Dynamics of Metabolic Responses to Optimise Recombinant Productions 220
5.4.2 Perturbing Central Regulatory Systems of the Principal Elements 221
5.5 Enhancing Production: A Matter of Quantity or of Quality State? 221
5.5.1 Cellular Stress and Integrity of the Cellular System 222
5.5.2 Novel Signalling Functions for "Old" Molecules 223
5.6 Conclusions 224
References 225

Abstract

Industrial microbiology is proposing an increasing number of bio-based processes that are ready to move from the validation to the demonstration step, with the industrial world being more open to this opportunity for a change. The challenge is therefore to make such processes viable and competitive. When moving from the lab to the industrial scale, the degree of complexity is

P. Branduardi (✉)
Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, 20126 Milan, Italy
e-mail: paola.branduardi@unimib.it

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_5

increasing, and the engineered cell factories very often display emerging properties that can be explained only from a systems perspective. Unfortunately, cellular rewiring often leads to a lower accumulation of the desired product. Synthetic biology is willing to take advantage from the knowledge on mechanisms involved in cellular homeostasis and, thanks to the principles of abstraction, modularity and standardisation, translate them into more efficient cell factories. Indeed, this novel approach to potentiate the power of metabolic engineering can be applied not only to a specific metabolic pathway but can be extended to networks indirectly connected to the pathway of interest. In this chapter, some of the principal synthetic tools developed to regulate or redirect the remodelling of cell factories, from genomic to metabolic level, with the aim to obtain higher titers, yield and productivity of bio-based products will be described and commented.

5.1 Introduction

Industrial microbiology is moving from a promising field of application to a real alternative to chemical synthesis, as witnessed by the increasing number of processes emerging at industrial scale [see, for a recent review, Porro et al. (2014)]. The majority of the latest announcements relate to recombinant productions. This fact highlights at least two important aspects: (1) the efforts in improving metabolic engineering approaches and tools turn out to be more and more successful and (2) the industrial world seems to be more open than in the past for a change. Indeed, despite all the optimisation strategies for development of an effective cell factory that can be imagined at lab scale, the real viability test of a process is only possible approaching industrial scale. Therefore, industries need to open up to the possibility to test the new biocatalysts and eventually to consider how to combine operative conditions with biological needs and vice versa. For industrial use, the biocatalyst is forced to produce a non-natural product (or, even if natural, to accumulate it up to non-physiological titer) in a non-natural environment (where the catalyst might tend to be metabolically inactive). Notably, the coexistence of these two inappropriate conditions triggers different (and possibly coordinated) cellular responses that involve the cell as a whole, having as a consequence an effect on metabolism and therefore on production. Despite the increasing number of studies on single responsive elements as well as on genome-scale networks, especially devoted to well-established cell factories such as *Escherichia coli* or *Saccharomyces cerevisiae*, it is still hard to predict how the cells will reprogramme their state as a response to the process conditions. In line with that, a significant number of reports are proposing adaptive laboratory evolution (ALE) as an efficient strategy to select the desired phenotype [for recent reviews, see Dragosits and Mattanovich (2013) and Winkler and Kao (2014)]. This is not in contrast with strategies proposing a rational design, since the analysis of the novel set-up of the evolved

cell factory can contribute to a better understanding of the whole system and/or to shed light on still uncovered responsive elements. As a remark, it has to be said that the efficacy of ALE is strongly related to the conditions applied for strain selection, which should be intended to mimic as much as possible the relevant process conditions.

Considering that, would it imply that the role of metabolic engineering or, more generally, of a tailored manipulation is limited to the initial stage of strain construction?

The ultimate idea is to come up with a synthesis of the two strategies, resulting in the construction of synthetic cells (as reported in Gibson et al. (2010) and within the programme Synthetic Yeast 2.0, http://syntheticyeast.org), specifically dedicated to a particular production or function. The outcomes of these projects will reverberate not only in applications with social impact but also in novel insights at the fundamental level of knowledge, opening the possibility to extend the concepts from single-cell organisms to more complex systems, such as cellular communities, possibly organised in differentiated organs or pluricellular organisms.

However, this scenario is still far from being real, while the need of a turn for establishing a viable bioeconomy is more and more urgent in our society. This is how synthetic biology can also come to the aid of industrial biotechnology. To make a complex story as simple as possible, a recombinant cell factory is usually genetically modified to (over)express endogenous or heterologous genes and/or to impair endogenous gene(s) function, to obtain the desired product. These modifications, per se, imply a rewiring of (genetic) circuits, necessary to cope with the newly imposed set-up. In addition, when cell factories are cultivated in the production medium and particularly under industrially relevant conditions, they have to cope with constantly fluctuating and often limiting external and internal parameters along the fermentation time. This will also have an effect on the rewiring of circuits, which is partially dependent on the strain's potential and partially on the severity of process constraints. Unfortunately, the ultimate result of the cellular adaptation is very often not optimal for process success. Synthetic biology, through the study of master regulator key elements and the reconstruction of such elements as parts, could allow us to regulate or redirect some of the cellular rewiring. This can turn into higher production, yield and productivity (Van Dien 2013), which matches with our need to reduce time-to-market.

Here we recapitulate the principles, together with the description of some synthetic tools and approaches (with a preference for *S. cerevisiae* and *E. coli* as chassis, due to the availability and applicability of tools), which have been developed to elicit cellular phenotypes to be selected for matching with specific purposes.

5.2 Rewiring the Cellular Set-Up at Genomic Level

Based on currently available information, the largest number of manipulations for the construction of a cell factory is targeted at genomic level. In this context, synthetic biology is developing combinatorial approaches, often with the final aim of streamlining genomes, and consequently transcriptomic responses. Here

we recapitulate some of the most promising approaches, which follow the principles of modularity and standardisation.

The perspective is that in a near future, the achieved knowledge, at least regarding the most used cell factories, will allow a tailored design, where predictions can be translated into results with less dependence on vast use of screening procedures.

5.2.1 Towards Minimal Cell, If Ever

In addition to the above-mentioned project on *S. cerevisiae* 2.0, genome reduction has already been translated into a desired phenotype by manipulating *E. coli* cells. This is, for example, the case for strain MDS42, tested as cell factory for lentiviral expression after removal of about 40 insertion sequences (Csörgo et al. 2012). Interestingly, a further reduction was targeted to the genes encoding error-prone DNA polymerases: this loss-of-function resulted into low-mutation-rate strains, which means genetic stability. Indeed, the maintenance of the desired genome over the whole process is definitely a must for standardisation and quality of production, especially if the cell factory is the result of desired and tailored recombinant manipulations. However, it is also interesting to notice that variations on genome structures elicited by detrimental process conditions and often observed at the end of fermentation processes [duplications, mutations, aneuploidy, etc.; see as example Puig et al. (2000)] can be considered as precious source of information due to our still poor knowledge about cellular intrinsic potential.

What is very interesting in examining the literature is that the "minimal" cell can be tailored only in connection with defined environmental conditions: referring to the previous example, it is clear that a strain manipulated as such very likely turns out with a different fitness if compared and in competition with natural strains. Following the same line of thoughts, the lack of genes encoding for biofilm structural components (May and Okabe 2011) can be a way of reducing genomic functions, but only if this loss-of-function is not going to compromise the final performance of the strain. The multiple techniques for genome reduction currently under development are therefore highly relevant [see, as recent examples, Xue et al. (2014) and Suzuki et al. (2015)], but most of them remain to be tested at production level.

5.2.2 Genome Engineering and Combinatorial Libraries for Tuned (Recombinant) Expression

Experimental evolution can be an efficient strategy to obtain a desired phenotype, as demonstrated by many recent applications of adaptive laboratory evolution (ALE) [see, as examples, Marietou et al. (2014); Patzschke et al. (2015) and Tilloy et al. (2014)]. However, at laboratory scale, the natural variation is limited, such as the timescale of such experiments to achieve a profound rearrangement of the

genome is limited. Considering literature data about cell factory and selective conditions, an average of 0.97 SNPs in an adapting genome for every 100 generations of evolution was recently calculated (Dettman et al. 2012). It can be estimated that cultivations for ALE typically last from 200 to 1000 generations. Therefore, considering the average of mutation frequency and genome dimension, this can result into 4–20 independent mutations per population (see Pál et al. (2014) and references therein). This is well exemplified by the experiments by Barrick and coauthors, showing that an *E. coli* strain acquired only 45 mutations over 20,000 generation of adaptation to glucose minimal medium (Barrick et al. 2009) and further pointing out a complex relation between genomic and adaptive evolution. Moreover, it is very difficult to establish an appropriate control for mutational processes and to distinguish beneficial mutations from neutral ones.

Synthetic biology approaches are implementing genome engineering, with at least two different aims: (1) to increase the number of variants generated in a limited time scale and (2) to facilitate the traceability of such mutations, to consequently facilitate the description (if not the comprehension) of the results. Remarkably, genome engineering can generate modifications that have never been explored in nature or that might have occurred in ancient time, when divisions among the three cellular lines were still at the very beginning and the incidence of horizontal gene transfer was more relevant.

Multiplex automated genome engineering (MAGE) allows the targeted editing of many different locations mediated by the replacement of the natural sequences with generated oligonucleotides that can operate an allelic exchange (Wang et al. 2009). This recombination, which can be iteratively repeated in cycles, generates a combinatorial genomic diversity. This technique was efficiently applied on *E. coli* to optimise the production of lycopene, an industrially relevant carotenoid. Recently, the MAGE technology has been implemented based on the exploitation of suicide plasmids, to make it applicable not only to few laboratory strains but also to industrial bacterial strains (Ryu et al. 2014).

MAGE can be conceivably combined with CAGE (conjugative assembly genome engineering), a method of large-scale transfer of engineered (and marked) genomic modules to assemble them into a single genome, which is based on the appropriate positioning of *oriT* conjugational sites and selection markers (Ma et al. 2014). The combination can allow reconstructing a (whole) genome where the genetic code has been modified, or expanded, for example, for the production of proteins including non-natural amino acids.

As mentioned before, the traceability of the introduced variation(s) is highly desirable. Trackable multiplex recombineering (TRMR) is such a method where DNA barcodes, detectable by microarray analyses, have been included in the recombineering cassettes (Mansell et al. 2013).

In view of evoking unravelled cellular potential to accomplish industrially relevant features, Alper et al. (2006) have proposed a strategy for reprogramming the transcriptional set-up by mutagenesis applied on key transcriptional elements. This technique, named global transcription machinery engineering (gTME), is indeed based first on the identification of transcription factors (TFs) that are

known as master regulators of general rewiring events in the cell, second in the construction of a library of mutated versions of said TFs (by error-prone PCR), third in the transformation of this library into the target cell factory, and finally in the selection under desired conditions of the strain(s), where the mutation(s) will evoke the desired novel transcriptional set-up. Eventually, the selected strains are investigated both for the mutations occurring at the TF level and more generally at the transcriptomic level, to possibly transfer the identified perturbations into the parental strain.

This technique has been first applied to *S. cerevisiae* to improve ethanol tolerance and volumetric productivity (Alper et al. 2006), followed by another example in *E. coli* (Alper and Stephanopoulos 2007), where the mutagenesis applied on a sigma factor also elicited different industrially relevant phenotypes, namely, ethanol tolerance, lycopene overproduction and an example of acquisition of multiple phenotypes (specifically, ethanol and sodium dodecyl sulfate tolerance). Different selection approaches are then proposed, combining the more classical growth assays on plates with chemostat cultivations. A number of different reports have been published, exploiting the principle of the strategy to accomplish a complex phenotype rearrangement [see as recent examples Tan et al. (2015) and Zhao et al. (2014)].

Similarly, but taking advantage of a different mechanism, a transposon-mediated mutant library has been used to identify strains with novel traits (Kim et al. 2011). Given their relatively simple design and their high compatibility at transducing genetic material, transposons are highly attractive to be used as parts in synthetic biology.

5.3 The Long Road from mRNA to Proteins

Despite the critical regulation of gene expression at transcriptional level, all following events during gene expression are also crucial for reaching the desired phenotype. In this context, it is important to underline a general difference between production of pharma proteins and the production of chemicals, especially those implying different catalytic steps. In the first case, usually the heterologous gene is expressed at highest levels to maximise production (despite increasing evidence on the consequent metabolic burden suggests caution). In the second case, the relative level of all the enzymes involved in the metabolic pathway of interest has to be carefully balanced in terms of processing ability, allowing an optimal flux to the desired product, thereby avoiding the accumulation of intermediates, possibly toxic to the cells. Moreover, and this is true also for pharma-protein production, all the mechanisms responsible for translational events, sensing and repairing are of pivotal importance. The study of the structural elements present on mRNA molecules (Kwok et al. 2015; Mitchell and Parker 2014) and the implication of mRNA distribution, docking, storage and metabolism (Decker and Parker 2012; Walters and Parker 2014) are suggesting novel targets for controlling protein production.

To further extend these concepts, it has to be said that the classical dogma of the information flow from DNA to RNA to protein to direct cellular activities does not take epigenetics into account, neither noncoding RNAs, protein-protein interactions, nor post-translational modifications. These levels of regulation are currently being analysed by systems biology studies as emergent properties of the system but rarely targeted for synthetic approaches yet.

Nevertheless, synthetic biology is developing strategies to regulate these events, both with finite purpose and with the aim to elicit general responses and phenotypic configurations. Here we will focus on this second aspect.

5.3.1 Post-translational Regulation of Levels and Local Concentrations of Proteins

When using bacterial cell factories, the reconstruction of the desired metabolic pathway by grouping necessary genes in the form of an operon ensures coordination, but also a similar level of expression, since they are all under the control of the same promoter. However, this does not imply that the enzymatic levels are optimised for the production of interest. The possibility of a post-transcriptional tuning of protein levels by playing with the translation initiation region (TIR) present on mRNA molecules has been demonstrated for mevalonate production in *E. coli* (Pfleger et al. 2006). The authors demonstrated that by generating libraries of tunable intergenic regions (TIGRs), recombining various post-transcriptional control elements such as mRNA secondary structures, RNase cleavage sites and ribosome-binding site (RBS) sequences, it was possible to vary the relative expression of different genes of one or even two orders of magnitude. Remarkably, this can be translated into the possibility of fishing for strains where the protein levels are optimally concerted for the desired production.

Another level of general reorganisation that can be critical for recombinant productions is the localisation of the heterologous pathway. Substrate availability, local concentrations and subcellular compartmentalisation are all reasons in favour of a rational design implying the targeting of necessary enzymes on macromolecular scaffolds. In this view, the RNA role can be extended to the novel function of constituting a three-dimensional scaffold harbouring aptamers for docking the desired enzymatic functions, as it was demonstrated for hydrogen production in *E. coli* (Delebecque et al. 2011, 2012).

5.3.2 Rewiring Cellular Differentiation by Modulating Post-translational Modifications

Post-translational modifications of proteins, counting more than 600 different types in eukaryotic cells (as annotated in the RESID database, http://pir.georgetown.edu/resid/), are largely responsible for cellular responses and coordinated activities [see, for a review, Deribe et al. (2010)]. Very likely, one of the most abundant

modifications is phosphorylation, involved in signal transduction cascades and in coordinating external stimuli with cellular responses. The possibility to exert a control on such cascades can determine the success of a biotechnological process, as in the case of differentiation events. Indeed, one of the critical elements when working with sporulating bacteria is the possibility to control the triggering of this differentiation pathway. Vishnoi et al. (2013) pointed out the importance of the timing in the phosphorylation of the *Bacillus subtilis* master regulator of sporulation initiation, Spo0A. By controlling the induction of the kinase KinC, together with Spo0A, they could elicit sporulation even under ideal nutrient conditions, demonstrating the possibility to rewire even a fundamental cellular mechanism of survival. This study, as well as other similar, opens up the possibility to create synthetic relays and switches, which could be applied to industrial strains, when the triggering of sporulation can be critical for the stability of the production, as in the case of clostridial host organisms, very well known for acetone-butanol-ethanol fermentation (Patakova et al. 2013).

5.4 Metabolic Controls: How to Turn Them to Our Purposes and Not Against?

All the cellular mechanisms that have been mentioned up to this point need to be in strong connection and coordination with cellular metabolism. A cell operates its many and various functions through a flexible and efficient metabolism. As a consequence, knowing how the cellular metabolism is regulated and how it can be manipulated is of pivotal importance to optimise microbial cells for industrial bioprocesses. Metabolic engineering has provided several successful examples of de novo production of different classes of metabolites, optimisation of endogenous products and exploitation of many different (including nonnatural) substrates, among others (Table 5.1). However, it is clear that the designed metabolic pathways will unavoidably interfere with the native metabolism and with its control. Even more, the new metabolites produced will have an active role in the general novel set-up and response of the recombinant cell. To shortly list the principal effects, metabolite concentration can regulate enzymatic activity (different mechanisms), can lead to their covalent modifications and can influence translational and transcriptional regulation (at different levels), as recently reviewed by Wegner et al. (2014). All these responses may occur simultaneously, but acting at different levels, and with different timescales, both in the triggering and in the extinguishing of consequences. Noticeably, even the codon optimisation strategies started to take into account that there is not just a single cellular set-up that can be considered as an absolute reference. Generally speaking, codon frequency can differ, within the same cell, among genes encoding for enzymes belonging to different pathways (*i.e.* primary versus secondary metabolism). Even more, in *S. cerevisiae* it has been recently demonstrated that the levels of the different tRNA species vary along the cell cycle progression, contributing to the triggering and to the crossing of checkpoints (Frenkel-Morgenstern et al. 2012). Thanks to these studies, a

Table 5.1 Recent examples that exemplify the successful development of cell factories optimised for the production of some of the main bio-based chemicals

	Microorganism	Titer	Reference
Biofuels			
Ethanol (from lignocellulose)	*S. cerevisiae*	~60 g/L	Several studies, see review Sedlak and Ho (2004); Yamada et al. (2013)
n-Butanol	*E. coli*	30 g/L	Shen et al. (2011)
Isobutanol	*E. coli*	50 g/L	Baez et al. (2011)
Triacylglycerols (TAGs, to biodiesel)	*Yarrowia lipolytica*	25 g/L	Blazeck et al. (2014)
Fatty acid ethyl esters (FAEE)	*S. cerevisiae*	34 mg/L	Shi et al. (2014)
Chemicals			
1,3-Propanediol (1,3-PDO)	*E. coli*	135 g/L	Nakamura and Whited (2003)
Succinic acid	*Corynebacterium glutamicum*	146 g/L	Okino et al. (2008)
2,3-Butanediol (BDO)[a]	*Bacillus amyloliquefaciens*	132.9 g/L	Yang et al. (2013)
R-2,3-Butanediol (BDO)[a]	*S. cerevisiae*	100 g/L	Lian et al. (2014)
Styrene	*S. cerevisiae*	29 mg/L	McKenna et al. (2014)
Fine chemicals			
Amorphadiene	*S. cerevisiae*	40 g/L	Westfall et al. (2012)
Resveratrol	*S. cerevisiae*	391 mg/L	Sydor et al. (2010)
Artemisinic acid	*S. cerevisiae*	25 g/L	Paddon et al. (2013)
Omega-3 eicosapentaenoic acid (EPA)	*Y. lipolytica*	25 % DCW[b]	Xie et al. (2015)

[a]Also as fuel or fuel additive
[b]Calculated as percentage on cell dry weight (CDW)

condition-specific codon optimisation approach has been recently designed for improving heterologous productions in *S. cerevisiae* (Lanza et al. 2014).

The profound interconnection of metabolism with the other cellular functions makes it difficult to draw a line distinguishing specific techniques. However, the examples reported here started from metabolic master regulator for sensing and connection with metabolic fluxes as the main target for eliciting a cellular rewiring.

5.4.1 Mimicking the Dynamics of Metabolic Responses to Optimise Recombinant Productions

When active (growing, not resting) engineered cells are used for a commercially viable process of commodities production, a very high yield of production [at least 80 % of the theoretical yield according to the metrics of Van Dien (2013)] needs to be obtained. This means that the flux to the product has to be maximised, leading to an inescapable trade-off between growth and production. Channelling most of the substrate into the product can impair not only growth as final effect but can also lead to cofactor imbalance and toxic intermediate accumulation, and finally, it can reduce the cellular intrinsic resilience. A good compromise can be achieved by separating growth and production phase, which has been more easily obtained by modifying process conditions and not the expression pattern of the engineered cells. Synthetic biology is now offering the possibility to dynamically switch the metabolic configuration from "growth" to "production" by the design of genetic circuits, implying sensors and actuators (Michener et al. 2012).

A relevant number of genetic circuits are already available and continuously augmenting [see for a recent review Venayak et al. (2015)]. They can be divided in circuits with an on-off control of a two-stage fermentation process and circuits that can ensure a continuous control. The first approach utilises on-off methods that can involve physical parameters (such as temperature, pH) or chemicals (i.e. inducers) as triggers. Solomon and coauthors, for example (Solomon et al. 2012), optimised gluconate production by an engineered *E. coli* strain. Gluconate is a classical example of a product that depends on central carbon metabolism but at the same time is drawing flux from glycolysis. In this situation, tuning the expression of the endogenous competing enzymes can be the solution. To reach this goal despite the tight regulation distinguishing the glycolytic metabolism, the authors exogenously manipulated glucokinase (Glk) levels, either with an engineered antisense RNA or with an inverting gene circuit, inhibiting Glk activity by up to 25 % and 50 %, respectively. Using these techniques, gluconate production increased and acetate production decreased, as well as specific growth rate (up to 50 %), but with no difference in final biomass accumulation. This elegant and successful example of dynamic control would need to be verified at industrial scale, since not only the final titer but also the total fermentation time represents a crucial issue for commodities production.

The second approach makes use of dynamic sensors, so that the circuit can respond to environmental changes and control the metabolism accordingly. In the last few years, some examples have been described in the literature. Among them is a report on the dynamic control of lycopene production (Farmer and Liao 2000) where the lycopene biosynthetic pathway was under the positive control of a sensor responsive to acetyl-phosphate, a metabolite whose concentration indicates a high glycolytic flux.

A significant number of works relates with the cellular management of toxic compounds, which is a crucial point of viable industrial bioprocesses. Indeed, toxic compounds can derive from the starting substrate used as feedstock, or the toxicity

is determined by the accumulation of the product of interest. This is typically the case when alcoholic biofuels are produced, heavily interfering with many cellular components, starting from membranes. Dunlop and coauthors designed an in silico model, employing a search for a robust phenotype and considering the timing of cellular response, where solvent-resistant bacterial efflux pumps are under the control of a biofuel-responsive promoter, resulting in genetic architectures improved for biofuel production (Dunlop et al. 2010).

At the current stage of development, these systems have to be accurately customised, both in the construction as in the validation and demonstration steps: a future development can be envisaged in the combination of genetic circuits with fermentation strategies, aiming at an integrated contribution of systems and synthetic biology to guide successful industrialisation of bioprocesses.

5.4.2 Perturbing Central Regulatory Systems of the Principal Elements

The techniques just described underline the complexity of manipulating a native metabolism with the final aim of a complete redirection of the carbon flux to the desired product. An alternative way to overcome this limitation is to count on the cellular capability to reorganise when a central mechanism (also defined as non-pathway components, like transcription or translation) is perturbed. This was well described by McKee and coauthors: by perturbing the carbon storage regulator (Csr) system, previously described to cause a profound reorganisation of *E. coli* central metabolism (Edwards et al. 2011), they were able to obtain a twofold improvement of n-butanol production with a concomitant acetate decrease (McKee et al. 2012). The basis for these results remains to be understood, since the prominent effect of the perturbation was an increase in amino acid levels. However, the approach appears intriguing as it could be applied to other prokaryotic cells, being Csr a conserved system.

5.5 Enhancing Production: A Matter of Quantity or of Quality State?

A great deal of focus in cell factory design was and still is devoted to ensure quantitative settings (e.g., in terms of gene copy number, protein levels and so on), as an essential prerequisite for high production levels. However, a relevant number of evidence suggest a physiological balancing as the more correct strategy to be pursued (see also some of the previous examples). Therefore, in the last years, growing attention is being directed to the quality state of macromolecules, instead of their quantitative state. This concept becomes even more central when the cell factory has to operate within a process of production, encountering detrimental situations that can exacerbate unstable cellular settings. Indeed, many classes of macromolecules are very often found to be profoundly impaired by limiting

situations, resulting in lipid peroxidation, protein carbonylation, reactive species accumulation and cofactor imbalance. In these conditions, the titer of the desired product might be augmented by increasing the copy number of genes directly responsible of its production and by modulating their expression, mRNA stability, number of influx and efflux transporters, but all these efforts are doomed if the macromolecular quality control is not ensured. This is then another aspect where systems biology can suggest strategies for the development of synthetic tools to be applied in supporting the mainstream of production and to ensure the functionality of the system as a whole.

5.5.1 Cellular Stress and Integrity of the Cellular System

Only a functional proteome allows the proper expression of a cellular being and a rapid and appropriate response to altered intracellular or extracellular conditions. A functional proteome is therefore a prerequisite for a stable phenotype of production when a cell factory is challenged during an industrial process. To cite Prof. Radman: "the proteome sustains and maintains life, whereas the genome ensures the perpetuation of life by renewing the proteome, a process contingent on a preexisting proteome that repairs, replicates, and expresses the genome" (Krisko and Radman 2013). Regrettably, very often exactly the conditions of an industrial process elicit damages to the proteome. In most cases, oxidative stress plays an important detrimental role, and even when the production is anaerobic, the biomass propagation phase is aerobic. More generally, the occurrence of free radicals is quite common in industrial process condition, triggering a cascade of negative effects on different cellular networks. There is a growing number of examples where engineering cellular components belonging to the general mechanisms of scavenging activities result in an improved phenotype. This is the case of *S. cerevisiae* strains engineered for ascorbic acid production, resulting in cells more robust against a series of stressors, among them organic acids (Martani et al. 2013), or overexpressing gene(s) of the glutathione pathway resulting in a less sensitive phenotype to toxic compounds released from lignocellulose pretreatment (Ask et al. 2013). Such modifications determine a complete reorganisation of the cells, influencing many pathways, not only directly related to the operated modification or to the desired productive pathway. Therefore, a combined systems and synthetic approach will possibly lead in a near future to easily customisable tools for augmenting cellular robustness against stressors. An important caveat recalls what has been previously mentioned about the dynamics of the systems, which refers both to the intracellular and to the extracellular environment. Therefore, it has to be taken into account that different stressors are acting on the cells over time, but not always the same stressors, not always with the same intensity.

5.5.2 Novel Signalling Functions for "Old" Molecules

The potential and versatility of RNA molecules has been exploited from nature since its appearance, probably even before the definition of cellular units. We are in a continuous process of discovery of the many and different RNA functions inside cells: the deeper our knowledge gets, the more synthetic strategies for applying the novel information can be designed and adapted to the cell factory of interest. This is the case with microRNAs (miRNAs), small noncoding RNAs that can participate in the modulation of protein expression [see, for a review, Inui et al. (2010)], can act as mediators among cells [see, for a review, Kohlhapp et al. (2015)] and were also suggested to have a role in conferring robustness to the cells (Ebert and Sharp 2012). In particular, this role has been described in cells facing environmental fluctuation during development (Li et al. 2009), and miRNAs have been also depicted as stabilisers of fluctuations in gene expression (Chang et al. 2011; Herranz and Cohen 2010). Moreover, recent findings suggest miRNAs as essential for lifespan determination in the model organisms, *Caenorhabditis elegans* and *Drosophila* [as reviewed by Kato and Slack (2013)]. Considering the importance of robustness, ageing and ageing-related issues for industrial processes, miRNAs appear as a very promising target for remodelling cell factories in this direction.

A completely different perspective for seeing the signalling would be not to consider individual cells but the whole population. Indeed, if cells belonging to different subpopulations could be devoted to diverse functions, this would decrease their metabolic burden and simplify the construction of genetic circuits. However, the necessary assumption is that such subpopulations have to be able to coordinate their functions through regulated intercellular communication exchanges, which should also act at reducing the possible noise of random fluctuations. Intercellular crosstalk exists and is widely used in all the three biological kingdoms of life, therefore also in microbial systems, but exactly because of the fundamental and ancient origin of the system, its engineering is far from being trivial.

As a starting point, the synthetic reconstruction of natural communication systems can be considered: their application is now primarily intended to validate the possibility to translate them in controlled circuits, and mainly in bacterial hosts, despite the first applications are emerging together with the extension of the concept from unicellular to pluricellular eukaryotic systems.

The bacterial quorum sensing (QS) system was the first to be translated in synthetic circuits and in particular the one based on LuxR/LuxI from *Vibrio fischeri*, where acyl-homoserine lactone (AHL, synthetised by LuxI) mediates intercellular coupling by binding to LuxR: in turn, the complex controls the expression of target genes [LuxI as first (Waters and Bassler 2005)].

When expressed in *E. coli*, this system allowed the cellular density-controlled expression of the killer gene ccdB (You et al. 2004) to autonomously regulate cellular density, or of the heterologous invasin [from *Yersinia pseudotuberculosis* (Anderson et al. 2006)] to invade tumour cells after injection in such tissue, with the possibility to target a specific delivery or response. A more complex system, including the coordinated expression of a protease that negatively regulates LuxI

expression by catalysing the degradation of AHL, allowed the creation of a synthetic oscillator (Danino et al. 2010). Examples of how to further implement this technology for inducing synchronised oscillations between bacterial populations have been reported (Liu et al. 2011; Prindle et al. 2012). The system can be further modulated if the genetic components of the LuxI/LuxR QS are separated in two populations acting as "sender" and "receiver" of signal, respectively.

Despite being still not directly connected with industrial production processes, the implementation of circuits of communication among cellular (sub)populations could inspire novel strategies for the exploitation of microbial communities, considering the final aim of developing consolidated bioprocess-based biorefineries (Shong et al. 2012).

5.6 Conclusions

The still incomplete knowledge about genes and their products, and therefore about cellular regulatory networks and their connections, is currently limiting the prediction of the final results when a novel biocatalyst is genetically generated. The process conditions, which are often detrimental for the cells and always producing fluctuations in chemical-physical parameters, further contribute to evoke cellular answers very difficult to be predicted. Screening and selection can therefore be a strategy to retrieve the desired phenotype, but there are at least two *caveats*: (1) to reproduce an industrial production process including all the limiting factors at lab scale is not trivial, resulting in the possible selection of an undesired or suboptimal strain and (2) the evoked phenotype might be not stable enough to constitute a viable and robust biocatalyst. On the other hand, the innate capability of cellular systems to adapt, to express emergent properties, to possibly communicate at intercellular level and to profoundly rearrange their settings at different level is still largely underestimated and unexploited (Lajoie et al. 2013).

One important element that has been only accidentally mentioned is timing: transcriptional regulations can be finely tuned and regulated, but they will result in slow responses, and quite similar situations will be encountered by manipulating translation, since protein half-life is in the average of hours. Therefore, circuits acting at these levels are matching the need for responses acting on the same order of time. Diversely, post-translational control mechanisms are effective on the order of fraction of seconds and can therefore be efficaciously applied to follow metabolic fluctuations, which occur on the order of minutes.

The urgency for a real turn into bioeconomy and for the establishment of viable biorefineries, possibly integrated in the territory, can also accomplish the rising need of organic waste management, which in turn can determine a more equal availability of feedstock. The uneven nature of sustainable biomasses together with industrial process requirements makes the development of robust and versatile biocatalysts unavoidable.

Systems and synthetic biology are currently playing a pivotal role, by successfully tuning their activities in providing a novel array of customisable, robust and

versatile cell factories with defined and predictable properties, and are additionally promising to move from synthetic cells to synthetic consortia.

Acknowledgements PB thankfully acknowledges Michael Sauer for the fruitful discussions and for critically reviewing the work. PB acknowledges the financial support by the FAR (Fondo di Ateneo per la Ricerca, University of Milano-Bicocca) and by the project grant SysBioNet, Italian Roadmap Research Infrastructures 2012.

References

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9:258–267

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314:1565–1568

Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355:619–627

Ask M, Mapelli V, Höck H, Olsson L, Bettiga M (2013) Engineering glutathione biosynthesis of *Saccharomyces cerevisiae* increases robustness to inhibitors in pretreated lignocellulosic materials. Microb Cell Fact 12:87

Baez A, Cho KM, Liao JC (2011) High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with in situ product removal. Appl Microbiol Biotechnol 90:1681–1690

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247

Blazeck J, Hill A, Liu L, Knight R, Miller J, Pan A, Otoupal P, Alper HS (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. Nat Commun 5:3131

Chang JG, Yang DM, Chang WH, Chow LP, Chan WL, Lin HH, Huang HD, Chang YS, Hung CH, Yang WK (2011) Small molecule amiloride modulates oncogenic RNA alternative splicing to devitalize human cancer cells. PLoS One 6, e18643

Csörgo B, Fehér T, Tímár E, Blattner FR, Pósfai G (2012) Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. Microb Cell Fact 11:11

Danino T, Mondragón-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. Nature 463:326–330

Decker CJ, Parker R (2012) P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol 4:a012286

Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333:470–474

Delebecque CJ, Silver PA, Lindner AB (2012) Designing and using RNA scaffolds to assemble proteins in vivo. Nat Protoc 7:1797–1807

Deribe YL, Pawson T, Dikic I (2010) Post-translational modifications in signal integration. Nat Struct Mol Biol 17:666–672

Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF, Kassen R (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol Ecol 21:2058–2077

Dragosits M, Mattanovich D (2013) Adaptive laboratory evolution – principles and applications for biotechnology. Microb Cell Fact 12:64

Dunlop MJ, Keasling JD, Mukhopadhyay A (2010) A model for improving microbial biofuel production using a synthetic feedback loop. Syst Synth Biol 4:95–104

Ebert MS, Sharp PA (2012) Roles for microRNAs in conferring robustness to biological processes. Cell 149:515–524

Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T (2011) Circuitry linking the Csr and stringent response global regulatory systems. Mol Microbiol 80:1561–1580

Farmer WR, Liao JC (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat Biotechnol 18:533–537

Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, Hou YM, Jensen LJ (2012) Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol 8:572

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56

Herranz H, Cohen SM (2010) MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. Genes Dev 24:1339–1344

Inui M, Martello G, Piccolo S (2010) MicroRNA control of signal transduction. Nat Rev Mol Cell Biol 11:252–263

Kato M, Slack FJ (2013) Ageing and the small, non-coding RNA world. Ageing Res Rev 12:429–435

Kim HS, Kim NR, Yang J, Choi W (2011) Identification of novel genes responsible for ethanol and/or thermotolerance by transposon mutagenesis in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 91:1159–1172

Kohlhapp FJ, Mitra AK, Lengyel E, Peter ME (2015) MicroRNAs as mediators and communicators between cancer cells and the tumor microenvironment. Oncogene 1–12. doi:10.1038/onc.2015.89

Krisko A, Radman M (2013) Phenotypic and genetic consequences of protein damage. PLoS Genet 9, e1003810

Kwok CK, Tang Y, Assmann SM, Bevilacqua PC (2015) The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends Biochem Sci 40:221–232

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342:357–360

Lanza AM, Curran KA, Rey LG, Alper HS (2014) A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst Biol 8:33

Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW (2009) A microRNA imparts robustness against environmental fluctuation during development. Cell 137:273–282

Lian J, Chao R, Zhao H (2014) Metabolic engineering of a *Saccharomyces cerevisiae* strain capable of simultaneously utilizing glucose and galactose to produce enantiopure (2R,3R)-butanediol. Metab Eng 23:92–99

Liu C, Fu X, Liu L, Ren X, Chau CK, Li S, Xiang L, Zeng H, Chen G, Tang LH, Lenz P, Cui X, Huang W, Hwa T, Huang JD (2011) Sequential establishment of stripe patterns in an expanding cell population. Science 334:238–241

Ma NJ, Moonan DW, Isaacs FJ (2014) Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. Nat Protoc 9:2285–2300

Mansell TJ, Warner JR, Gill RT (2013) Trackable multiplex recombineering for gene-trait mapping in *E. coli*. Methods Mol Biol 985:223–246

Marietou A, Nguyen AT, Allen EE, Bartlett DH (2014) Adaptive laboratory evolution of *Escherichia coli* K-12 MG1655 for growth at high hydrostatic pressure. Front Microbiol 5:749

Martani F, Fossati T, Posteri R, Signori L, Porro D, Branduardi P (2013) Different response to acetic acid stress in *Saccharomyces cerevisiae* wild-type and l-ascorbic acid-producing strains. Yeast 30:365–378

May T, Okabe S (2011) Enterobactin is required for biofilm development in reduced-genome *Escherichia coli*. Environ Microbiol 13:3149–3162

McKee AE, Rutherford BJ, Chivian DC, Baidoo EK, Juminaga D, Kuo D, Benke PI, Dietrich JA, Ma SM, Arkin AP, Petzold CJ, Adams PD, Keasling JD, Chhabra SR (2012) Manipulation of the carbon storage regulator system for metabolite remodeling and biofuel production in *Escherichia coli*. Microb Cell Fact 11:79

McKenna R, Thompson B, Pugh S, Nielsen DR (2014) Rational and combinatorial approaches to engineering styrene production by *Saccharomyces cerevisia*e. Microb Cell Fact 13:123

Michener JK, Thodey K, Liang JC, Smolke CD (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. Metab Eng 14:212–222

Mitchell SF, Parker R (2014) Principles and properties of eukaryotic mRNPs. Mol Cell 54:547–558

Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 14:454–459

Okino S, Noburyu R, Suda M, Jojima T, Inui M, Yukawa H (2008) An efficient succinic acid production process in a metabolically engineered *Corynebacterium glutamicum* strain. Appl Microbiol Biotechnol 81:459–464

Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Jiang H, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496:528–532

Pál C, Papp B, Pósfai G (2014) The dawn of evolutionary genome engineering. Nat Rev Genet 15:504–512

Patakova P, Linhova M, Rychtera M, Paulova L, Melzoch K (2013) Novel and neglected issues of acetone–butanol–ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol Adv 31:58–67

Patzschke A, Steiger MG, Holz C, Lang C, Mattanovich D, Sauer M (2015) Enhanced glutathione production by evolutionary engineering of *Saccharomyces cerevisiae* strains. Biotechnol J. doi:10.1002/biot.201400809

Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24:1027–1032

Porro D, Branduardi P, Sauer M, Mattanovich D (2014) Old obstacles and new horizons for microbial chemical production. Curr Opin Biotechnol 30:101–106

Prindle A, Samayoa P, Razinkov I, Danino T, Tsimring LS, Hasty J (2012) A sensing array of radically coupled genetic 'biopixels'. Nature 481:39–44

Puig S, Querol A, Barrio E, Pérez-Ortín JE (2000) Mitotic recombination and genetic changes in *Saccharomyces cerevisia*e during wine fermentation. Appl Environ Microbiol 66:2057–2061

Ryu YS, Biswas RK, Shin K, Parisutham V, Kim SM, Lee SK (2014) A simple and effective method for construction of *Escherichia coli* strains proficient for genome engineering. PLoS One 9, e94266

Sedlak M, Ho NW (2004) Production of ethanol from cellulosic biomass hydrolysates using genetically engineered *Saccharomyces yeast* capable of cofermenting glucose and xylose. Appl Biochem Biotechnol 113–116:403–416

Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, Liao JC (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. Appl Environ Microbiol 77:2905–2915

Shi S, Valle-Rodríguez JO, Siewers V, Nielsen J (2014) Engineering of chromosomal wax ester synthase integrated *Saccharomyces cerevisiae* mutants for improved biosynthesis of fatty acid ethyl esters. Biotechnol Bioeng 111:1740–1747

Shong J, Jimenez Diaz MR, Collins CH (2012) Towards synthetic microbial consortia for bioprocessing. Curr Opin Biotechnol 23:798–802

Solomon KV, Sanders TM, Prather KL (2012) A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng 14:661–671

Suzuki Y, Assad-Garcia N, Kostylev M, Noskov VN, Wise KS, Karas BJ, Stam J, Montague MG, Hanly TJ, Enriquez NJ, Ramon A, Goldgof GM, Richter RA, Vashee S, Chuang RY, Winzeler EA, Hutchison CA, Gibson DG, Smith HO, Glass JI, Venter JC (2015) Bacterial genome reduction using the progressive clustering of deletions via yeast sexual cycling. Genome Res 25:435–444

Sydor T, Schaffer S, Boles E (2010) Considerable increase in resveratrol production by recombinant industrial yeast strains with use of rich medium. Appl Environ Microbiol 76:3361–3363

Tan FR, Dai LC, Wu B, Qin H, Shui ZX, Wang JL, Zhu QL, Hu QC, Ruan ZY, He MX (2015) Improving furfural tolerance of *Zymomonas mobilis* by rewiring a sigma factor RpoD protein. Appl Microbiol Biotechnol 99(12):5363–5371

Tilloy V, Ortiz-Julien A, Dequin S (2014) Reduction of ethanol yield and improvement of glycerol formation by adaptive evolution of the wine yeast *Saccharomyces cerevisiae* under hyperosmotic conditions. Appl Environ Microbiol 80:2623–2632

Van Dien S (2013) From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. Curr Opin Biotechnol 24:1061–1068

Venayak N, Anesiadis N, Cluett WR, Mahadevan R (2015) Engineering metabolism through dynamic control. Curr Opin Biotechnol 34:142–152

Vishnoi M, Narula J, Devi SN, Dao HA, Igoshin OA, Fujita M (2013) Triggering sporulation in *Bacillus subtilis* with artificial two-component systems reveals the importance of proper Spo0A activation dynamics. Mol Microbiol 90:181–194

Walters R, Parker R (2014) Quality control: is there quality control of localized mRNAs? J Cell Biol 204:863–868

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898

Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21:319–346

Wegner A, Meiser J, Weindl D, Hiller K (2014) How metabolites modulate metabolic flux. Curr Opin Biotechnol 34:16–22

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109:111–118

Winkler JD, Kao KC (2014) Recent advances in the evolutionary engineering of industrial biocatalysts. Genomics 104:406–411

Xie D, Jackson EN, Zhu Q (2015) Sustainable source of omega-3 eicosapentaenoic acid from metabolically engineered *Yarrowia lipolytica*: from fundamental research to commercial production. Appl Microbiol Biotechnol 99:1599–1610

Xue X, Wang T, Jiang P, Shao Y, Zhou M, Zhong L, Wu R, Zhou J, Xia H, Zhao G, Qin Z (2014) MEGA (Multiple Essential Genes Assembling) deletion and replacement method for genome reduction in *Escherichia coli*. ACS Synth Biol. doi:10.1021/sb500324p

Yamada R, Hasunuma T, Kondo A (2013) Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. Biotechnol Adv 31:754–763

Yang T, Rao Z, Zhang X, Xu M, Xu Z, Yang ST (2013) Improved production of 2,3-butanediol in *Bacillus amyloliquefaciens* by over-expression of glyceraldehyde-3-phosphate dehydrogenase and 2,3-butanediol dehydrogenase. PLoS One 8, e76149

You L, Cox RS, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. Nature 428:868–871

Zhao H, Li J, Han B, Li X, Chen J (2014) Improvement of oxidative stress tolerance in *Saccharomyces cerevisiae* through global transcription machinery engineering. J Ind Microbiol Biotechnol 41:869–878

Microbial Platform Cells for Synthetic Biology

6

Dong-Woo Lee and Sang Jun Lee

Contents

6.1 Introduction 230
6.2 The Constituents of Synthetic Cells 231
6.2.1 Biological Parts 231
6.2.2 DNA Sequencing 232
6.2.3 Gene Synthesis 232
6.2.4 Genetic Circuits 233
6.3 Conventional Microbial Platform Cells 234
6.3.1 Genome Editing 234
6.3.2 Minimal Genomes 235
6.3.3 Classical Chassis Cells 236
6.3.4 Products of Synthetic Microbial Cells 237
6.4 Extremophiles as New Synthetic Cellular Platforms 239
6.4.1 Diversity of Extremophiles 239
6.4.2 Recent Applications of Metabolic Engineering and Synthetic Biology in Extremophiles 240
6.4.3 Extremophiles as Novel Platform Cells for Biofuel Production 242
6.5 Perspectives 243
References 245

D.-W. Lee, Ph.D. (✉)
School of Applied Biosciences, Kyungpook National University, Daegu 41566, Korea
e-mail: leehicam@knu.ac.kr

S.J. Lee, Ph.D. (✉)
Biosystems and Bioengineering Program, University of Science and Technology (UST), Daejeon 34113, Korea

Infection and Immunity Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Korea
e-mail: leesj@kribb.re.kr

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_6

Abstract

In synthetic biology, biological systems are considered assemblies of modules, parts, and devices that are controlled by synthetic circuits. In biotechnological aspects, the area of synthetic biology includes (1) the synthesis of genes, genomes, and life; (2) the synthesis of new metabolic pathways; (3) modular engineering of proteins; and (4) the development of biosensors, most of which can be verified in synthetic cells. To achieve the various goals of synthetic biology, we need to know how to design and engineer microbial synthetic cells properly. To this end, there have been frequent discussions of chassis cells, which may be used in versatile ways. We need to define the meaning of chassis cells clearly and determine how to choose and produce microbial platform cells for industrial applications in the future. In this chapter, we summarize (1) the constituents of synthetic cells, (2) the current progress of development of conventional platform cells, and (3) the development of new cellular platforms based on extremophiles.

6.1 Introduction

Fuels, which encompass a variety of feedstocks for refinery chemicals and pharmaceuticals, have been obtained from fossil resources for more than 100 years. The need for and production of petroleum and its accessory products have increased as a consequence of the growing global population. Because fossil resources are finite, obtaining the necessary amounts of energy and building blocks from fossil oils will not be possible in the near future. Much attention has been paid to alternative renewable energy sources from natural resources such as sunlight, wind, tides, geothermal heat, and biomass. Biotechnologists have attempted to solve the problems associated with sustainable energy from biomass using biological knowledge and technologies.

In the 1970s, when the principles of DNA sequencing (Maxam and Gilbert 1977; Sanger et al. 1977) and recombinant DNA technology (Cohen et al. 1973; Nathans and Smith 1975) were first reported, scientists started to manipulate genes and produce upgraded cells with intended functions. However, technical bottlenecks associated with the creation of synthetic cells still existed at that time. Nowadays, the remarkable development of nucleotide sequencing methods has enabled the generation of enormous amounts of genome information that is freely available for use in genetic and cellular engineering. In addition, cutting-edge gene synthesis (Hughes et al. 2011) and genome assembly technologies (Gibson 2009) have enabled researchers to synthesize cells with novel functions easily. During the last decade, scientists have initiated the field of "synthetic biology" for the production of bioenergy, refinery chemicals, and pharmaceuticals from synthetic cells.

In synthetic biology, biological systems are recognized as an assembly of modules, parts, and devices that are controlled by synthetic circuits. In biotechnological aspects, the area of synthetic biology includes (1) the synthesis of genes,

genomes, and life; (2) the synthesis of new metabolic pathways; (3) modular engineering of proteins; and (4) the development of biosensors, most of which can be verified in synthetic cells. To achieve the various goals of synthetic biology, we need to know how to design and engineer microbial synthetic cells properly. To this end, there have been frequent discussions about chassis cells, which may be used in versatile ways; however, we need to define the meaning of chassis cells clearly and determine how to choose and generate microbial platform cells for future industrial applications. In this chapter, we summarize (1) the constituents of synthetic cells, (2) the current progress of conventional platform cells, and (3) the development of new cellular platforms based on extremophiles.

6.2 The Constituents of Synthetic Cells

6.2.1 Biological Parts

To facilitate the engineering of biological systems and the generation of artificial or synthetic cells with relevant functions, biological parts can be combined and inserted into the cells in a "Lego-like" assembly (Shetty et al. 2008). Most biological parts can be defined as DNA sequences, because new functions or phenotypes can be conferred to synthetic cells by simply introducing new genes or altering the native genome. Biological parts operating at transcriptional levels include promoters, enhancers, operators, riboswitches, attenuators, and transcriptional terminators. At the translational level, ribosomal binding sites, protein-coding sequences, and various reporter genes can be used. The sizes of biological parts range from a few to several hundred base pairs, and to ensure their proper operation in synthetic cells, the nucleotide sequences of biological parts should not be altered. For example, in the case of promoters in the galactose operon, even single nucleotide exchanges can alter the promoter strength dramatically and turn transcription on or off (Lewis et al. 2014).

The Registry of Standard Biological Parts is available online (http://parts.igem.org/). Assembly of basic biological parts can be used to generate new and larger composite parts, devices, and systems with complex functions, which can be introduced into synthetic cells using a variety of customized vectors. Because biological systems are modular, in the larger scope of biological parts, receptors, transducers, and effectors in signal transduction pathways can be included and engineered for synthetic biology purposes (Furukawa and Hohmann 2013; Morey et al. 2011).

Practical biological parts can be generated using the enormous quantities of information in genomic databases. However, if intact biological parts are not suitable and/or available, synthetic parts required for the intended regulation of biological systems can be generated. For example, synthetic small RNAs can be designed to identify and modulate the expression of target genes for metabolic engineering in *Escherichia coli* (Na et al. 2013; Yoo et al. 2013). In addition, proteins such as transcription factors or enzymes with desired ligand or substrate

specificities can be generated using molecular evolution technology, which is also useful for the engineering of synthetic parts (Eriksen et al. 2014). Molecular engineering using proteins or RNAs as scaffolds can also be applied to pathway engineering in synthetic cells (Conrado et al. 2008; Delebecque et al. 2011). For example, mevalonate pathway enzymes have been organized on intracellular scaffolds to enable the efficient production of isoprenoids (Dueber et al. 2009).

6.2.2 DNA Sequencing

The essential regions of DNA sequences were determined by Sanger DNA sequencing for approximately 20 years before the emergence of massive pyrosequencing techniques in 1998 (Ronaghi et al. 1998). Since then, a number of high-throughput sequencing platforms, such as Genome Analyzer, SOLiD, CGA, and PacBio RS, have been produced by various companies, and new technologies such as nanopore sequencing are now emerging (Myllykangas et al. 2012). The development of cutting-edge sequencing technologies enabled researchers to generate genome sequence databases, which are reservoirs of genetic information for cellular constituents and are indispensable to the identification and harnessing of specific sequences for modules, parts, clusters, and circuitries of relevant synthetic cells (Yang et al. 2009; Hachiya et al. 2011).

Microbial genome projects have provided tremendous amounts of genetic information regarding culturable microbes; however, most microbes in nature are still non-culturable (Zhang et al. 2012). Recently, single-cell genome sequencing technology has been used for genomic studies of non-culturable bacteria and metagenomic researches (Embree et al. 2014; Rodrigue et al. 2009; Yilmaz and Singh 2012); this technology uses non-PCR-based multiple displacement amplification by the highly processive phage Φ29 DNA polymerase (Blanco et al. 1989). In addition to supplying sequence information for biological parts and devices, DNA sequencing also aids the verification of fabricated synthetic systems and facilitates the rapid identification of genetic changes in microbial populations (Eren et al. 2013).

6.2.3 Gene Synthesis

With the development of current technologies and the reduced cost of de novo gene synthesis, a 1 kb synthetic gene can be commercially supplied within a week. Methods for the chemical synthesis of oligonucleotides (Yu et al. 2012), PCR- or ligation-mediated assembly (Xiong et al. 2008), and error correction (Ma et al. 2012) have been developed for rapid gene synthesis (Hughes et al. 2011).

The genetic code is almost universal with few variations and is compatible with all living things. However, the codon usage frequency is different in an individual genus or species. Prior to the development of gene synthesis technologies, *E. coli* strains containing extra copies of tRNAs for rare codons were engineered for the

efficient expression of heterologous genes encoding polypeptides. Nowadays, instead of engineering host strains, DNA sequences encoding biological parts of interest can be designed and synthesized using various algorithms that optimize codon usage and mRNA secondary structure (Gould et al. 2014). Although codon usage optimization is an attractive approach for high-level protein expression that maintains the primary polypeptide structure in synthetic cells, it does not preclude the possibility of generating proteins with altered conformations and substrate specificities caused by differential folding, which might be attributed to ribosome stalling at synonymous codons (Kimchi-Sarfaty et al. 2007; Tsai et al. 2008).

6.2.4 Genetic Circuits

The most distinct advantages of biological systems are their abilities to self-replicate and self-regulate. Once synthetic microbial cells have been established, their characteristic behaviors and genetic information, which are programmed in the genetic circuitry, can be maintained for generations. At the early stage of genetic circuit development, toggles and oscillators were designed and operated in cells (Judd et al. 2000). Toggle switches are synthetic bistable gene regulatory networks that consist of promoters and repressors arranged in a mutually inhibitory fashion (Gardner et al. 2000). Repressilator, a genetic oscillator, has three repressible promoters, each of which transcribes the repressor of one of the other promoters (Elowitz and Leibler 2000). Because biological systems are modular and hierarchical, synthetic genetic circuits are expected to be able to function exquisitely in complex biological systems. However, variations of switching in these circuits can occur due to stochastic effects caused by intrinsic and extrinsic noise (Sprinzak and Elowitz 2005). A strong bistable toggle switch with an ultrasensitive response was created by linking distinct positive feedback loops (Chang et al. 2010). Recently, a synthetic transcriptional AND gate was designed using split T7 RNA polymerase (Schaerli et al. 2014; Shis and Bennett 2013). Because two or three T7 RNA polymerase fragments can form a single functional protein complex (Segall-Shapiro et al. 2014), the T7 promoter can only be activated in the presence of two input signals, which could function as a band-pass filter in cells (Schaerli et al. 2014).

The application range of genetic circuits is diverse. In synthetic biology, biosensors are one of the most frequently used types of genetic circuit. Various biological modules and parts can be combined to construct biosensors capable of detecting metabolites (Zhang and Keasling 2011), products of enzymes (Choi et al. 2014), and hazardous materials (Kim et al. 2014b) in the fields of biochemical engineering, environment research, and healthcare. For metabolic engineering, metabolite sensors can be designed based on metabolite-responsive transcription factors or riboswitches for screening of microbes with desired performance. A riboswitch-based synthetic selection device has been developed to facilitate the evolution of high producers of L-lysine or L-tryptophan (Yang et al. 2013). In addition, a synthetic RNA device known as a Riboselector is a useful tool for strain

improvement in the areas of biorefinery and biomedicine (Jang et al. 2015). A high-throughput generalized enzyme screening system has been developed using artificial genetic circuits (Choi et al. 2014). A modified DmpR transcription factor can recognize *p*-nitrophenol that is often found as covalently linked to various chromogenic substrates (Pavel et al. 1994). The formation of non-metabolizable *p*-nitrophenol is proportional to the in vivo activity of the relevant enzyme, which can intracellularly turn on DmpR-mediated transcription; hence, this system can be used to expedite the screening and evolution of enzymes using fluorescence-activated cell sorting (Jeong et al. 2012).

A modular cell-based biosensor has been engineered to sense and integrate multiple hazardous heavy metal signals using AND logic circuits that can function as a genetic filter and amplifier (Wang et al. 2013). The AND gated sensor can filter out nonspecific input signals, thereby enhancing the selectivity of sensing and the signal-to-noise ratio. In addition, heavy metal microbial sensors have been grown in microfluidic chambers in a chemostat-like manner to enable the sensitive detection of molecular analytes on a chip (Kim et al. 2014b).

6.3 Conventional Microbial Platform Cells

6.3.1 Genome Editing

To achieve the aims of synthetic biology, biological parts or circuits should be properly introduced and driven in chassis cells from *E. coli* through mammals. Microbial cells are considered synthetic biological workhorses because they have several advantages over cells derived from higher organisms, including well-known characteristics, familiar genetic tools, cheaper growth medium, and high-yield fermentation, all of which have been described in numerous reports (Cronan 2001; Zonneveld 1986). Microbial synthetic platform cells can be improved through two genomic approaches; specifically, the genomes of classical platform cells such as *E. coli* and *Saccharomyces cerevisiae* can be edited and reduced by forward genome engineering, and new synthetic platform cells can be constructed by a reverse engineering approach based on the de novo synthesis of a genome.

The concept of synthetic biology emerged in the 1970s, and a number of synthetic biology tools have subsequently been developed in *E. coli*, one of the most popular microbial chassis cells (Szybalski and Skalka 1978). Although various genetic tools using plasmids and phages have been developed, the *E. coli* genome project, which was completed in 1997 (Blattner et al. 1997), accelerated the method of highly sophisticated genomic engineering. Single or multiple bases at specific sites in a genome can be edited by oligonucleotide-directed recombineering (Oppenheim et al. 2004; Swingle et al. 2010; Thomason et al. 2014). Multiplex automated genome engineering is a powerful high-throughput genome editing technique (Wang et al. 2009) that enables the simultaneous modification of multiple locations on a chromosome in a single cell or across a population of cells via allelic replacement. For example, in a previous study, this technique enabled the site-

specific replacement of 314 stop codons with synonymous TAA codons in the *E. coli* genome (Isaacs et al. 2011). In addition, in *Bacillus* species, site-specific point mutations can be introduced into the genome using PCR-amplified fragments consisting of the desired mutations, a drug-resistant gene, and a counter-selectable marker (Fabret et al. 2002; Dong and Zhang 2014).

Chromosomal genes can be inactivated by PCR-mediated recombineering, which facilitates the generation of specific gene mutations for microbial functional genomics (Datsenko and Wanner 2000). The Keio collection, comprising knockouts of individual genes in *E. coli*, provided an insight into the limit of the minimal genome and the gene sets that are essential for cell growth (Baba et al. 2006; Yamamoto et al. 2009), which is indispensable for the design of synthetic genomes in microbial chassis cells with better performance.

6.3.2 Minimal Genomes

In the context of structural genomics, even after researchers design and engineer the desired genome, deformation and instability can be caused by intrinsic mobile DNA elements such as insertion sequence elements, transposases, defective phages, integrases, and site-specific recombinases (Frost et al. 2005). To stabilize the genome and metabolism of a chassis cell, "jumping" DNA elements must be deleted. Indeed, the use of a Tn5-targeted Cre/IoxP excision method (Yu et al. 2002) and high-throughput systematic genome mutagenesis by the Tn5 transposon (Kang et al. 2004) has enabled the creation of individual *E. coli* deletion and insertion mutants without loss of normal growth. However, multiple rounds of DNA elimination could leave adjacent target sequences of recombination, such as loxP and FRT, causing the limitation of further excision in the genome.

In addition to affecting genomic stability, mobile elements in chromosomal DNA can also jump into plasmid DNA vaccines. These vaccines can be stably overproduced in highly dense cultures of modified *E. coli* BL21 strains that are deficient in the *endA* and *recA* genes (Phue et al. 2008). To produce homogenous DNA vaccines lacking mobile DNA fragments, surrogate hosts with clean genomes should be harnessed. Recently, the scar-absent excision method was developed to generate the clean and minimal genome of the *E. coli* MDS42 strain lacking mobile DNA elements or recombination target sequences (Posfai et al. 2006); this outcome was accomplished by lambda red-mediated chromosomal integration of I-SceI restriction sites at specific locations, followed by homologous recombinational repair of chromosome double-strand breaks induced by I-SceI cleavage (Posfai et al. 1999). Lipopolysaccharide-encoding genes are not essential and have been deleted from the genome of microbial cells in which pyrogen-free therapeutic proteins or DNA vaccines can be produced safely (Posfai et al. 2006).

6.3.3 Classical Chassis Cells

As chassis cells refer to the conceptual extension of the living cells, "plug-in" and "plug-out" genetic circuits and new functionalities should be predictably carried out in this biological chassis. On the other hand, platform cells can be defined as aim-specific strains containing a set of biological components shared among different microorganisms. A wealth of biological data has made *E. coli* an attractive model organism to study and develop as a microbial platform. Recombinant human insulin was first made in *E. coli* via recombinant DNA technology and was approved by the US Food and Drug Administration (Johnson 1983). Since then, *E. coli* has been used widely as a therapeutic protein producer. The lack of posttranslational modifications of proteins produced in *E. coli* can be overcome by a number of approaches. For example, selective in vitro glycosylation has enabled the generation of homogeneous glycoforms of human erythropoietin (Macmillan et al. 2001), and human growth hormone can be secreted into the oxidized periplasm of *E. coli* to promote disulfide bridge formation (de Oliveira et al. 1999). In addition, *E. coli* metabolism can be modified or harnessed to produce valuable primary and secondary metabolites (Fig. 6.1). Organic acids (Forster and Gescher 2014), branched amino acids (Park and Lee 2010), and isoprenoids (Harada and Misawa 2009) can also be produced in *E. coli*. An in silico metabolic network of *E. coli* has been constructed and used for metabolic flux analysis (Orth et al. 2011). By stoichiometric coupling of biomass formation to the production of desired products such as succinate, lactate, and 1,3-propanediol, the in silico OptKnock framework can be used to suggest gene deletion strategies that lead to the overproduction of metabolites in *E. coli* cells (Burgard et al. 2003).

The genus *Corynebacterium* was originally developed as a platform cell for the sole purpose of L-glutamate production (Kimura 2002). Accumulated knowledge of the genetics and physiology of this genus expedited the production of various metabolites in *C. glutamicum* (Matano et al. 2014) and enabled the consolidated production of L-lysine from raw cornstarch by a strain of *C. glutamicum* that secretes heterologous α-amylase (Tateno et al. 2007). For bio-based production of polyamides, commonly known as nylons, *C. glutamicum* was reprogrammed to produce large quantities of 1,5-diaminopentane, also called cadaverine, which could be converted from L-lysine by codon-optimized *E. coli* L-lysine decarboxylase (Kind et al. 2014). Recently, engineering of the 2-keto acid metabolic pathway enabled *C. crenatum* to produce higher alcohols such as isobutanol, 2-methyl-1-butanol, and 3-methyl-1-butanol for biofuel, using hydrolysates of duckweed as feedstock (Su et al. 2015).

From the view point of synthetic biology, eukaryotic microbial cells have several advantages over prokaryotic cells. First, their higher recombination frequencies facilitate high-throughput genome assembly in yeast cells (Chao et al. 2015). Second, glycosylation is often essential for the folding and activity of native proteins, and pharmaceutical proteins such as therapeutic antibodies can be produced via humanized glycosylation in yeast cells (Vogl et al. 2013). Third, the yields of complex metabolites such as artemisinic acid are often higher in yeast

cells than prokaryotic cells (Paddon and Keasling 2014). Despite these advantages, intracellular compartmentalization caused by multi-membrane structures is a problem that must be overcome in eukaryotic systems. For example, in *Aspergillus niger*, the mitochondrial expression of aconitase and cis-aconitate decarboxylase, rather than the cytosolic expression of these proteins, enhances the production of itaconic acid, an unsaturated dicarboxylic acid that has a high potential as a biochemical building block (Blumhoff et al. 2013).

The generation of new microbial cells with shuffled genomes is classically achieved by protoplast fusion (Leja et al. 2011); this method has been used frequently to improve the phenotypes of industrial strains, such as increasing the acid and glucose tolerances of *Lactobacillus* (Gong et al. 2009). Genome shuffling is cost-effective and is not considered to generate genetically modified organisms in the food industry. In addition, the "domino method" of bottom-up genome assembly has been used to build a 134.5 kb rice chloroplast genome using a *Bacillus subtilis* genome vector (Itaya et al. 2008). In 2010, Gibson et al. reported the design, synthesis, and assembly of a 1.08 Mb *Mycoplasma mycoides* genome via homologous recombination in yeast and described its transplantation into recipient cells to create an artificial cell controlled by a chemically synthesized genome (Gibson et al. 2010a). Mitochondrial genomic DNA was also chemically synthesized along the same lines (Gibson et al. 2010b). Recently, a functional eukaryotic chromosome, synIII, was synthesized based on the native *S. cerevisiae* chromosome III (Annaluru et al. 2014). In the near future, it may be possible to generate any prokaryotic or eukaryotic microbial cell with a synthetic genome.

6.3.4 Products of Synthetic Microbial Cells

One of the most important goals in synthetic biology is the biosynthesis of biofuels, refinery chemicals, and pharmaceuticals (Martin et al. 2003; Keasling 2010; Huo et al. 2011; Zhang et al. 2011). For decades, microbial cells have been regarded as cellular factories that are capable of generating valuable products from cheaper raw materials. From ancient times, microbial cells have been harnessed for food fermentation to produce alcohols, organic acids, and other compounds; however, because of the exhaustion of fossil resources, their purposes have now changed to the production of biofuel (Zabed et al. 2014; Durre 2007) and building blocks for biopolymers (Sauer et al. 2008; Jung et al. 2010). Alcohols can be synthesized via CoA-dependent pathways, as typified by acetone-butanol-ethanol fermentation in *Clostridial* strains (Fig. 6.1). *E. coli* has been engineered to convert amino acids to alcohols via keto acid pathways (Atsumi et al. 2008), allowing protein waste to be used as a biofuel feedstock (Huo et al. 2011). Biodiesels can also be produced via microbial fatty acid pathways (Steen et al. 2010), and mevalonate or non-mevalonate pathways for isoprenoid precursors have been used for the production of various substitutes for jet fuel (Peralta-Yahya et al. 2011; Wang et al. 2010). Recently, even short chain alkanes, also called gasoline, have been synthesized biochemically in *E. coli* (Choi and Lee 2013).

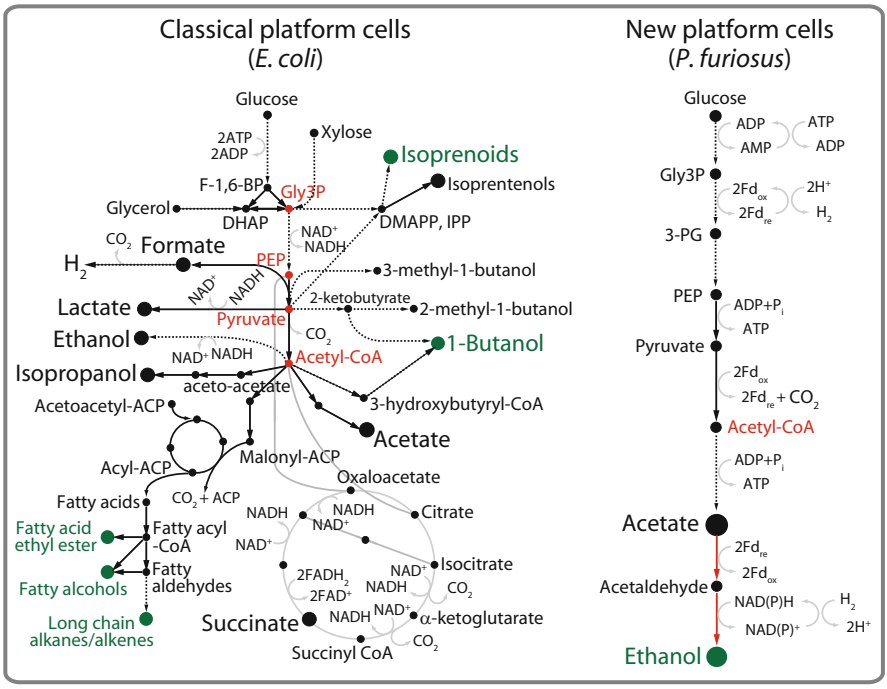


Fig. 6.1 Metabolic pathways for the production of biofuels in *E. coli* (*left panel*) and *P. furiosus* (*right panel*). The single *arrows* represent the one-step conversions, and the *dashed lines* represent multiple reaction steps. The *red circles* and *arrows* represent the metabolic intermediates and synthetic reactions, respectively. The *green circles* represent high-value products produced via synthetic pathways in the corresponding platform cells

A number of reports have described the engineering of metabolic pathways that are tightly coupled to the cellular energy and redox balance, which is one of the key factors that determines the yields and productivities of biofuels and fermentation products (Johnson and Schmidt-Dannert 2008; Trinh et al. 2008; Portnoy et al. 2010; Lan and Liao 2011, 2012; Kim et al. 2013). In addition to process development, the optimization of cellular energy and redox status should be taken into account to increase the yields and productivities of cell factories.

Many drugs are rarely found in nature and are difficult to synthesize chemically due to their complex structures. For many decades, varieties of natural biochemicals such as antibiotics have been produced in microorganisms using conventional random mutagenesis techniques (Baltz 2001). More recently, cutting-edge synthetic biology technologies have enabled the production of complex pharmaceuticals, including antibiotics, anticancer and antitumor agents, antioxidants, and antiparasitic drugs, in heterologous microbial cells such as *E. coli*, *B. subtilis, Streptomyces* sp., and *S. cerevisiae* (Lee et al. 2009). In *E. coli*, the production of artemisinic acid, the precursor of an antimalarial drug,

was attempted by heterologous expression of eukaryotic cytochrome P450 (Chang et al. 2007). The marginal conversion of amorphadiene to artemisinic acid in *E. coli* has been successfully overcome in baker's yeast, indicating that the expression of some plant enzymes is better in eukaryotic platform cells than prokaryotic systems (Paddon and Keasling 2014). Microbial platforms of natural and semisynthetic opioids have also been developed in *S. cerevisiae* (Thodey et al. 2014).

6.4 Extremophiles as New Synthetic Cellular Platforms

6.4.1 Diversity of Extremophiles

Extremophilic microorganisms (extremophiles) are ubiquitous organisms that thrive optimally in extreme environments characterized by high temperatures (−2 to 15 °C or 60–110 °C), high or low pH (<4 or >9), high ionic strength (2–5 M NaCl), or high pressure (>38 MPa) (Rothschild and Mancinelli 2001). Across several decades, extremophiles encompassing prokaryotic and archaeal microorganisms have played crucial roles in basic evolutionary and phylogenetic research into the origin of life (Woese et al. 1990; Sarmiento et al. 2013). Comparisons of 16S and 18S rRNA gene sequences demonstrated that the majority of extremophiles are members of the archaea, although extremophilic bacteria (e.g., the genus of *Aquifex* and *Thermotoga*) are also known (Brown and Doolittle 1997). Enzymes produced by extremophiles (extremozymes), which have extreme stability, have also served as novel biocatalysts in a variety of biotechnological applications (Egorova and Antranikian 2005; Demirjian et al. 2001). Previously, many researchers in this field have focused on the isolation and identification of extremophiles and extremozymes, understanding of their structure-function relationships at a molecular level, and enzyme engineering to create novel biocatalysts with enhanced stability and altered specificity (Vetriani et al. 1998; Vieille and Zeikus 2001). In recent years, remarkable advances in fermentation technology and synthetic biology tools, including metabolic engineering for extremophiles that were previously regarded as unamenable to conventional genetic manipulation, have expanded their roles in the food, laundry, and chemical industries (Hawkins et al. 2013; Brunecky et al. 2013).

During the past decade, improved genome sequence availability, newfound bioinformatics approaches, cheap DNA synthesis, and rapid assembly technologies have enabled researchers to rewire the metabolic pathways of extremophiles and/or build synthetic parts into their endogenous pathways and genetic circuits. Accordingly, extremophiles can be used as an alternative source of microbial platform cells for the production of biofuels because they are stable and active under industrial conditions (Hawkins et al. 2013). Moreover, several extremophiles, including archaea, have novel metabolic pathways that have not been identified in other prokaryotic microorganisms (Liu et al. 2012; Sato and Atomi 2011); these pathways may be attractive and beneficial for the industrial use of extremophiles with novel activities and applications. Thus, despite the relatively short history and

lack of knowledge of extremophiles compared with other conventional platform microorganisms (e.g., *E. coli*, *B. subtilis*, *S. cerevisiae*, etc.), such an analogical transition in the field of extremophiles has begun with emerging synthetic biology and metabolic engineering approaches aimed at utilizing the parts-to-circuits-to-function paradigm.

Since the late 1960s, the isolation of extremophiles such as *Thermus aquaticus* (Brock and Freeze 1969) and *Methanococcus jannaschii* (Jones et al. 1983) suggested the existence of a tremendous variety of microorganisms in extreme environments that might exceed expectations. Subsequently, DNA sequencing technology-aided metagenomic approaches revealed that this scenario might be true. In 1996, sequencing of the first archaeal genome (that of *M. jannaschii*) helped researchers to characterize the very early forms of life on Earth and encouraged investigations of the unusual features of extremophiles that have not been observed in other microorganisms (Bult et al. 1996). However, most extremophiles are difficult to culture and isolate in large quantities; hence, they are not attractive as platform cells for conventional applications or as model organisms for basic studies. During the last decade, the development of next-generation sequencing technology has enabled researchers to access genetic information for numerous examples of unusual forms of life. Subsequently, hundreds of genome sequences of extremophiles have been generated, and many novel genetic contexts have been functionally annotated in more detail than ever before. The prominent features of the genomic information from extremophiles unveiled the early forms of energy metabolism (e.g., a modified Embden-Meyerhof pathway and archaea-specific autotrophic carbon fixation) (Sakuraba et al. 2004; Siebers and Schönheit 2005; Berg et al. 2010), the mechanism of evolutionary adaptation to environmental changes (Takami et al. 2004), the occurrence of lateral gene transfer (Nelson et al. 1999), and the features of archeoviruses (Pina et al. 2011), all of which are applicable to industrial demands, including the production of biofuels and high-value metabolites. Recently, the successful production of biohydrogen (H_2) and cellulosic biomass-based ethanol using hyperthermophilic archaea and a combination of metabolic engineering, synthetic biology, and systems biology shed light on extremophiles as novel platform microbial cells that are even better than conventional platform cells for specific aims, including the sustainable consolidated production of biofuels (Hawkins et al. 2013).

6.4.2 Recent Applications of Metabolic Engineering and Synthetic Biology in Extremophiles

Compared with the use of microorganisms such as *E. coli* and yeast in metabolic engineering and synthetic biology, advances in the use of extremophiles for these applications have been slow. Nevertheless, extremophile synthetic biology is progressing brightly. Because synthetic part identification and de novo synthesis are possible, rapid synthesis methods, including the harnessing of extremophile homologous recombination and gene disruption systems, can replace traditional

cloning and enable combinatorial circuit assembly (Thiel et al. 2014; Sato et al. 2003). Very recently, rapid and combinatorial strain engineering in extremophiles has led to the development of new platform cells for viable biofuel production (Farkas et al. 2012). Indeed, several proof-of-concept approaches have been performed using complex extremophiles as the synthetic platform cells (Lipscomb et al. 2014; Keller et al. 2013). These successful case studies have driven the extremophile synthetic biology field to tackle products that are more practical and have high economic value.

Very recently, the chemolithoautotrophic archaeon *Metallosphaera sedula*, which has an optimum growth temperature of 73 °C (pH 2.0), was investigated to examine the metabolic pathway connections between CO_2 fixation and central metabolism (Auernik et al. 2008; Hawkins et al. 2014). For growth on metal sulfides or H_2, *M. sedula* employs the unique carbon fixation 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle, which consists of 13 enzymes (Berg et al. 2007; Han et al. 2012). On the basis of a metabolic flux analysis (Estelmann et al. 2011), an in vivo growth analysis using a gas-intensive bioreactor system together with an in vitro assay using heterologously expressed metabolic enzymes in the 3HP/4HB cycle demonstrated that the carbon fixation and tricarboxylic acid cycles are responsible for the generation of central carbon precursor molecules via acetyl-CoA (Hawkins et al. 2014; Huber et al. 2008). This study provided the framework for altering the balance between cataplerotic and anaplerotic reactions, thereby aiding future metabolic engineering efforts aimed at producing biofuels and organic chemicals. Furthermore, the *Pyrococcus furiosus* ΔACSIα strain was recently engineered to generate 3HP from CO_2 and acetyl-CoA by the heterologous expression of three enzymes from the CO_2 fixation cycle of *M. sedula*. The autotrophically engineered strain exhibits a threefold increase in specific 3HP production. This study indicates that the hyperthermophilic archaeon will be a good model system for future metabolic engineering that is comparable to previous mesophilic synthetic platform cells (Thorgersen et al. 2014).

Lignocellulosic crops are regarded as sustainable and renewable, but they have some technically limiting factors, including recalcitrance of lignin-containing cellulosic biomass, resistance to enzymatic hydrolysis, and the presence of five carbon sugars (Demain et al. 2005; Lee et al. 2013). To overcome these hurdles, consolidated bioprocessing of plant biomass has been developed using cellulolytic *Clostridium thermocellum* (Wiegel et al. 1985) for the production of ethanol (Lynd et al. 2008). In this regard, extremely thermophilic *Caldicellulosiruptor* strains could have a high potential for the degradation of cellulosic biomass for biofuel production (Blumer-Schuette et al. 2011). The distinct advantages of in vivo high-temperature production of biofuels include reduced contamination risk, facilitated removal of volatile products, and a wide temperature range to modulate and balance both the engineered pathway and the host's metabolism (Basen et al. 2012; Keller et al. 2013; Abdel-Banat et al. 2010). Previously, the advanced biofuel butanol was produced via metabolic engineering of carbon and nitrogen metabolic modules in mesophiles (Huo et al. 2011; Atsumi et al. 2008a). Very recently, a hybrid synthetic pathway assembled from genes derived from three different sources was introduced

into a hyperthermophilic organism and used to convert glucose to 1-butanol via acetyl-CoA (Keller et al. 2015). Another remarkable feature is that the butanol production temperature was 30 °C lower than the optimal archaeal growth temperature, indicating that the module could also be used in thermophiles capable of degrading plant biomass.

6.4.3 Extremophiles as Novel Platform Cells for Biofuel Production

The heterotrophic and hyperthermophilic archaeon *P. furiosus* is a new addition to the growing list of genetically tractable microorganisms that are suitable for metabolic engineering to produce liquid fuels and industrial chemicals. During the last decade, *P. furiosus*, which grows optimally at 100 °C (Fiala and Stetter 1986), has been examined as a microbial platform for the production of H_2. This chemolithoorganotrophic archaeon can obtain energy from the conversion of carbohydrates to acetate, CO_2, and H_2 by a respiratory membrane-bound hydrogenase (MBH), which is encoded by a 14-gene operon (Schut et al. 2013). MBH-mediated energy conservation occurs through H_2 production coupled to the oxidation of reduced ferredoxin and the generation of a Na^+ ion gradient (Sapra et al. 2003). Very recently, a remarkable study was performed using the intact, functional, 14-subunit respiratory membrane-bound [NiFe]-hydrogenase complex of *P. furiosus* for H_2 production (McTernan et al. 2014). Under anaerobic conditions, an intact respiratory MBH complex with a mass of 310 kDa was expressed and purified to couple H_2 production to the oxidation of ferredoxin and pumping of Na^+ ions. Subsequently, the recombinant catalytically active cytoplasmic subcomplex of MBH was successfully engineered and expressed in *P. furiosus* by differential transcription of the *MBH* operon (McTernan et al. 2015). This study suggested that archaeal energy-transducing complexes can be engineered as novel modules for other conventional platform cells. Overall, the recent successful attempts described above represent significant achievements in the use of extremophiles as alternative microbial platform cells.

Previously, bioethanol production was performed using a few engineered mesophilic platform cells (Nielsen and Keasling 2011; Peralta-Yahya et al. 2011). In addition, the *AdhA* gene encoding bacterial alcohol dehydrogenase has been inserted into an engineered archaeon (*P. furiosus*) to promote the successful production of bioethanol (Basen et al. 2014) (Fig. 6.1). The synthetic aldehyde ferredoxin oxidoreductase/AdhA pathway can also be used for thermophilic bioalcohol production from exogenously added aliphatic and aromatic carboxylic acids, implying that the most primitive of all life forms could serve as a potentially game-changing platform cell factory. Thermophiles such as *Geobacillus* strains may be other potential host organisms for high-temperature processes. In contrast to previous attempts to engineer mesophilic host microorganisms for efficient cellulose degradation through thermal adaptation (Lynd et al. 2008; Lee et al. 2013), Suzuki et al. (2013) developed a *G. kaustophilus* HTA426 strain with the ability to degrade cellulose paper and insoluble starch at high temperatures by expressing the

genes encoding cellulase and α-amylase using a new inducible expression system. In addition, the use of extremophiles over mesophiles as host organisms has enabled researchers to circumvent the overall low efficiency of photosynthesis and the production of sugar intermediates. For example, in a recent study, heterologous expression of five genes involved in the carbon fixation cycle enabled *P. furiosus* to grow autotrophically using CO_2, which it is incapable of doing naturally (Keller et al. 2013). Such a unique approach demonstrated that the engineered hyperthermophilic archaeon is able to use H_2 gas and incorporate CO_2 into 3HP, indicating that the low potential reducing power from non-biomass sources, such as hydrogen gas or electricity, can be redirected to reduce CO_2 directly into products. In addition, chromosomal integration of the heterologous 18-subunit membrane-bound formate H_2 lyase complex into *P. furiosus* allowed the host cell to efficiently convert formate to H_2 gas as a source of a reductant for the microbial production of biofuels (Lipscomb et al. 2014). As such, the use of extremophiles could circumvent limiting factors for H_2 production such as low solubility and poor gas mass transfer rates. In addition, such an accomplishment highlights the potential versatility of extremophiles for metabolic engineering applications.

6.5 Perspectives

In the era of synthetic biology, researchers can construct microbial platform cells using genome editing, synthesis, and assembly technologies. In addition, to achieve the goals of synthetic biology, biological parts can be introduced into platform cells using sophisticated genetic tools (Fig. 6.2). Thus, the integration of synthetic biology tools with advanced metabolic engineering has enabled the exploitation of extremophiles as new platform cells. Regardless of the use of classical or new synthetic platform cells, one of the most critical factors associated with controlling genes and regulatory circuitries is the requirement to minimize stochastic effects and maximize the intended cellular performance. However, in frequent cases, artificial biological systems are unlikely to be operated properly due to unknown reasons, which may be solved by using multi-omics technology. Such a systematic analysis involving the genome, transcriptome, proteome, and metabolome, known as "systems and synthetic biology," can lead to successful synthetic biology approaches in the future.

With regard to the sustainability of synthetic cells, genome dynamics should be considered carefully (Umenhoffer et al. 2010). In particular, cells are evolving continuously and will continue to do so, even in non-stressful environments due to spontaneous mutations caused by errors in DNA replication, spontaneous lesions, and mobile genetic elements (Foster 2007). When microorganisms encounter environmental stresses that retard their cellular growth, microbial adaptation may occur through genomic mutations beyond our knowledge and/or observations in the laboratory or in nature (Kim et al. 2014a). In addition, under certain adverse conditions, synthetic cells could be altered due to adaptive genome mutations,

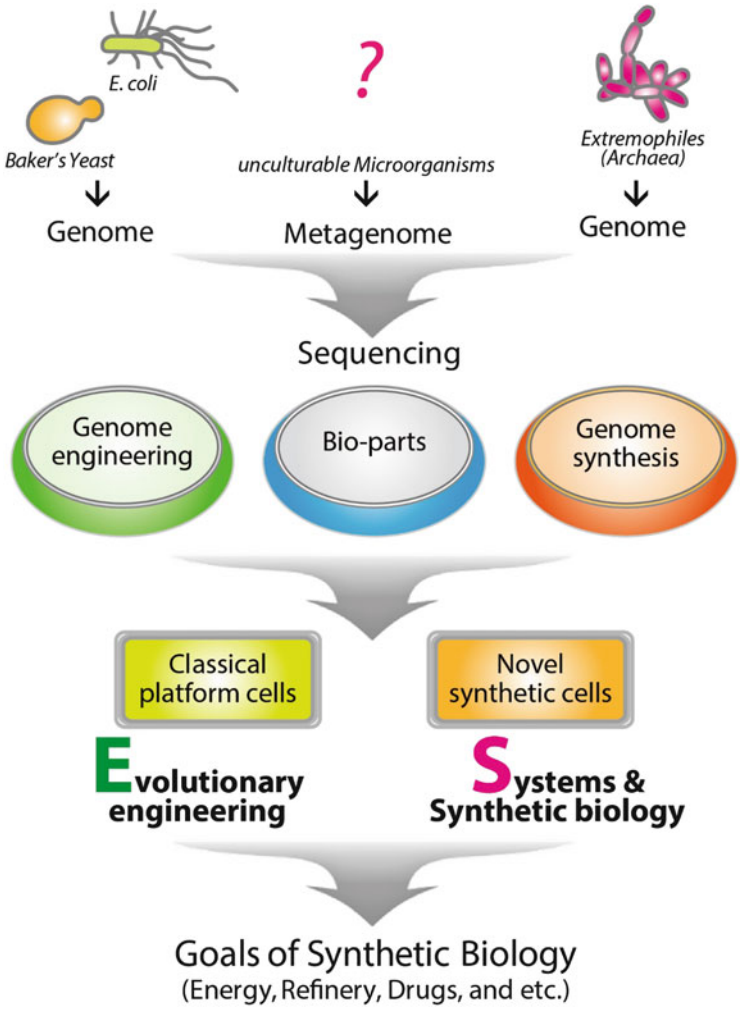


Fig. 6.2 Schematic diagram of the workflow for the development of microbial platform cells in the field of synthetic biology

leading to a defect in their intended function. Consequently, adaptive mutants could arise suddenly and dominate the original microbial population.

Overall, the optimization and improvement of biological parts and circuitries in microbial synthetic cells could be achieved by repeated rounds of adaptation and genome sequence verification; this so-called evolutionary engineering will provide well-matched biological parts and robust microbial platform cells.

References

Abdel-Banat BM, Hoshida H, Ano A, Nonklang S, Akada R (2010) High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microbiol Biotechnol 85(4):861–867. doi:10.1007/s00253-009-2248-5

Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, Yeluru A, Murugan S, Kandavelou K, Agier N, Fischer G, Yang K, Martin JA, Bilgel M, Bohutski P, Boulier KM, Capaldo BJ, Chang J, Charoen K, Choi WJ, Deng P, DiCarlo JE, Doong J, Dunn J, Feinberg JI, Fernandez C, Floria CE, Gladowski D, Hadidi P, Ishizuka I, Jabbari J, Lau CY, Lee PA, Li S, Lin D, Linder ME, Ling J, Liu J, Liu J, London M, Ma H, Mao J, McDade JE, McMillan A, Moore AM, Oh WC, Ouyang Y, Patel R, Paul M, Paulsen LC, Qiu J, Rhee A, Rubashkin MG, Soh IY, Sotuyo NE, Srinivas V, Suarez A, Wong A, Wong R, Xie WR, Xu Y, Yu AT, Koszul R, Bader JS, Boeke JD, Chandrasegaran S (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344(6179):55–58. doi:10.1126/science.1249252

Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451(7174):86–89. doi:10.1038/nature06450

Auernik KS, Maezato Y, Blum PH, Kelly RM (2008) The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon Metallosphaera sedula provides insights into bioleaching-associated metabolism. Appl Environ Microbiol 74(3):682–692. doi:10.1128/AEM.02019-07

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008. doi:10.1038/msb4100050

Baltz RH (2001) Genetic methods and strategies for secondary metabolite yield improvement in actinomycetes. Antonie Van Leeuwenhoek 79(3–4):251–259

Basen M, Sun J, Adams MW (2012) Engineering a hyperthermophilic archaeon for temperature-dependent product formation. mBio 3(2):e00053-00012. doi:10.1128/mBio.00053-12

Basen M, Schut GJ, Nguyen DM, Lipscomb GL, Benn RA, Prybol CJ, Vaccaro BJ, Poole FL II, Kelly RM, Adams MW (2014) Single gene insertion drives bioalcohol production by a thermophilic archaeon. Proc Natl Acad Sci U S A 111(49):17618–17623. doi:10.1073/pnas.1413789111

Berg IA, Kockelkorn D, Buckel W, Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science 318(5857):1782–1786. doi:10.1126/science.1149976

Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hugler M, Alber BE, Fuchs G (2010) Autotrophic carbon fixation in archaea. Nat Rev Microbiol 8(6):447–460. doi:10.1038/nrmicro2365

Blanco L, Bernad A, Lazaro JM, Martin G, Garmendia C, Salas M (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem 264(15):8935–8940

Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of Escherichia coli K-12. Science 277(5331):1453–1462

Blumer-Schuette SE, Ozdemir I, Mistry D, Lucas S, Lapidus A, Cheng JF, Goodwin LA, Pitluck S, Land ML, Hauser LJ, Woyke T, Mikhailova N, Pati A, Kyrpides NC, Ivanova N, Detter JC, Walston-Davenport K, Han S, Adams MW, Kelly RM (2011) Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria Caldicellulosiruptor hydrothermalis, Caldicellulosiruptor kristjanssonii, Caldicellulosiruptor kronotskyensis,

Caldicellulosiruptor owensensis, and Caldicellulosiruptor lactoaceticus. J Bacteriol 193 (6):1483–1484. doi:10.1128/JB.01515-10
Blumhoff ML, Steiger MG, Mattanovich D, Sauer M (2013) Targeting enzymes to the right compartment: metabolic engineering for itaconic acid production by Aspergillus niger. Metab Eng 19:26–32. doi:10.1016/j.ymben.2013.05.003
Brock TD, Freeze H (1969) Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. J Bacteriol 98(1):289–297
Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 61(4):456–502
Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MW, Lunin VV, Himmel ME, Bomble YJ (2013) Revealing nature's cellulase diversity: the digestion mechanism of Caldicellulosiruptor bescii CelA. Science 342 (6165):1513–1516. doi:10.1126/science.1244273
Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science 273(5278):1058–1073
Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84 (6):647–657. doi:10.1002/bit.10803
Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD (2007) Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. Nat Chem Biol 3(5):274–277. doi:10.1038/nchembio875
Chang DE, Leung S, Atkinson MR, Reifler A, Forger D, Ninfa AJ (2010) Building biological memory by linking positive feedback loops. Proc Natl Acad Sci U S A 107(1):175–180. doi:10.1073/pnas.0908314107
Chao R, Yuan Y, Zhao H (2015) Recent advances in DNA assembly technologies. FEMS Yeast Res 15(1):1–9
Choi YJ, Lee SY (2013) Microbial production of short-chain alkanes. Nature 502(7472):571–574. doi:10.1038/nature12536
Choi SL, Rha E, Lee SJ, Kim H, Kwon K, Jeong YS, Rhee YH, Song JJ, Kim HS, Lee SG (2014) Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. ACS Synth Biol 3(3):163–171. doi:10.1021/sb400112u
Cohen SN, Chang ACY, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A 70(11):3240–3244
Conrado RJ, Varner JD, DeLisa MP (2008) Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol 19(5):492–499. doi:10.1016/j.copbio.2008.07.006
Cronan JE (2001) Escherichia coli as an experimental organism. In: eLS. Wiley doi:10.1002/9780470015902.a0002026.pub2
Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640–6645. doi:10.1073/pnas.120163297
de Oliveira JE, Soares CR, Peroni CN, Gimbo E, Camargo IM, Morganti L, Bellini MH, Affonso R, Arkaten RR, Bartolini P, Ribela MT (1999) High-yield purification of biosynthetic human growth hormone secreted in Escherichia coli periplasmic space. J Chromatogr A 852 (2):441–450
Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333(6041):470–474. doi:10.1126/science.1206938
Demain AL, Newcomb M, Wu JH (2005) Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 69(1):124–154. doi:10.1128/MMBR.69.1.124-154.2005

Demirjian DC, Morıs-Varas F, Cassidy CS (2001) Enzymes from extremophiles. Curr Opin Chem Biol 5(2):144–151. doi:10.1016/S1367-5931(00)00183-6
Dong H, Zhang D (2014) Current development in genetic engineering strategies of Bacillus species. Microb Cell Fact 13:63. doi:10.1186/1475-2859-13-63
Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27(8):753–759. doi:10.1038/nbt.1557
Durre P (2007) Biobutanol: an attractive biofuel. Biotechnol J 2(12):1525–1534. doi:10.1002/biot.200700168
Egorova K, Antranikian G (2005) Industrial relevance of thermophilic Archaea. Curr Opin Microbiol 8(6):649-655. doi:10.1016/j.mib.2005.10.015
Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338. doi:10.1038/35002125
Embree M, Nagarajan H, Movahedi N, Chitsaz H, Zengler K (2014) Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. ISME J 8(4):757–767. doi:10.1038/ismej.2013.187
Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol/Br Ecol Soc 4(12). doi:10.1111/2041-210x.12114
Eriksen DT, Lian J, Zhao H (2014) Protein design for pathway engineering. J Struct Biol 185 (2):234–242. doi:10.1016/j.jsb.2013.03.011
Estelmann S, Hugler M, Eisenreich W, Werner K, Berg IA, Ramos-Vera WH, Say RF, Kockelkorn D, Gad'on N, Fuchs G (2011) Labeling and enzyme studies of the central carbon metabolism in Metallosphaera sedula. J Bacteriol 193(5):1191–1200. doi:10.1128/JB.01155-10
Fabret C, Ehrlich SD, Noirot P (2002) A new mutation delivery system for genome-scale approaches in Bacillus subtilis. Mol Microbiol 46(1):25–36
Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW, Westpheling J (2012) Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants. Appl Environ Microbiol 78(13):4669–4676. doi:10.1128/AEM.00936-12
Fiala G, Stetter K (1986) Pyrococcus furiosus sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch Microbiol 145(1):56–61. doi:10.1007/BF00413027
Forster AH, Gescher J (2014) Metabolic engineering of Escherichia coli for production of mixed-acid fermentation end products. Front Bioeng Biotechnol 2:16. doi:10.3389/fbioe.2014.00016
Foster PL (2007) Stress-induced mutagenesis in bacteria. Crit Rev Biochem Mol Biol 42 (5):373–397. doi:10.1080/10409230701648494
Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3(9):722–732. doi:10.1038/nrmicro1235
Furukawa K, Hohmann S (2013) Synthetic biology: lessons from engineering yeast MAPK signalling pathways. Mol Microbiol 88(1):5–19. doi:10.1111/mmi.12174
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342. doi:10.1038/35002131
Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res 37(20):6984–6990. doi:10.1093/nar/gkp687
Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA III, Smith HO, Venter JC (2010a) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56. doi:10.1126/science.1190719
Gibson DG, Smith HO, Hutchison CA III, Venter JC, Merryman C (2010b) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7(11):901–903. doi:10.1038/nmeth.1515

Gong J, Zheng H, Wu Z, Chen T, Zhao X (2009) Genome shuffling: progress and applications for phenotype improvement. Biotechnol Adv 27(6):996–1005

Gould N, Hendy O, Papamichail D (2014) Computational tools and algorithms for designing customized synthetic genes. Front Bioeng Biotechnol 2. doi:10.3389/fbioe.2014.00041

Hachiya T, Irie N, Arakawa K (2011) Effective utilization of next-generation sequencers. Jikken Igaku 29(11):1789–1796

Han Y, Hawkins AS, Adams MW, Kelly RM (2012) Epimerase (Msed_0639) and mutase (Msed_0638 and Msed_2055) convert (S)-methylmalonyl-coenzyme A (CoA) to succinyl-CoA in the Metallosphaera sedula 3-hydroxypropionate/4-hydroxybutyrate cycle. Appl Environ Microbiol 78(17):6194–6202. doi:10.1128/AEM.01312-12

Harada H, Misawa N (2009) Novel approaches and achievements in biosynthesis of functional isoprenoids in Escherichia coli. Appl Microbiol Biotechnol 84(6):1021–1031. doi:10.1007/s00253-009-2166-6

Hawkins AS, McTernan PM, Lian H, Kelly RM, Adams MWW (2013) Biological conversion of carbon dioxide and hydrogen into liquid fuels and industrial chemicals. Curr Opin Biotechnol 24(3):376–384. doi:10.1016/j.copbio.2013.02.017

Hawkins AB, Adams MW, Kelly RM (2014) Conversion of 4-hydroxybutyrate to acetyl coenzyme A and its anapleurosis in the Metallosphaera sedula 3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway. Appl Environ Microbiol 80(8):2536–2545. doi:10.1128/AEM.04146-13

Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, Eisenreich W, Fuchs G (2008) A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis. Proc Natl Acad Sci U S A 105(22):7851–7856. doi:10.1073/pnas.0801043105

Hughes RA, Miklos AE, Ellington AD (2011) Gene synthesis: methods and applications. Methods Enzymol 498:277–309. doi:10.1016/b978-0-12-385120-8.00012-7

Huo YX, Cho KM, Rivera JG, Monte E, Shen CR, Yan Y, Liao JC (2011) Conversion of proteins into biofuels by engineering nitrogen flux. Nat Biotechnol 29(4):346–351. doi:10.1038/nbt.1789

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333(6040):348–353. doi:333/6040/348 [pii] 10.1126/science.1205822

Itaya M, Fujita K, Kuroki A, Tsuge K (2008) Bottom-up genome assembly using the Bacillus subtilis genome vector. Nat Methods 5(1):41–43. doi:10.1038/nmeth1143

Jang S, Yang J, Seo SW, Jung GY (2015) Riboselector: riboswitch-based synthetic selection device to expedite evolution of metabolite-producing microorganisms. Methods Enzymol 550:341–362. doi:10.1016/bs.mie.2014.10.039

Jeong YS, Choi SL, Kyeong HH, Kim JH, Kim EJ, Pan JG, Rha E, Song JJ, Lee SG, Kim HS (2012) High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. Protein Eng Des Selection 25(11):725–731. doi:10.1093/protein/gzs071

Johnson IS (1983) Human insulin from recombinant DNA technology. Science 219 (4585):632–637

Johnson ET, Schmidt-Dannert C (2008) Light-energy conversion in engineered microorganisms. Trends Biotechnol 26(12):682–689. doi:S0167-7799(08)00239-4 [pii] 10.1016/j.tibtech.2008.09.002

Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS (1983) Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch Microbiol 136 (4):254–261. doi:10.1007/BF00425213

Judd EM, Laub MT, McAdams HH (2000) Toggles and oscillators: new genetic circuit designs. BioEssays 22(6):507–509. doi:10.1002/(sici)1521-1878(200006)22:6<507::aid-bies3>3.0.co;2-0

Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of Escherichia coli for the production of polylactic acid and its copolymers. Biotechnol Bioeng 105(1):161–171. doi:10.1002/bit.22548

Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, Blattner FR (2004) Systematic mutagenesis of the *Escherichia coli* genome. J Bacteriol 186(15):4921–4930. doi:10.1128/JB.186.15.4921-4930.2004 186/15/4921 [pii]

Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330 (6009):1355–1358. doi:330/6009/1355 [pii] 10.1126/science.1193990

Keller MW, Schut GJ, Lipscomb GL, Menon AL, Iwuchukwu IJ, Leuko TT, Thorgersen MP, Nixon WJ, Hawkins AS, Kelly RM, Adams MW (2013) Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. Proc Natl Acad Sci U S A 110(15):5840–5845. doi:10.1073/pnas.1222607110

Keller MW, Lipscomb GL, Loder AJ, Schut GJ, Kelly RM, Adams MWW (2015) A hybrid synthetic pathway for butanol production by a hyperthermophilic microbe. Metab Eng 27:101–106. doi:10.1016/j.ymben.2014.11.004

Kim HJ, Hou BK, Lee SG, Kim JS, Lee DW, Lee SJ (2013) Genome-wide analysis of redox reactions reveals metabolic engineering targets for D-lactate overproduction in Escherichia coli. Metab Eng 18:44–52. doi:10.1016/j.ymben.2013.03.004

Kim HJ, Jeong H, Hwang S, Lee MS, Lee YJ, Lee DW, Lee SJ (2014a) Short-term differential adaptation to anaerobic stress via genomic mutations by Escherichia coli strains K-12 and B lacking alcohol dehydrogenase. Front Microbiol 5:476. doi:10.3389/fmicb.2014.00476

Kim M, Lim JW, Kim HJ, Lee SK, Lee SJ, Kim T (2014b) Chemostat-like microfluidic platform for highly sensitive detection of heavy metal ions using microbial biosensors. Biosens Bioelectron 65c:257–264. doi:10.1016/j.bios.2014.10.028

Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315 (5811):525–528. doi:10.1126/science.1135308

Kimura E (2002) Triggering mechanism of L-glutamate overproduction by DtsR1 in coryneform bacteria. J Biosci Bioeng 94(6):545–551

Kind S, Neubauer S, Becker J, Yamamoto M, Volkert M, Abendroth G, Zelder O, Wittmann C (2014) From zero to hero – production of bio-based nylon from renewable resources using engineered Corynebacterium glutamicum. Metab Eng 25:113–123. doi:10.1016/j.ymben.2014.05.007

Lan EI, Liao JC (2011) Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. Metab Eng 13(4):353–363. doi:S1096-7176(11)00047-4 [pii] 10.1016/j.ymben.2011.04.004

Lan EI, Liao JC (2012) ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. Proc Natl Acad Sci U S A 109(16):6018–6023. doi:1200074109 [pii] 10.1073/pnas.1200074109

Lee SY, Kim HU, Park JH, Park JM, Kim TY (2009) Metabolic engineering of microorganisms: general strategies and drug production. Drug Discov Today 14(1–2):78–88. doi:10.1016/j.drudis.2008.08.004

Lee SJ, Lee SJ, Lee DW (2013) Design and development of synthetic microbial platform cells for bioenergy. Front Microbiol 4:92. doi:10.3389/fmicb.2013.00092

Leja K, Myszka K, Czaczyk K (2011) Genome shuffling: a method to improve biotechnological processes. Biotechnologia 92(4):345–351

Lewis DE, Le P, Adhya S (2014) Differential role of base pairs on gal promoters strength. J Mol Biol. doi:10.1016/j.jmb.2014.12.010

Lipscomb GL, Schut GJ, Thorgersen MP, Nixon WJ, Kelly RM, Adams MW (2014) Engineering hydrogen gas production from formate in a hyperthermophile by heterologous production of an 18-subunit membrane-bound complex. J Biol Chem 289(5):2873–2879. doi:10.1074/jbc.M113.530725

Liu Y, Beer LL, Whitman WB (2012) Sulfur metabolism in archaea reveals novel processes. Environ Microbiol 14(10):2632–2644. doi:10.1111/j.1462-2920.2012.02783.x

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE (2008) How biotech can transform biofuels. Nat Biotech 26(2):169–172. http://www.nature.com/nbt/journal/v26/n2/suppinfo/nbt0208-169_S1.html

Ma S, Saaem I, Tian J (2012) Error correction in gene synthesis technology. Trends Biotechnol 30 (3):147–154. doi:10.1016/j.tibtech.2011.10.002

Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL (2001) Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. Chem Biol 8(2):133–145

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21(7):796–802. doi:10.1038/nbt833 nbt833 [pii]

Matano C, Uhde A, Youn JW, Maeda T, Clermont L, Marin K, Kramer R, Wendisch VF, Seibold GM (2014) Engineering of Corynebacterium glutamicum for growth and L-lysine and lycopene production from N-acetyl-glucosamine. Appl Microbiol Biotechnol 98(12):5633–5643. doi:10.1007/s00253-014-5676-9

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74 (2):560–564

McTernan PM, Chandrayan SK, Wu CH, Vaccaro BJ, Lancaster WA, Yang Q, Fu D, Hura GL, Tainer JA, Adams MW (2014) Intact functional fourteen-subunit respiratory membrane-bound [NiFe]-hydrogenase complex of the hyperthermophilic archaeon Pyrococcus furiosus. J Biol Chem 289(28):19364–19372. doi:10.1074/jbc.M114.567255

McTernan PM, Chandrayan SK, Wu CH, Vaccaro BJ, Lancaster WA, Adams MW (2015) Engineering the respiratory membrane-bound hydrogenase of the hyperthermophilic archaeon Pyrococcus furiosus and characterization of the catalytically active cytoplasmic subcomplex. Protein Eng Des Selection 28(1):1–8. doi:10.1093/protein/gzu051

Morey KJ, Antunes MS, Albrecht KD, Bowen TA, Troupe JF, Havens KL, Medford JI (2011) Developing a synthetic signal transduction system in plants. Methods Enzymol 497(Synthetic Biology, Part A):581–602. doi:10.1016/B978-0-12-385075-1.00025-1

Myllykangas S, Buenrostro J, Ji H (2012) Overview of sequencing technology platforms. In: Rodríguez-Ezpeleta N, Hackenberg M, Aransay AM (eds) Bioinformatics for high throughput sequencing. Springer, New York, NY, pp 11–25. doi:10.1007/978-1-4614-0782-9_2

Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. Nat Biotechnol 31(2):170–174. doi:10.1038/nbt.2461

Nathans D, Smith HO (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. Annu Rev Biochem 44:273–293. doi:10.1146/annurev.bi.44.070175.001421

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature 399(6734):323–329. doi:10.1038/20601

Nielsen J, Keasling JD (2011) Synergies between synthetic biology and metabolic engineering. Nat Biotech 29(8):693–695

Oppenheim AB, Rattray AJ, Bubunenko M, Thomason LC, Court DL (2004) In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. Virology 319(2):185–189. doi:10.1016/j.virol.2003.11.007

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Mol Syst Biol 7:535. doi:10.1038/msb.2011.65

Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol 12(5):355–367. doi:10.1038/nrmicro3240

Park JH, Lee SY (2010) Fermentative production of branched chain amino acids: a focus on metabolic engineering. Appl Microbiol Biotechnol 85(3):491–506. doi:10.1007/s00253-009-2307-y

Pavel H, Forsman M, Shingler V (1994) An aromatic effector specificity mutant of the transcriptional regulator DmpR overcomes the growth constraints of Pseudomonas sp. strain CF600 on para-substituted methylphenols. J Bacteriol 176(24):7550–7557

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483. doi:10.1038/ncomms1494

Phue JN, Lee SJ, Trinh L, Shiloach J (2008) Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with Escherichia coli K (DH5alpha). Biotechnol Bioeng 101(4):831–836. doi:10.1002/bit.21973

Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. FEMS Microbiol Rev 35 (6):1035–1054. doi:10.1111/j.1574-6976.2011.00280.x

Portnoy VA, Scott DA, Lewis NE, Tarasova Y, Osterman AL, Palsson BO (2010) Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655. Appl Environ Microbiol 76(19):6529–6540. doi: AEM.01178-10 [pii] 10.1128/AEM.01178-10

Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR (1999) Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res 27 (22):4409–4415

Posfai G, Plunkett G, III, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. Science 312(5776):1044–1046. doi:1126439 [pii] 10.1126/science.1126439

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW (2009) Whole genome amplification and de novo assembly of single bacterial cells. PLoS One 4(9). doi:10.1371/journal.pone.0006864

Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. Science 281(5375):363, 365

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature 409 (6823):1092–1101. doi:10.1038/35059215

Sakuraba H, Goda S, Ohshima T (2004) Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. Chem Rec 3(5):281–287. doi:10.1002/tcr.10066

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74(12):5463–5467

Sapra R, Bagramyan K, Adams MW (2003) A simple energy-conserving system: proton reduction coupled to proton translocation. Proc Natl Acad Sci U S A 100(13):7545–7550. doi:10.1073/pnas.1331436100

Sarmiento F, Mrazek J, Whitman WB (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc Natl Acad Sci U S A 110(12):4726–4731. doi:10.1073/pnas.1220225110

Sato T, Atomi H (2011) Novel metabolic pathways in Archaea. Curr Opin Microbiol 14 (3):307–314. doi:10.1016/j.mib.2011.04.014

Sato T, Fukui T, Atomi H, Imanaka T (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. J Bacteriol 185 (1):210–220

Sauer M, Marx H, Mattanovich D (2008) Microbial production of 1,3-propanediol. Recent Pat Biotechnol 2(3):191–197

Schaerli Y, Gili M, Isalan M (2014) A split intein T7 RNA polymerase for transcriptional AND-logic. Nucleic Acids Res 42(19):12322–12328. doi:10.1093/nar/gku884

Schut GJ, Boyd ES, Peters JW, Adams MW (2013) The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. FEMS Microbiol Rev 37(2):182–203. doi:10.1111/j.1574-6976.2012.00346.x

Segall-Shapiro TH, Meyer AJ, Ellington AD, Sontag ED, Voigt CA (2014) A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. Mol Syst Biol 10:742. doi:10.15252/msb.20145299

Shetty RP, Endy D, Knight TF Jr (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5. doi:10.1186/1754-1611-2-5

Shis DL, Bennett MR (2013) Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. Proc Natl Acad Sci U S A 110(13):5028–5033. doi:10.1073/pnas.1220157110

Siebers B, Schönheit P (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Curr Opin Microbiol 8(6):695–705. doi:10.1016/j.mib.2005.10.014

Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438(7067):443–448. doi:10.1038/nature04335

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463(7280):559–562. http://www.nature.com/nature/journal/v463/n7280/suppinfo/nature08721_S1.html

Su H, Jiang J, Lu Q, Zhao Z, Xie T, Zhao H, Wang M (2015) Engineering Corynebacterium crenatum to produce higher alcohols for biofuel using hydrolysates of duckweed (Landoltia punctata) as feedstock. Microb Cell Fact 14:16. doi:10.1186/s12934-015-0199-3

Suzuki H, Yoshida K, Ohshima T (2013) Polysaccharide-degrading thermophiles generated by heterologous gene expression in Geobacillus kaustophilus HTA426. Appl Environ Microbiol 79(17):5151–5158. doi:10.1128/AEM.01506-13

Swingle B, Markel E, Costantino N, Bubunenko MG, Cartinhour S, Court DL (2010) Oligonucleotide recombination in Gram-negative bacteria. Mol Microbiol 75(1):138–148. doi:10.1111/j.1365-2958.2009.06976.x

Szybalski W, Skalka A (1978) Nobel prizes and restriction enzymes. Gene 4(3):181–182

Takami H, Takaki Y, Chee GJ, Nishi S, Shimamura S, Suzuki H, Matsui S, Uchiyama I (2004) Thermoadaptation trait revealed by the genome sequence of thermophilic Geobacillus kaustophilus. Nucleic Acids Res 32(21):6292–6303. doi:10.1093/nar/gkh970

Tateno T, Fukuda H, Kondo A (2007) Direct production of L-lysine from raw corn starch by Corynebacterium glutamicum secreting Streptococcus bovis alpha-amylase using cspB promoter and signal sequence. Appl Microbiol Biotechnol 77(3):533–541. doi:10.1007/s00253-007-1191-6

Thiel A, Michoud G, Moalic Y, Flament D, Jebbar M (2014) Genetic manipulations of the hyperthermophilic piezophilic archaeon Thermococcus barophilus. Appl Environ Microbiol 80(7):2299–2306. doi:10.1128/AEM.00084-14

Thodey K, Galanie S, Smolke CD (2014) A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol 10(10):837–844. doi:10.1038/nchembio.1613

Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL (2014) Recombineering: genetic engineering in bacteria using homologous recombination. Curr Protoc Mol Biol 106:1.16.1–1.16.39. doi:10.1002/0471142727.mb0116s106

Thorgersen MP, Lipscomb GL, Schut GJ, Kelly RM, Adams MWW (2014) Deletion of acetyl-CoA synthetases I and II increases production of 3-hydroxypropionate by the metabolically-engineered hyperthermophile Pyrococcus furiosus. Metab Eng 22:83–88. doi:10.1016/j.ymben.2013.12.006

Trinh CT, Unrean P, Srienc F (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74 (12):3634–3643. doi:AEM.02708-07 [pii] 10.1128/AEM.02708-07

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J Mol Biol 383(2):281–291. doi:10.1016/j.jmb.2008.08.012

Umenhoffer K, Feher T, Baliko G, Ayaydin F, Posfai J, Blattner FR, Posfai G (2010) Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. Microb Cell Fact 9:38. doi:10.1186/1475-2859-9-38

Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, Britton KL, Rice DW, Klump HH, Robb FT (1998) Protein thermostability above 100 degrees C: a key role for ionic interactions. Proc Natl Acad Sci U S A 95(21):12300–12305

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65(1):1–43. doi:10.1128/MMBR.65.1.1-43.2001

Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Curr Opin Biotechnol 24(6):1094–1101. doi:10.1016/j.copbio.2013.02.024

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460(7257):894–898. doi:nature08187 [pii] 10.1038/nature08187

Wang C, Yoon S-H, Shah AA, Chung Y-R, Kim J-Y, Choi E-S, Keasling JD, Kim S-W (2010) Farnesol production from Escherichia coli by harnessing the exogenous mevalonate pathway. Biotechnol Bioeng 107(3):421–429. doi:10.1002/bit.22831

Wang B, Barahona M, Buck M (2013) A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. Biosens Bioelectron 40 (1):368–376. doi:10.1016/j.bios.2012.08.011

Wiegel J, Mothershed CP, Puls J (1985) Differences in xylan degradation by various noncellulolytic thermophilic anaerobes and Clostridium thermocellum. Appl Environ Microbiol 49(3):656–659

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87(12):4576–4579

Xiong A-S, Peng R-H, Zhuang J, Liu J-G, Gao F, Chen J-M, Cheng Z-M, Yao Q-H (2008) Non-polymerase-cycling-assembly-based chemical gene synthesis: Strategies, methods, and progress. Biotechnol Adv 26(2):121–134. doi:10.1016/j.biotechadv.2007.10.001

Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, Furubayashi A, Kinjyo S, Dose H, Hasegawa M, Datsenko KA, Nakayashiki T, Tomita M, Wanner BL, Mori H (2009) Update on the Keio collection of Escherichia coli single-gene deletion mutants. Mol Syst Biol 5:335. doi:10.1038/msb.2009.92

Yang MQ, Athey BD, Arabnia HR, Sung AH, Liu Q, Yang JY, Mao J, Deng Y (2009) High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. BMC Genom 10 Suppl 1:I1

Yang J, Seo SW, Jang S, Shin SI, Lim CH, Roh TY, Jung GY (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. Nat Commun 4:1413. doi:10.1038/ncomms2404

Yilmaz S, Singh AK (2012) Single cell genome sequencing. Curr Opin Biotechnol 23(3):437–443. doi:10.1016/j.copbio.2011.11.018

Yoo SM, Na D, Lee SY (2013) Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli. Nat Protoc 8(9):1694–1707. doi:10.1038/nprot.2013.105

Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS, Kim SC (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat Biotechnol 20 (10):1018–1023. doi:10.1038/nbt740 nbt740 [pii]

Yu T, Bao X, Piao W, Peng J, Li W, Yang C, Xing M, Zhang Y, Qi J, Xu L, Xu L, Liu Q (2012) Recent patents on oligonucleotide synthesis and gene synthesis. Recent Pat DNA Gene Seq 6 (1):10–21

Zabed H, Faruq G, Sahu JN, Azirun MS, Hashim R, Boyce AN (2014) Bioethanol production from fermentable sugar juice. ScientificWorldJournal 2014:957102. doi:10.1155/2014/957102

Zhang F, Keasling J (2011) Biosensors and their applications in microbial metabolic engineering. Trends Microbiol 19(7):323–329. doi:S0966-842X(11)00092-8 [pii] 10.1016/j.tim.2011.05.003

Zhang F, Rodriguez S, Keasling JD (2011) Metabolic engineering of microbial pathways for advanced biofuels production. Curr Opin Biotechnol 22(6):775–783. doi:S0958-1669(11)00087-5 [pii] 10.1016/j.copbio.2011.04.024

Zhang X-m, Teng Z-c, Ren K (2012) Advances in 'non-culturable' bacteria. Anhui Nongye Kexue 40(10):5811–5812, 5819

Zonneveld BJM (1986) Cheap and simple yeast media. J Microbiol Methods 4(5–6):287–291. doi:10.1016/0167-7012(86)90040-0

Synthetic Biology Assisting Metabolic Pathway Engineering

7

Hans Marx, Stefan Pflügl, Diethard Mattanovich, and Michael Sauer

Contents

7.1 Introduction 256
7.1.1 Metabolic Engineering 256
7.1.2 Pathway Engineering 257
7.1.3 Getting Synthetic 258
7.2 Levels of Action 258
7.2.1 Genome Level 258
7.2.2 Transcription Level 260
7.2.3 Translation Level 266
7.2.4 Protein Level 269
7.2.5 Spatial Organisation 269
7.2.6 Combined Approaches 275

H. Marx (✉)
Department of Biotechnology, BOKU-VIBT University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

Christian Doppler Research Laboratory for Biotechnology of Glycerol, Muthgasse 18, 1190 Vienna, Austria
e-mail: hans.marx@boku.ac.at

S. Pflügl • D. Mattanovich
Department of Biotechnology, BOKU-VIBT University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, 1190 Vienna, Austria

M. Sauer
Department of Biotechnology, BOKU-VIBT University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

Christian Doppler Research Laboratory for Biotechnology of Glycerol, Muthgasse 18, 1190 Vienna, Austria

Austrian Centre of Industrial Biotechnology (ACIB GmbH), Muthgasse 11, 1190 Vienna, Austria

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_7

7.3 Commercial Application 275
7.4 Conclusions 276
References 277

Abstract

Synthetic biology is mainly focused on the design and construction of new biological parts. Assembly of these parts aids with the creation of synthetic biological pathways. This ultimately leads to an overlap of disciplines in which the synthetic parts are applied to a microbial production host and metabolic engineering for production of value-added compounds. This chapter intends to highlight the supportive input of synthetic biology parts, devices and modules to metabolic pathway engineering for the production of chemicals and fuels. Synthetic bricks find applications on the genome, transcriptome and proteome levels. By exemplifying several applications on the different levels, this chapter intends to give a brief overview of recent developments in the field; furthermore, upcoming developments in protein scaffolding and spatial organisation of pathways in bacterial microcompartments are discussed.

7.1 Introduction

7.1.1 Metabolic Engineering

Especially in the domain of pathway engineering for the production of metabolites, chemicals or biofuels, the two disciplines of metabolic engineering and synthetic biology interlock. The creation and engineering of metabolic pathways for the economically feasible production of desired compounds from renewable resources is a major aim of the discipline of metabolic engineering. In this context, the microbial cell is often compared to a chemical factory, which converts a given substrate via several consecutive steps into one or more products. The advantage of microbial cell factories is the incorporation of these steps into one "vessel" or chassis. Lee et al. (2012c) state the expectation that materials currently produced by the petrochemical industry will be replaced by increasingly diverse chemicals and materials produced in microbial cell factories.

One aim of metabolic engineering is to take an existing metabolic pathway, which leads to the desired product and to maximise the flux towards the product. In order to do so, bottlenecks in the pathway have to be identified, for example, by metabolic flux analysis and opened up by genetic or protein engineering to enhance enzyme activity and turnover of intermediate substrates towards the product. Competing pathways that drain substrates from the pathway of interest have to be deleted when possible, or at least the flux through the competing pathways has to be reduced to the minimum necessary in order to maintain the cell's viability.

The genes and enzymes directly involved in the cells metabolism only account for a small part of the open reading frames of a genome (roughly 12–20 %).

Much more account for the regulatory network of the cell (Nielsen 2003). However, looking at a metabolic map of an organism, it seems that certain branches of the network are rather far apart from each other, but one should not forget that even distant reactions can be interconnected by common cofactors such as ATP, or NADH and NADPH, which link the carbon metabolism to energy and redox metabolism. Furthermore, certain metabolites (like glucose, pyrophosphate, coenzyme A, NH_3) participate in more than ten different reactions, which depicts the tight connection even more. The overall cell metabolism therefore has to be seen as the entity of reactions converting carbon sources to cell building blocks as well as reactions supplying the cell with energy (ATP) and redox equivalents (NADH) (Nielsen 2003). Even more important for engineering the production of chemical compounds is the fact that all of the interesting metabolites and macromolecules produced by a microbial cell are derived from 12 precursor metabolites (Nielsen and Jewett 2008; Förster et al. 2003). These precursor metabolites represent sugar phosphates, intermediates of glycolysis and the TCA cycle, thereby depicting the field in which the pathway engineer has to intervene: the core of the tightly regulated central carbon metabolism.

7.1.2 Pathway Engineering

To gain a substantial product output, Yadav et al. (2012) have outlined the different levels to act on in order to enhance the flux through a metabolic pathway. One step is to increase the substrate uptake followed by eliminating competing pathways leading to the formation of undesired by-products. An enhanced flow of precursor metabolites as well as the sufficient supply of cofactors has to be provided. The activity of enzymes in the pathway has to be optimised to ensure a high flow rate from substrate via several intermediates to the product, and finally, the export of the product to the cultivation medium has to be ensured. Furthermore, when designing a metabolic pathway, one has to keep in mind that thermodynamics pose constraints on the direction in which the enzymatic reactions will lead (Noor et al. 2014), as well as the prerequisite that the oxidation state of the substrates will be balanced with the oxidation state of the products.

The transition of a classically metabolically engineered pathway to a pure synthetic pathway is fluent. Nielsen and Keasling (2011) illustrated this by describing on one hand a pathway in a naturally occurring production organism which has to be engineered by "classical methods" to eliminate by-product formation and the other aforementioned perturbations, on the other hand, a purely synthetic organism that is designed and constructed for the special production task only. Clearly, most of the approaches to date lie in between these two opposing examples (Lee et al. 2012b).

At the time being, successful examples for metabolic engineering are very much dependent on the host background used and therefore not directly applicable to other potential production organisms (Yadav et al. 2012). One of the major advantages that synthetic biology can add to metabolic engineering is the generalisation of parts and devices that perform independently of the host/chassis

background or at least the parts and devices are already characterised in the background of the desired chassis. This chapter intends to highlight the supportive input of synthetic biology parts, devices or systems to metabolic pathway engineering for the production of chemicals and fuels.

7.1.3 Getting Synthetic

Although synthetic biology is more focused on the design and construction of new biological parts, the assembly of these parts could or will result in the creation of a synthetic biological pathway. This ultimately leads to an overlap of disciplines in which the synthetic parts are applied to retrofit a microbial production host to the metabolic engineering goal of producing value-added compounds (Nielsen et al. 2014).

Synthetic biological parts are assembled into synthetic devices, which can be further assembled to modules. By the lack of a completely synthetic minimal cell, these modules have to be integrated in a host chassis, which has the advantage of providing the necessary building block metabolites, cofactors and energy. For further information on the construction of chassis cells, see the respective chapter in this book. In doing so, a completely synthetic pathway can be tested and will show the production of a new compound, most probably in a low amount at the beginning (König et al. 2013). The disadvantage is the not completely predictable interaction of the module with the chassis metabolic network. For this reason, synthetic biologists have to apply the same omics technologies as the metabolic engineer to predict or verify the effect of the module in the hosts' background and in a further step to increase the flux towards product formation. In this respect, systems biology will guide the design and optimisation of modules for synthetic pathway construction, and vice versa, the application of better devices and modules will allow systems biology to gain deeper insight into chassis metabolic networks.

7.2 Levels of Action

7.2.1 Genome Level

The production of lactic acid with *Saccharomyces cerevisiae* is an excellent example for an unwanted by-product pathway competing with the production pathway, tightly connected to the central carbon metabolism. By introducing a heterologous L-lactate dehydrogenase (L-LDH) in *S. cerevisiae*, pyruvate is directly converted to L-lactic acid. The competitor for pyruvate is the native pathway to ethanol starting with pyruvate decarboxylase (PDC) converting pyruvate into acetaldehyde which is further converted to ethanol. Engineering strategies which applied knockouts of the three PDC genes (PDC1, PDC5 and PDC6) present in *S. cerevisiae* observed a very reduced growth rate and fitness of the production strain (Sauer et al. 2010). Yamanishi and Matsuyama (2012) proposed a Cre-lox-

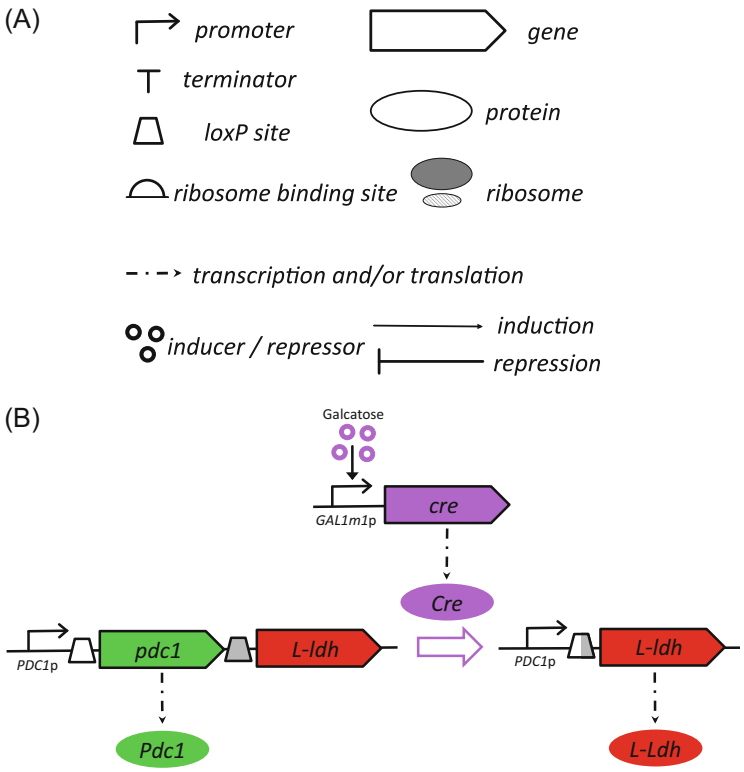


Fig. 7.1 (**a**) Symbols for parts and reactions used throughout this chapter. (**b**) Upon induction with galactose, the Cre recombinase is expressed, and the PDC gene flanked by loxP sites is excised from the genome. The excision of PDC results in the expression of L-LDH under the control of the native PDC promoter

based switch for dividing the process into a growth and a production phase. For this purpose, PDC5 was deleted from the host genome, PDC6 was left unchanged due to lower activity and PDC1 coding sequence was replaced by a cassette consisting of PDC1 coding sequence flanked by two loxP sites followed by the coding sequence of a L-LDH (Fig. 7.1). The Cre recombinase was also inserted on the genome under the control of an inducible promoter. Upon induction, Cre recombinase is expressed and cuts out the PDC1 coding sequence from the genome simultaneously putting the L-LDH coding sequence under the control of the PDC1 promoter, thereby inhibiting ethanol production and activating L-lactic acid production. Although lactate yields could be increased from 8.3 % to 85.4 %, the authors state that the control and induction strategy would need further improvement, due to leakiness of the system. This result depicts the problematic application of genomic intervention during the production process. The following examples will depict more convenient

ways of redirecting carbon flux towards the desired product by synthetic parts and devices.

7.2.2 Transcription Level

7.2.2.1 Promoter Strength

Flux imbalances in a synthetic pathway can be overcome by tuning the promoter strength for each of the pathways catalysing enzymes. Varying the promoter strength for each enzyme one at a time is a tedious and time-consuming effort. To enhance the speed in such an approach, Du et al. (2012) developed the customised optimisation of metabolic pathways by combinatorial transcriptional engineering (COMPACTER) strategy. As a case study, COMPACTER was employed to engineer *S. cerevisiae* for xylose utilisation. As xylose is the most abundant C-5 sugar from lignocellulosic feedstocks, it is important to make it accessible for microorganisms devoted to a biorefinery application. The fungal xylose utilising pathway based on reductase-hydrogenase reactions was favoured in comparison to the bacterial isomerase reaction-based pathway. The fungal xylose pathway applied consists of three consecutive reactions. The xylose reductase (XR in this study from *Candida shehatae*) reduces xylose to xylitol, further on xylitol is converted to xylulose by xylitol dehydrogenase (XDH from *Candida tropicalis*), and finally, xylulokinase (XKS from *Pichia pastoris*) phosphorylates xylitol to channel it into the pentose phosphate pathway (PPP). So there are three enzymes whose transcript levels have to be orchestrated to balance the flux to the PPP and further on into glycolysis. For each enzyme, another constitutive *S. cerevisiae*-derived promoter was applied. From each promoter (PDC1 for XR, TEF1 for XDH and ENO2 for XKS expression), a mutagenised library was constructed by nucleotide analogue mutagenesis (Alper et al. 2005). These promoter mutants were assembled with their respective enzyme coding sequences and terminator sequences onto a single-copy plasmid by the DNA assembler method (Shao et al. 2009). Each plasmid created harboured the three expression cassettes for XR, XDH and XKS representing a mutant pathway for xylose utilisation. The reason for assembling the mutant pathway on a single-copy plasmid, instead of inserting it to the genome of *S. cerevisiae*, was the possibility to conveniently transfer an identified favourable mutant pathway to another strain background. The identification of a favourable mutant pathway was conducted by plating transformants of DNA assembler plasmids on xylose and selecting favourable phenotypes by colony size. The phenotypes created were characterised in liquid culture testing their performance on xylose utilisation and ethanol production under aerobic conditions. Expression levels of XR, XDH and XKS were determined by quantitative PCR. Du and coworkers tested the COMPACTER in two different strain backgrounds: a laboratory strain and the industrial “Classic Turbo Yeast”. For the laboratory strain, the selected phenotype had an improved ethanol yield of 60 % and a 70 % faster xylose consumption rate coupled to a 1.5 times faster ethanol production rate compared to the same strain background with the

non-mutated promoter equipped xylose utilisation pathway. The findings for the industrial strain were even more promising exhibiting a xylose consumption rate of 0.92 g/L/h and an ethanol yield of 0.26 g/g xylose. The benefit of COMPACTER is underlined by the finding that the industrial strain with the non mutated promoter pathway performed very poor by a non detectable ethanol production and total consumption of only 9 % of xylose provided. However, the most important outcome of the study is the comparison of expression levels in the laboratory strain and the industrial strain, which has shown that in the industrial strain, a higher XDH level than XR level is necessary for xylose consumption. In contrary, in the industrial strain, a higher XR than XDH level is favourable. Furthermore, by exchanging the best performing mutant pathways to the other strain, a decrease in performance was detected highlighting the host background dependency of synthetic pathways, which can be attributed to the overall availability of cofactors and variable stress responses of diverse host strains (Du et al. 2012). The synthetic character of this example lies first of all in the brick wise assembly of expression cassettes and further on a whole synthetic pathway for expanding the substrate range of a desired production host as well as the transferability of a selected combination to another chassis strain. The universal applicability of the COMPACTER has been proven by its application to a cellobiose utilising pathway additionally.

An even more advanced approach of improving xylose utilisation has been published recently by Latimer et al. (2014). To point out the differences to the aforementioned example:

1. Instead of using mutated promoter libraries, the authors used five characterised promoters (Lee et al. 2013) spanning evenly distributed the strength of three orders of magnitude including the strongest constitutive promoter known in *S. cerevisiae*. Plasmids harbouring the pathways with varying promoters for every single gene were constructed using the Golden Gate assembly technique (Engler et al. 2009).
2. All the necessary pathway genes were isolated from *Scheffersomyces stipitis*.
3. Additionally to the genes for XR, XDH and XKS, four genes of the nonoxidative PPP and the gene for pyruvate kinase were tested for overexpression. Expression cassettes were cloned onto two different plasmids, one containing the xylose utilisation pathway (resulting in 5^3 125 possible combinations) and one containing the PPP (5^5 3125 possible combinations) to allow the screening of 5^8 possible combinations. For identification of preferable combinations, the previously established TRAC method (Lee et al. 2013) was applied.
4. For enrichment of favourable promoter gene combinations, both aerobic and anaerobic conditions were used. Especially the anaerobic enrichment is of industrial interest owing to fermentations which are rather economically feasible when performed in large scale under anaerobiosis.

The results from this study show a clear difference in promoter strength for the three-gene pathway (xylose utilisation only) to the eight-gene pathway (xylose

utilisation plus PPP) selected under aerobic conditions. While the eight-gene pathway was enriched towards strong promoters driving XR expression, the three-gene pathway was enriched for medium strength promoter activity. Additionally, the eight-gene pathway strains showed a much faster growth on xylose. Authors hypothesised that the strain background used might be limited in the PPP especially in the expression of transaldolase (TAL), as the enrichment showed a preferable combination of the strongest promoter with TAL. Furthermore, the authors also suggest that the limiting PPP activity might explain the differences between the laboratory strain and the industrial strain from Du et al. (2012). Consequently, using a strain limited in PPP activity will only show a local optimum of expression tuning when the three-gene pathway is applied. Finally, the authors tested aerobically enriched strains under anaerobic industrially relevant conditions resulting in inferior performance of these strains, underlining the importance of adjusting selection criteria to the industrially relevant framework. This example highlights that it is often not enough to equip a production chassis with an additional synthetic pathway to broaden the substrate spectrum. One might not forget to tune the chassis native metabolism in order to benefit from the new substrate.

A more advanced concept in this context is multivariate modular metabolic engineering (MMME). This approach is based on grouping a given biosynthetic pathway into modules. For each module, the expression of every enzyme is adjusted to ascertain an optimal carbon flux through the module. Then the expression levels of the entire modules, which complete the pathway, can be varied against each other in order to find the optimal level for each module to maximise the pathway's productivity. This approach helps to find the right compromise between the necessity of knowledge for a complete rational engineering approach and the necessity for extensive screening for a (simple) combinatorial approach. MMME aims for a rationally based approach with moderate knowledge of the system. Therefore, MMME is applicable independent of host and pathway but still limiting the design space in which analyses have to be done to abandon the need for a high-throughput method. Biggs et al. (2014) reviewed a collection of examples where MMME was applied for the enhanced biosynthetic production of taxadiene, resveratrol, β-carotene, fatty acids and more, highlighting the universal applicability independent of host or product.

7.2.2.2 Toggle Switch

Precursor metabolites like acetyl-CoA are tightly connected to the central carbon metabolism. In a variety of chemical and fuel production processes (e.g. mevalonate pathway to artemisinic acid, 1-butanol, cadaverine, isobutanol), acetyl-CoA serves as a starting point of biosynthesis. Ultimately, there is a keen competition for building blocks of the desired product with the cells' needs for growth and energy. A knockout of competing genes from the cellular metabolism in this case is not feasible, because most processes depend on a moderate to high cell density during the production to maintain a high titer and yield of the production process. Several microbial chemical production processes show a distinct partition of biomass and product formation, for instance, the microbial production of itaconic

acid with *Aspergillus terreus* and citric acid production with *Aspergillus niger* or *Yarrowia lipolytica*. This partition is mostly owed to the limitation of compounds in the production medium other than the carbon source. To create a similar situation of disconnecting the growth phase from the production phase, Soma et al. (2014) applied a synthetic toggle switch for the production of isopropanol. To turn off the drain of acetyl-CoA into the TCA cycle at a certain time point, the *gltA* gene coding for citrate synthase was deleted from the *E. coli* chassis genome and supplied on a so-called switching plasmid. The switching plasmid carries a $P_L lacO_1$ promoter which controls the expression of a TetR repressor which inhibits transcription of *gltA* from a $P_L tetO_1$ promoter. Furthermore, the isopropanol pathway itself is encoded on an operon which is also controlled by $P_L lacO_1$ (Fig. 7.2a). This design enables the simultaneous activation of the isopropanol pathway as well as the repression of citrate synthase expression upon the addition of IPTG to the culture broth. To have a tight control over the two $P_L lacO_1$ promoters, an additional repressor source was added by a plasmid expressing the *lacI* repressor. Before testing the performance of the isopropanol pathway, the authors characterised the effect of the toggle switch controlled *gltA* expression in a strain without the isopropanol pathway plasmid. When the switch was not induced by IPTG, the respective strain performed comparable to a control strain considering growth, glucose consumption and acetate production. When IPTG was added at different time points of cultivation, the functionality of the switch was observed via an increased formation of acetate and an inhibitory effect on growth, especially when the addition was done at the early time points of cultivation (0 h, 6 h and 9 h). The effect of the toggle switch was also monitored by measurement of intracellular metabolite levels which revealed a decrease in citrate level down to 3.6 % compared to a control strain, and the same drop was observed for α-ketoglutarate to 5.2 %. Acetyl-CoA levels were increased by a factor of 3.2, which is in line with the aforementioned acetate accumulation, since acetyl-CoA serves as a precursor of acetate. Furthermore, also the acetyl-CoA precursor pyruvate increased 15.9-fold compared to the control strain. As the proof of concept was done, the effect of acetyl-CoA provision was tested for the production of isobutanol. Results show that especially switching off *gltA* at early time points of cultivation (6 h and 9 h) has a significant effect on isopropanol production. IPTG addition at 6 h and 9 h resulted in titers of 48.3 mM and 50.9 mM isopropanol, respectively. The yields were 47.3 and 48.5 mol/mol%, 2.8- to 3.1-fold higher compared to the respective control strain, whereat the theoretical maximum yield for isopropanol production from glucose is 1 mol isopropanol per 1 mol glucose.

By this attempt, Soma et al. (2014) demonstrated the applicability of a synthetic toggle switch for rerouting metabolites, closely connected to the central carbon metabolism, towards a synthetic pathway. The isopropanol pathway was selected as an example, but the applicability to other synthetic pathways using acetyl-CoA as a precursor metabolite is obvious. Moreover, the insertion of a toggle switch like this is feasible in further branch points of the central carbon metabolism, where synthetic pathways draw their precursor metabolites from.

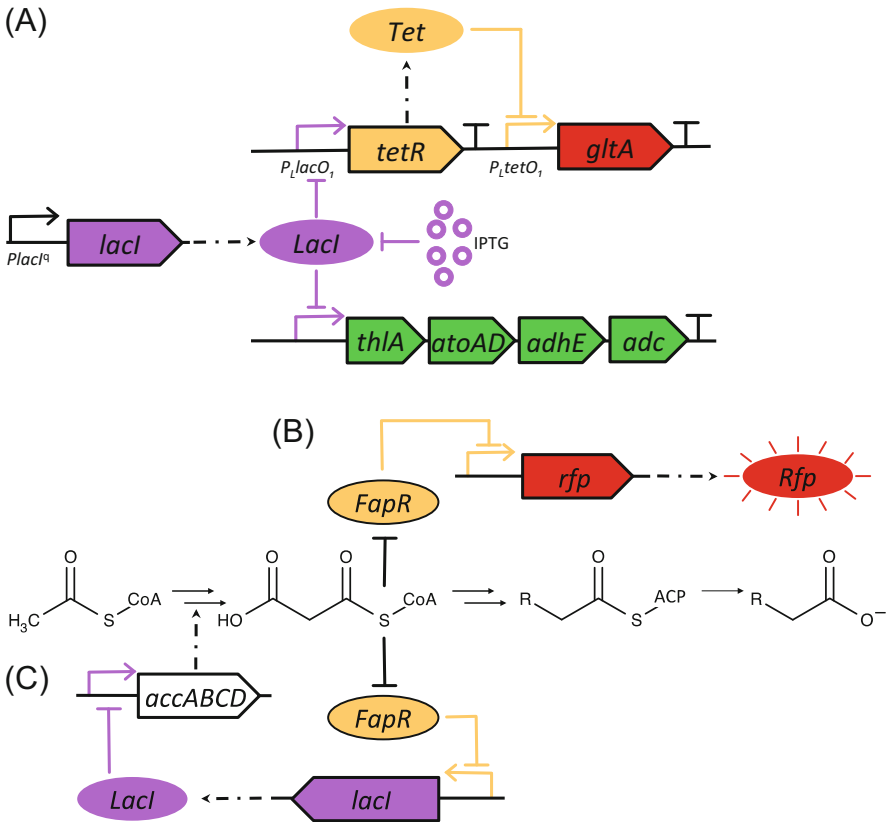


Fig. 7.2 (**a**) LacI represses the expression of TetR and the biosynthesis pathway of isopropanol. Upon induction with IPTG, TetR represses the expression of gltA (citrate synthase), while the expression of the isopropanol pathway is induced. (**b**) The malonyl-CoA-responsive transcription factor FapR represses the expression of Rfp. When malonyl-CoA is present in the system, it binds to the FapR transcription factor, thereby releasing FapR from the promoter and turning on Rfp expression. Fluorescence intensity of Rfp is directly correlated to the level of malonyl-CoA present in the cell. (**c**) Depending on the intracellular malonyl-CoA pool, the expression of acetyl-CoA carboxylase (accABCD) is regulated. High levels of malonyl-CoA induce the expression of LacI, which represses the expression of acetyl-CoA carboxylase, thereby reducing the formation of malonyl-CoA

7.2.2.3 Malonyl-CoA Sensor-Actuator System

When engineering chassis for accumulation of precursor metabolites, the engineer has to be cautious. The accumulation of certain precursor metabolites can also be detrimental to cell growth and further on to productivity due to toxicity effects of the precursor/intermediate. The biosynthesis of malonyl-CoA is such an example. Acetyl-CoA carboxylase catalyses the formation of malonyl-CoA from acetyl-CoA. Malonyl-CoA is the precursor of fatty acid synthesis, in which the activity of acetyl-CoA carboxylase represents the rate-limiting step. By overexpression of

acetyl-CoA carboxylase, the production of fatty acids can be enhanced, but the accumulation of the precursor metabolite malonyl-CoA hampers cell growth. To balance the expression of acetyl-Co carboxylase with the pace of the fatty acid biosynthesis to circumvent malonyl-CoA accumulation, Liu et al. (2015) developed a malonyl-CoA sensor-actuator system. They constructed a negative feedback circuit which is able to up- or downregulate expression of acetyl-CoA carboxylase, depending on the intracellular malonyl-CoA pool. The design of the sensor-actuator circuit is based on the malonyl-CoA responsive transcription factor FapR from *Bacillus subtilis*. In its native host, FapR represses fatty acid and phospholipid biosynthesis by binding to a 17 bp DNA sequence. At the onset, Liu and coworkers inserted the FapR-binding site in the −10 region of a synthetic promoter controlling the expression of red fluorescent protein (RFP). When malonyl-CoA is present in the system, it binds to the FapR transcription factor, thereby releasing FapR from the promoter and turning on RFP expression. All the expression cassettes necessary for the verification of the operability of the sensor-actuator system as well as fatty acid production in *E. coli* were constructed on separate plasmids with compatible origins of replication. The expression of FapR was controlled by a P_{BAD} promoter, whereas the expression of acetyl-CoA carboxylase was controlled by an IPTG-inducible T7 promoter (Fig. 7.2b). Upon induction with different concentrations of IPTG, acetyl-CoA carboxylase is expressed and malonyl-CoA is formed. The surplus of malonyl-CoA present in the cell releases the FapR repressor from the synthetic promoter giving rise to an RFP signal which can be quantified. By measuring the intracellular malonyl-CoA levels and correlating them to RFP signal intensity, the researchers have established a quick method for quantifying intracellular malonyl-CoA levels. By mutating the FapR-binding site, it was even possible to create promoters with varying strength and dynamic ranges, which is essential for fitting the triggered response to the actual malonyl-CoA level in the system. The final aim was the construction of a negative feedback regulatory circuit. For this reason, the expression of the acetyl-CoA carboxylase was controlled by a LacI-repressed T7 promoter. The expression of LacI was controlled by the FapR-regulated promoter, thereby closing the loop. A further coexpression of a cytosolic thioesterase *tesA* enables the release of the free fatty acids from the acyl carrier protein of the fatty acid biosynthesis pathway (Fig. 7.2c). The resulting strain produced 2.03 g/L of free fatty acids which was 34 % higher in titer than the respective control strain. In line with this result also, the fatty acid productivity was increased by 33 %. Also in respect to growth rate, the sensor-actuator equipped strain entered the exponential growth phase earlier than the control strain, reflecting the relief of the growth inhibitory effect of malonyl-CoA accumulation. It will be interesting to see in the future whether this malonyl-CoA sensor-actuator device will be applied to other production systems that need malonyl-CoA precursors like flavonoids, polyketides and 3-hydroxypropionic acid production strains.

7.2.3 Translation Level

7.2.3.1 Synthetic Ribosome Binding Sites

Besides tuning the promoter strength of a given synthetic pathway (acting on transcription), the expression level of catalysing enzymes can be tuned by the insertion of synthetic ribosome binding sites (RBS, acting on translation). Lin et al. (2014) demonstrated this for the production of riboflavin in *E. coli*. The riboflavin biosynthetic pathway is a branched pathway starting from GTP and ribulose-5-phosphate as precursors. Several strategies have been applied to enhance riboflavin production. Among them, synthetic operons of the riboflavin pathway from *E. coli* and *Bacillus subtilis* have been tested on plasmids with different copy numbers. However, most importantly, riboflavin itself is a precursor for FMN (flavin mononucleotide) and FAD (flavin adenine dinucleotide), and FMN acts as a potent feedback regulator on the riboflavin biosynthesis pathway. To overcome this feedback inhibition, Lin et al. (2014) attempted to reduce the expression level of *ribF*, the bifunctional riboflavin kinase/FMN adenylyltansferase, which catalyses the formation of FMN and FAD from riboflavin. To this end, they constructed a library of RBS sequences weaker than the native RBS using the RBS calculator from Salis (2011). Introducing this RBS library to the production strain, nearly 2/3 of the presented strains exhibited a higher riboflavin production compared to the parental strain. The best clone accumulated a 77 % higher titer of the vitamin. Further improvements of the cell factory could involve an upregulation of the precursor pathways to GTP and ribulose-5-phosphate. In this case, stronger RBS sequences could be used to tune expression. This is also one of the tools which Ng et al. (2015) used to improve NADPH provision in *E. coli*, underlining the potential of the modification of translational efficiency by RBS engineering.

7.2.3.2 Dynamic Metabolic Valve

To redirect the flux of glucose away from glycolysis directly to a heterologous pathway for gluconate production by the activity of glucose dehydrogenase (Gdh), Solomon et al. (2012) engineered a dynamic metabolic valve tuning the expression of glucose kinase (Glk). To highlight the synthetic character of this approach, it is noteworthy that all the building blocks used in this study were obtained from the Registry of Standard Biological Parts (http://partsregistry.org). The authors tested the inducible transcription of antisense RNA (asRNA) to reduce Glk activity, thereby reducing the specific growth rate of the production host *E. coli*. asRNA constructs were designed by inverting the *glk* including 20 bp of the upstream region of glk encompassing the RBS. Transcription of asRNA was controlled by a Tet promoter (P_{tet}) (Fig. 7.3a). Four constructs of asRNA of different length were tested, whereby constructs of 100 bp and 1000 bp showed markedly reduction in specific growth rates upon induction with anhydrotetracycline (aTc). The 100 bp asRNA construct indicated a decreased Glk activity of 12 %, whereas the 1000 bp asRNA construct indicated a decrease of 25 %. The mode of action of these constructs seemed to be different. The 100 bp construct did not have any influence on mRNA levels of Glk, indicating a RBS occlusion mode of action. For the

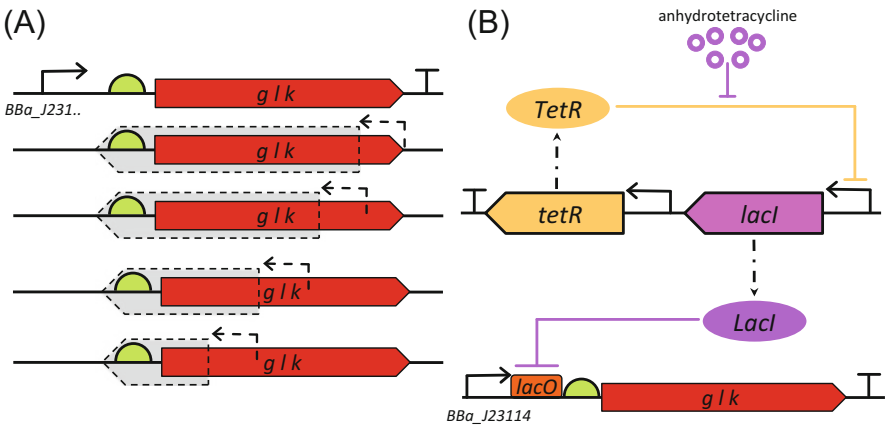


Fig. 7.3 (**a**) Illustrates the expression of antisense RNA of different lengths from an inducible promoter (*dashed arrows*), by inverting the coding sequence of glucose kinase Glk including 20 bp of the upstream region of *glk* encompassing the ribosome binding site. (**b**) The constitutively expressed tetR represses the expression of LacI which binds to the lac operator of the Glk expression cassette turning Glk expression off. Upon induction with anhydrotetracycline, tetR repression is released from LacI, thereby turning off the expression of Glk

1000 bp construct in contrary, a reduction of mRNA levels more than 85 % was observed, suggesting that binding of the asRNA increased mRNA degradation. Testing these constructs on glucose medium, the authors witnessed what they termed a "growth-mediated buffering" effect. The effect of inhibition of Glk reduces growth rate, thereby concentrating the remaining Glk in the cells. Glk is a fairly stable protein, and it could only be diluted in the culture by cell growth, which was detained upon asRNA transcription. In another attempt, Solomon et al. (2012) constructed an inverting gene circuit as a dynamic glucose valve activator consisting of a constitutively expressed tetR which represses expression of lacI. Upon aTc addition, TetR is repressed, thereby driving the expression of LacI which binds to the lac operator of the Glk expression cassette turning Glk expression off (Fig. 7.3b). Applying this circuit with an appropriate amount of 100 ng/mL aTc, an 80 % repression of mRNA levels was observed. Furthermore, the inverter was able to reduce growth by 50 %. Testing the inverter for the production pathway of gluconate, a 30 % increase in gluconate titers and additionally a 40 % reduction in acetate formation were observed, which accounts for carbon loss via by-product formation.

7.2.3.3 Riboswitches

A riboswitch is the regulatory part of an mRNA, located in the upstream region of a coding sequence. It consists of two elements, an aptamer and an expression platform. When a specific metabolite (e.g. lysine) binds to the aptamer region, a structural change of the expression platform is induced, which can block the access to, e.g. the ribosome binding site (Fig. 7.4a). Zhou and Zeng (2015) tested two

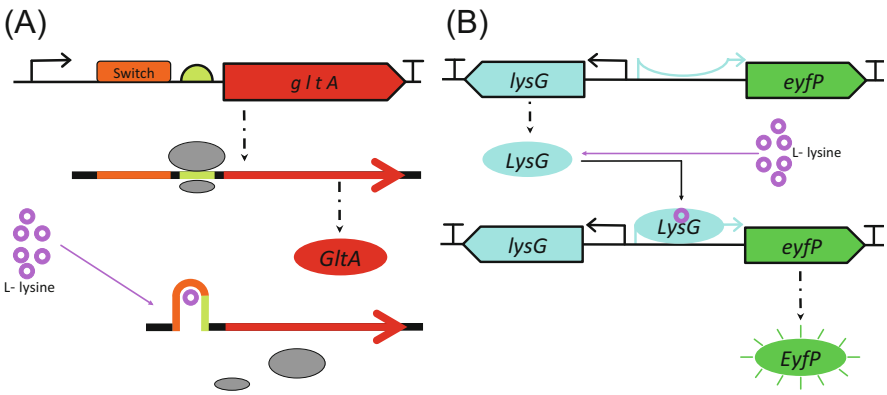


Fig. 7.4 (**a**) Intracellular accumulation of L-lysine induces a conformational change of the mRNA by binding to the aptamer region of the riboswitch controlling the translation of citrate synthase *gltA*. (**b**) The transcription factor LysG is activated by L-lysine and binds to the promoter region of lysE which controls the expression of eYFP. The fluorescence signal of eYFP enables the cell sorting for single cells with elevated levels of L-lysine

lysine riboswitches, one from *E. coli* and the other from *B. subtilis*, first in *E. coli* and further on in the lysine producer *Corynebacterium glutamicum*. Comparable to Soma et al. (2014) mentioned above, Zhou and Zeng (2015) wanted to gain control over expression of citrate synthase *gltA* but without the necessity of external addition of an inducer molecule. By replacing the regulatory sequence of the native *gltA* 5′UTR with the lysine riboswitches, they confirmed that the lysine pool present in *E. coli* was sufficient to downregulate *gltA* translation resulting in a striking reduction of cell growth and glucose consumption in nearly one order of magnitude. To further characterise the strength of the two riboswitches, a β-galactosidase reporter plasmid was constructed, finding that the *E. coli* riboswitch confers a much stronger lysine repression than the one from *B. subtilis*, tested in an *E. coli* background at least. After transferring both these riboswitch-regulated *gltA* expression cassettes to the chromosome of a *C. glutamicum* lysine overproducer strain, results confirmed the stronger repression of the *E. coli*-derived riboswitch. While the growth of lysine riboswitch carrying mutants was retarded, the glucose consumption rate was nearly as high as for the control strain. The redirection of the carbon flux towards product formation was confirmed by a 63 % and 38 % higher product yield for the *E. coli* and the *B. subtilis* riboswitch, respectively, compared to the control strain. Measurements of enzyme activities of citrate synthase indicated that compared to the control strain, the riboswitch mutants had only 30 % and 43 % of activity for the *E. coli* and *B. subtilis*, respectively.

Future perspectives for the development and application of riboswitches and a status quo have been recently reviewed by Berens et al. (2015). One of the striking massages is that an aptamer domain can be selected in vitro for virtually any given small molecule. This comes in very handy for sensing a metabolite of choice in a producing cell to regulate the activity of growth or by-product pathways.

Furthermore, metabolite levels could be made visible in the cell comparable to the RFP signal quantification described by Liu et al. (2015) above, whether for a quick quantification for intracellular metabolite levels or the online monitoring of the overall cell status during a production process. A further example of metabolite sensing via an optical signal is given in the following section.

7.2.4 Protein Level

Enzymes catalysing a biosynthetic overproduction pathway are often feedback inhibited by the respective product. To overcome such a feedback inhibitory effect, it is worthwhile to engineer the catalysing protein itself by eliminating or reducing the affinity of the enzyme to the inhibitory product. For this purpose, Schendzielorz et al. (2014) applied the metabolite sensor pSenLys-Spc. This sensor is based on the metabolite binding transcription factor LysG which binds to the promoter region of its target gene *lysE* upon activation by elevated levels of basic amino acids (L-lysine, L-arginine and L-histidine) in *C. glutamicum* (Fig. 7.4b). Fusing an *eyfp* (enhanced yellow fluorescent protein) coding sequences to the *lysE* promoter enabled a high-throughput sorting for cells with elevated intracellular product formation by FACS (fluorescence-activated cell sorting). To enable the accumulation of elevated intracellular product levels, a *C. glutamicum* strain with a chromosomal deletion of a basic amino acid exporter, carrying the pSenLys-Spc reporter, was employed. This strain was transformed with mutant libraries of the product-inhibited pacemaker enzymes for L-arginine (*argB*), L-lysine (lysC) and L-histidine (HisG). Sorting cells via FACS enabled the identification of mutants of each enzyme which showed reduced feedback inhibition, resulting in elevated product formation of 34 mM of L-arginine, 45 mM of L-lysine and 17 mM of L-histidine, compared to titers of 0.4 mM, 1.2 mM and 0.0 mM of the respective control strains.

7.2.5 Spatial Organisation

The picture of a bacterial cell as a bag of enzymes is being overhauled. Today, a wide variety of organisational features in bacteria and archaea are known, which have recently been reviewed by Saier (2013). In this section, only organisational forms related to metabolism like substrate channelling via enzyme aggregation, bacterial microcompartments and natural scaffolds will be discussed. The benefits and properties of these natural phenomena will be discussed in the context of current efforts aiming at their exploitation as synthetic biology tools to artificially create organisation in heterologously introduced metabolic pathways.

7.2.5.1 Substrate Channelling in Central Carbon Metabolism of Bacteria

The idea of the central carbon metabolism in bacteria as a result of random molecular collisions involving substrates, cofactors and enzymes is becoming increasingly unrealistic (Saier 2013). Recent studies found several protein-protein

interactions in glycolysis and TCA cycle of *Bacillus subtilis* (Commichau et al. 2009; Meyer et al. 2011). In glycolysis, an interaction between phosphofructokinase, phosphoglyceromutase and enolase has been identified, for the TCA cycle citrate synthase, isocitrate dehydrogenase and malate dehydrogenase on the one hand, and fumarase, aconitase and malate dehydrogenase on the other hand have been detected as interaction partners for catalysing sequential reactions (Commichau et al. 2009; Meyer et al. 2011). These interactions are thought to establish substrate channelling of intermediates of the central carbon metabolism to enhance metabolic flux through these pathways.

One emerging strategy to overcome problems in metabolic engineering is the colocalisation of enzymes similar to the multienzyme complexes found in nature. By bringing enzymes in close proximity, substrate channelling to prevent carbon loss due to diffusion or secretion of intermediates and side reactions in host cell networks is achieved (Lee et al. 2012a). Strategies aiming at colocalisation of enzymes usually involve the construction of a synthetic scaffold to which enzymes are attached. This creates a space with high enzyme "density" where intermediates are passed from one enzyme to the next, with reduced probability of leaking to the cytosol. Attaching enzymes to scaffolds furthermore allows for control of stoichiometry between different enzymes to balance flux in a pathway.

7.2.5.2 Protein Scaffolds

A first example of improving a heterologously expressed pathway with the aid of a scaffold was shown by Dueber et al. (2009). There, protein-protein interaction domains were used as a scaffold, and the enzymes of the mevalonate pathway, acetoacetyl-CoA thiolase from *E. coli*, together with hydroxymethylglutaryl-CoA synthase and hydroxymethylglutaryl-CoA reductase from *S. cerevisiae*, were attached to this scaffold by tagging the enzymes with peptide ligands specific for the interaction domains. The modularity of the interaction domains allowed for changing the stoichiometry of the enzymes within the scaffold. Controlling the stoichiometry of the overexpressed enzymes was used to optimise the flux through the mevalonate pathway, as the expression of the enzymes without the scaffold showed an imbalanced flux resulting in intermediate accumulation. In order to lower metabolic burden, pathway enzyme and scaffold expression were separated from each other and individually titrated for optimal production. Low induction of the pathway enzymes and high induction of the protein scaffold turned out to be optimal for mevalonate production. The pathway with optimised stoichiometry showed a 77-fold increase in mevalonate production compared to expression of the pathway without a scaffold. Other studies showed improved production of glucaric acid (fivefold) and butyrate (threefold) by the use of the same protein scaffold (Moon et al. 2010; Baek et al. 2013).

Another example for the application of protein scaffolds is the hydrolysis of cellulose. This reaction is carried out by anaerobic bacteria in a complex, multienzyme structure called cellulosome. The cellulosome usually consists of several cellulose-binding modules and a series of different cohesin-dockerin attached cellulases. A highly ordered structure and close proximity are key features of

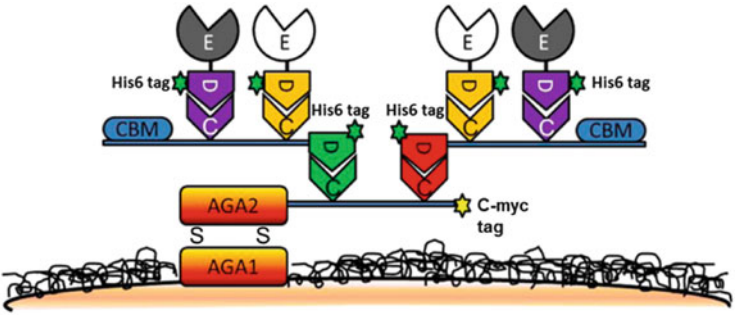


Fig. 7.5 Schematic diagram of the cellulosome complex assembled on the yeast surface using the adaptive assembly. The two adaptor scaffoldins (*red* and *blue*) serve as templates for enzyme recruitment to the yeast surface via specific interaction with the surface-displayed anchoring scaffoldin. *Green* (scaffoldin/dockerin from *Acetivibrio cellulolyticus*), *red* (scaffoldin/dockerin from *Bacteroides cellulosolvens*), *purple* (scaffoldin/dockerin from *Clostridium thermocellum*), *yellow* (scaffoldin/dockerin from *Ruminococcus flavefaciens*). *CBM* cellulose-binding module, *E* enzyme. Reprinted with permission from Tsai et al. 2013. ACS Synth Biol. 2013. Copyright 2013 American Chemical Society

natural cellulosomes. Tsai et al. (2013) recently showed the construction of a synthetic cellulosome displayed on yeast surface to increase cellulose hydrolysis 4.2-fold compared to the use of free enzymes. In their study, they combined cohesin-dockerin modules from different bacteria and created an artificial, tetravalent cellulosome. Briefly, an anchor scaffoldin was attached to the yeast cell surface, containing two cohesin domains. Connected to each of these cohesion domains was a dockerin tagged with two enzymatic subunits, endoglucanase and β-glucosidase, and a cellulose-binding module (Fig. 7.5). It was further shown that ethanol production increased twofold with the tetravalent cellulosome compared to a divalent cellulosome with similar enzyme loading. The authors hypothesised that enzyme proximity has a higher impact on cellulose hydrolysis than overall number of cellulolytic enzymes.

7.2.5.3 Nucleic Acids as Scaffolds

Plasmid-Based Scaffolds

Another approach proven to be feasible is the use of DNA as scaffold. Conrado et al. (2012) used plasmid DNA with synthetic binding sites as the scaffold and fused zinc fingers specific for these binding sites to the enzymes of a metabolic pathway. Production improvement using *E. coli* as host was largely dependent on the scaffold architecture, namely, the ratio of binding sites for the individual enzymes and the length of the spacer on the plasmid between the different binding sites. By varying the length of the spacer, the arrangement of the enzymes on the plasmid DNA is changed, with the optimal arrangement leading to close proximity of the different enzymes. Furthermore, the number of scaffold units on the plasmid was also found to be important, as in general, the plasmid copy number even with a

high-copy-number plasmid is significantly lower compared to the number of enzyme molecules produced in an overexpression strain. Therefore, multiple scaffold units were included on the plasmid to accommodate all enzyme molecules on the scaffold. Using optimal scaffold architecture for different pathways, 5-fold, 4.5-fold and 2.5-fold titer improvement for resveratrol, 1,2-propanediol and mevalonate, respectively, was observed compared to expression of the enzymes without scaffolds. Another study employed DNA scaffolding for improved production of N-acetylglucosamine and obtained a 2.5-fold titer improvement when glucosamine synthase and N-acetylglucosamine/N-acetyltransferase were colocalised on a DNA scaffold (Liu et al. 2014). The review by Lee et al. (2012a) named different interaction motifs as well as variable length linkers between binding sites without solubility issues as potential advantages of DNA scaffolds, whereas heavy enzyme modification by lengthy zinc finger domains and limitation of the number of scaffolds per cell by the plasmid copy number (~500) were mentioned as possible drawbacks.

RNA Molecule-Based Scaffolds

Single RNA molecules with two aptamers as well as RNA molecules that can be polymerised in one or two dimensions have been employed in a recent study (Sachdeva et al. 2014). For that purpose, a library of eight aptamers and corresponding RNA-binding domains was constructed. By tagging split green fluorescent protein fragments with RNA-binding domains, fixation to an RNA scaffold colocalised the fragments resulting in measurable fluorescence. The study further showed application of RNA scaffolds to a two-enzyme pentadecane pathway and a three- and four-enzyme pathway for succinate production, which for both metabolites resulted in a 1.8-fold increase in production.

Bacterial Outer Membrane Vesicles as Scaffolds

Another interesting approach for a scaffold has been shown by Park et al. (2014). Bacterial outer membrane vesicles (OMVs) were used as the scaffold for extracellular display of cellulolytic enzymes. In detail, an ice nucleation anchor was connected to the surface of the OMV surface. Attached to the anchor was a trivalent scaffold containing cohesin domains and a cellulose-binding domain. By interaction with the cohesin domains, enzymes tagged with the respective dockerins could be assembled in an ordered fashion on the surface of OMVs. Glucose production was shown to be improved 23-fold compared to expression of free enzymes.

7.2.5.4 Bacterial Microcompartments

Bacterial microcompartments (BMCs), macromolecular structures entirely made up from proteins, are simple organelles with a large number of different structures and functions (Chowdhury et al. 2014; Axen et al. 2014). A first example of a BMC was discovered in the 1950s, when carboxysomes were observed in the electron microscope (Chowdhury et al. 2014; Drews and Niklowitz 1956). Despite the knowledge about bacterial microcompartments in some bacteria, it has only recently become fully appreciated how widespread bacterial microcompartments

are as an organisational form in bacteria (Lee et al. 2012a). Assisted by bioinformatics and an increasing number of bacterial genome sequences available, today, BMCs have been discovered in hundreds of bacteria with a variety of metabolic functions assigned to them (Axen et al. 2014). The genetic information encoding these macromolecular complexes is either organised in operons (e.g. pdu, eut and α-carboxysomes) or spread throughout the bacterial genome (e.g. β-carboxysomes) (Frank et al. 2013). Operons have most likely been spread among different bacteria via horizontal gene transfer. BMCs are polyhedral in shape and have a diameter of approximately 100–150 nm. The protein layer forming the shell of BMCs is thought to be 3–4 nm thick and consists of up to 20,000 polypeptides of 10–20 different types, and so far, no evidence for the presence of lipids has been found (Cheng et al. 2008). Encapsulated by the shell, 10,000–15,000 protein molecules are found in the lumen of the BMC required for the metabolic function (Cheng et al. 2008; Yeates et al. 2008). The interior of the microcompartments is connected to the rest of the cell by small selective pores in the protein shell (Frank et al. 2013). These pores are thought to allow substrates and products to enter and leave BMCs, while toxic and volatile intermediates are retained within the lumen. This provides protection to the rest of the cell. Additionally, intermediate loss due to diffusion is thought to be reduced, and a higher metabolic efficiency is achieved by creation of an improved microenvironment in the lumen of the BMC.

Therefore, the main function of BMCs is thought to be the optimisation of metabolic pathways (Chowdhury et al. 2014). The models of the two best studied BMCs support this hypothesis. The anabolic carboxysome is involved in the carbon fixation in cyanobacteria and functions as part of a carbon dioxide-concentrating mechanism (CCM) (Cheng et al. 2008). CO_2 is concentrated within the carboxysome, which improves the catalytic efficiency of RuBisCO, the rate-limiting enzyme of the Calvin cycle. In addition, the shell of the carboxysome prevents entry of molecular oxygen and, thus, lowers photorespiration activity, a process otherwise draining up to 50 % of carbon fixed in the Calvin cycle (Cheng et al. 2008; Cannon et al. 2001). The second example, the catabolic pdu metabolosome, plays a key role in bacterial degradation of 1,2-propanediol and glycerol. Both substrates are converted into an alcohol and a carboxylic acid. The intermediate of this conversion is in both cases a toxic aldehyde. The pores of the pdu metabolosome are thought to be selective in retaining the intermediate within the lumen of the BMC while allowing substrates and products to enter and exit (Chowdhury et al. 2014). Therefore, the main function of the pdu metabolosome is thought to protect the cell from cytotoxicity and DNA damage by sequestering the toxic aldehyde intermediate (Sampson and Bobik 2008; Chowdhury et al. 2014).

Because of their properties, BMCs have been proposed as a potential synthetic biology tool for metabolic engineering strategies. Targeting of a complete, non-natural pathway composed of a number of different enzymes to a BMC might provide the same advantages as for the pathway naturally found in the BMC.

A prerequisite for exploitation of these structures as nanobioreactors would require targeting of foreign enzymes to the interior of a BMC. Encapsulation of

metabolic enzymes naturally located within a BMC is thought to be mediated by interaction of the enzymes with shell proteins during BMC assembly (Fan et al. 2010; Chowdhury et al. 2014). Fan and coworkers first observed short N-terminal extensions on pduP, a propionaldehyde dehydrogenase, when they performed sequence alignment of different pduP homologues. It could be shown that the deletion of the first 14 or 18 amino acids from pduP of *Salmonella enterica* had little effect on activity, but the enzyme largely lost its association with the BMC. When GFP was N-terminally tagged with the first 18 amino acids of pduP, fluorescence localised to the BMC was observed. This clearly proved the role of this peptide in encapsulation of enzymes in the BMC of *S. enterica*. Other luminal enzymes such as pduD and pduV were also shown to have similar N-terminal targeting sequences. However, for other encapsulated enzymes, no such N- or C-terminal extensions could be identified, and the mechanism of encapsulation of these enzymes is not yet understood. It is thought that the mechanism underlying encapsulation is an interaction between α-helices at the N-terminus of the enzyme and the C-terminus of one of the shell proteins and that manipulation of these interactions could allow for the encapsulation of multiple enzymes with a defined stoichiometry (Fan et al. 2012; Chowdhury et al. 2014). This would mark an important feature of a bespoke nanobioreactor. Another study showed a potential tool for rapid flow cytometric quantitation of protein encapsulation (Kim and Tullman-Ercek 2014). This fluorescence reporter-based system could be of great use for engineering targeting sequences as it allows fast evaluation of the targeting efficiency.

A first milestone in making bespoke BMC nanobioreactors was the study of Parsons et al. (2010), which showed recombinant production of empty BMCs in *E. coli* with a minimal set of six shell proteins, namely, pduA, pduB, pduJ, pduK, pduN and pduU, from *Citrobacter freundii*. Targeting of GFP to recombinant BMCs containing the fusion construct mCherry-pduA resulted in a colocalised red and green fluorescence signal, strongly indicating targeting of GFP to intact mCherry-labelled protein shells. Other studies later showed recombinant production of functional carboxysomes with a minimal set of ten genes, and empty recombinant microcompartments form ethanolamine (eut) metabolosome shell proteins in *E. coli* (Bonacci et al. 2012; Choudhary et al. 2012). The study of Lawrence et al. (2014) marked the first example of a simple BMC nanobioreactor in *E. coli*. The pyruvate decarboxylase and alcohol dehydrogenase of *Z. mobilis* were targeted with P18 and D60 targeting sequences, respectively, and expressed along with six shell proteins pduA, pduB, pduJ, pduK, pduN and pduU from *C. freundii*. This resulted in a 30 % higher production of ethanol compared to expression without the shell proteins.

For efficient exploitation of these fascinating structures as reaction chambers, an understanding of the pore selectivity is required. Knowledge about how pores function would allow adaption of pore properties for molecules associated with a foreign pathway. However, pore selectivity is still poorly understood. The models of shell proteins available show distinct differences regarding pore residues. Since one of the main hypotheses about BMCs is that these structures confer protection of

the cell by sequestering toxic intermediates, all pores from the different shell proteins must show similar selectivity in allowing substrates and products to enter and leave the BMC and sequestering toxic intermediates. Crystallographic studies of pduB of *L. reuteri* in the presence of glycerol showed blocking of the pore by three molecules of glycerol, which would allow only glycerol to enter the BMC and preventing the toxic aldehyde intermediates from relocating to the cytosol (Pang et al. 2012). How the products propionaldehyde-CoA and propanol are exported from the BMC is not yet understood.

The ability to control pore selectivity and therefore transport to and from the BMC would be a key requirement to make bespoke BMCs as nanobioreactors in metabolic engineering strategies. To date, available studies dealing with changing pore selectivity are very limited. Cai et al. (2014) showed the possibility of random mutagenesis of pore residues as well as combining shell proteins from different types of BMCs to create chimeric shells. The review of Lee et al. (2012a) has suggested random pore mutations coupled with an effective screening to alter pore selectivity. Transcriptional regulators as biosensors for desired metabolites coupled to the expression of an antibiotic resistance gene or a cytotoxic protein would allow for positive and negative selection (Lee et al. 2012a). Additionally, engineering of FRET sensors for pathway metabolites and sequestering of these sensors in BMCs have been proposed to directly measure metabolite permeability (Lee et al. 2012a).

7.2.6 Combined Approaches

Metabolic regulation of natural cells is generally extremely tight and robust due to the combination of regulatory events on all levels. As outlined before, excellent tools and devices have been constructed acting on different levels of the cellular networks. Finally, the metabolic engineer has to do what nature does—to combine different approaches on different levels to obtain optimal results. Clearly, here we are looking into the future as the development of synthetic tools, and a systems wide understanding of cellular networks is just emerging. One example for a combination of approaches has recently been shown by Ng et al. (2015) who combined the construction of synthetic bacterial operons, RBS engineering and MAGE to obtain a library of more than 600 *E. coli* strains, which they screened for improved NADPH provision. Combining this further with an optimised terpenoid pathway increased the terpenoid titer by 97 %. This points nicely into the direction how synthetic biology will aid metabolic engineering very efficiently in the future.

7.3 Commercial Application

A remarkable story of commercialisation of a synthetic pathway-derived chemical is Genomatica's process for the production of 1,4-butanediol (BDO). The synthetic pathway to the highly reduced BDO (relative to the potential carbohydrate

substrates available for bioproduction) has been designed from scratch. By using a genome-scale metabolic model of *E. coli* and their in-house biopathway prediction algorithm, 10,000 possible ways were predicted to form BDO from *E. coli*'s central carbon metabolism. The reason for this vast number of predicted ways might be that the applied SimPheny Biopathway Predictor software uses transformation of functional groups of known chemistry, instead of known enzyme reactions. By visualising the predicted pathways, Genomatica's scientists had to decide which way to go. The decision was made by ranking the possible ways by predicted thermodynamic feasibility, maximum yield, length of pathway, steps without currently known enzymes and the number of steps from the central metabolism to the product (Yim et al. 2011). A decision was made to test two pathways starting from the TCA cycle intermediates succinate and α-ketoglutarate from where pathways were predicted to lead to the intermediate metabolite 4-hydroxybutyrate (4HB) and further on to BDO. Similar to the modulation described above, the predicted pathway was divided into an upstream part leading to 4HB and a downstream part leading from 4HB to BDO. For both upstream and downstream pathways, several heterologous enzymes from different organisms were introduced in *E. coli*, testing several plasmids with varying copy number. Combining up- and downstream parts, it was found to be favourable to have the upstream part on a medium copy level and the downstream part on a high copy plasmid. As soon as the complete pathway was established, scientists at Genomatica started to engineer the production host by eliminating competing by-product pathways predicted by the OptKnock algorithm. It was shown that a strain engineered in this way changed the product pattern in a fermentation run away from by-product formation towards the desired product. Titers of 18 g/L BDO were reported in comparison to the non engineered strain yielding 7.5 g/L. Finally, the applicability of this BDO-producing strain in a biorefinery context was proven by testing the strain on sucrose, xylose and crude mixed sugars from biomass hydrolysis.

7.4 Conclusions

The future is bright for synthetic biology applications in metabolic pathway engineering. Several examples depicted in this book chapter will find or already have found their way towards a commercialised application. The standardisation and characterisation of biological parts and devices and the concept of transferability between different laboratory strains and industrially relevant production organisms add a big plus to the recent examples of successful metabolic engineering, which are rather isolated showcases. As outlined in the text, synthetic devices can assist metabolic engineering efforts on all the possible levels of action. It is noteworthy that also recent metabolic engineering efforts tend to systematically standardise their tools like using tunable promoters, manipulating ribosome binding sites, diversifying gene order in an operon and scaffolding to colocalise enzymes participating in a pathway. Moreover, to circumvent imbalances in a heterologous pathway, enzymes with a comparable turnover are grouped into modules, and

further on modules with different turnovers are orchestrated by adjusting their concentration by applying promoters of different strength or varying the copy number of a module (Yadav et al. 2012).

As already mentioned in the beginning, there are no "stand-alone" definitions for metabolic engineering and synthetic biology, rather than an overlap of disciplines that will cross-fertilise in the future. With the ongoing improvement in both disciplines, the field is moving forward, and future will show if synthetic biology can keep up with the prediction to heal us, feed us and fuel us (Editorial 2014).

References

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102(36):12678–12683, Epub 2005 Aug 25

Axen SD, Erbilgin O, Kerfeld CA (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. PLoS Comput Biol 10(10), e1003898. doi:10.1371/journal.pcbi.1003898

Baek JM, Mazumdar S, Lee SW, Jung MY, Lim JH, Seo SW, Jung GY, Oh MK (2013) Butyrate production in engineered Escherichia coli with synthetic scaffolds. Biotechnol Bioeng 110(10):2790–2794. doi:10.1002/bit.24925, Epub 2013 Apr 22

Berens C, Groher F, Suess B (2015) RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. Biotechnol J 10 (2):246–257. doi:10.1002/biot.201300498

Biggs BW, De Paepe B, Santos CN, De Mey M, Kumaran Ajikumar P (2014) Multivariate modular metabolic engineering for pathway and strain optimization. Curr Opin Biotechnol 29:156–162. doi:10.1016/j.copbio.2014.05.005, Epub 2014 Jun 11

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA, Savage DF (2012) Modularity of a carbon-fixing protein organelle. Proc Natl Acad Sci U S A 109(2):478–483. doi:10.1073/pnas.1108557109

Cai F, Sutter M, Bernstein SL, Kinney JN, Kerfeld CA (2014) Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. ACS Synth Biol. doi:10.1021/sb500226j

Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM (2001) Microcompartments in prokaryotes: carboxysomes and related polyhedra. Appl Environ Microbiol 67(12):5351–5361

Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA (2008) Bacterial microcompartments: their properties and paradoxes. Bioessays 30(11–12):1084–1095. doi:10.1002/bies.20830

Choudhary S, Quin MB, Sanders MA, Johnson ET, Schmidt-Dannert C (2012) Engineered protein nano-compartments for targeted enzyme localization. PLoS One 7(3), e33342. doi:10.1371/journal.pone.0033342, Epub 2012 Mar 12

Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA (2014) Diverse bacterial microcompartment organelles. Microbiol Mol Biol Rev 78(3):438–468. doi:10.1128/MMBR.00009-14

Commichau FM, Rothe FM, Herzberg C, Wagner E, Hellwig D, Lehnik-Habrink M, Hammer E, Völker U, Stülke J (2009) Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. Mol Cell Proteomics 8 (6):1350–1360. doi:10.1074/mcp.M800546-MCP200, Epub 2009 Feb 3

Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, Turnšek J, Tomšič N, Avbelj M, Gaber R, Koprivnjak T, Mori J, Glavnik V, Vovk I, Benčina M, Hodnik V, Anderluh G, Dueber JE, Jerala R, DeLisa MP (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic Acids Res 40(4):1879–1889. doi:10.1093/nar/gkr888, Epub 2011 Oct 22

Drews G, Niklowitz W (1956) Beiträge zur Cytologie der Blaualgen. II. Zentroplasma und granulare Einschlüsse von Phormidium uncinatum. Arch. Mikrobiol 24:147–162

Du J, Yuan Y, Si T, Lian J, Zhao H (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res 40(18), e142. doi:10.1093/nar/gks549, Epub 2012 Jun 19

Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27(8):753–759. doi:10.1038/nbt.1557, Epub 2009 Aug 2

Editorial (2014) Milestones in synthetic (micro)biology. Nat Rev Microbiol 12:309

Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4(5), e5553. doi:10.1371/journal.pone.0005553, Epub 2009 May 14

Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA (2010) Short N-terminal sequences package proteins into bacterial microcompartments. Proc Natl Acad Sci U S A 107(16):7509–7514. doi:10.1073/pnas.0913199107

Fan C, Cheng S, Sinha S, Bobik TA (2012) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. Proc Natl Acad Sci U S A 109(37):14995–15000. doi:10.1073/pnas.1207516109, Epub 2012 Aug 27

Förster J, Famili I, Fu P, Palsson BØ, Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res 13(2):244–253

Frank S, Lawrence AD, Prentice MB, Warren MJ (2013) Bacterial microcompartments moving into a synthetic biological world. J Biotechnol 163(2):273–279. doi:10.1016/j.jbiotec.2012.09.002

Kim EY, Tullman-Ercek D (2014) A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. Biotechnol J 9(3):348–354. doi:10.1002/biot.201300391

König H, Frank D, Heil R, Coenen C (2013) Synthetic genomics and synthetic biology applications between hopes and concerns. Curr Genomics 14(1):11–24. doi:10.2174/1389202911314010003

Latimer LN, Lee ME, Medina-Cleghorn D, Kohnz RA, Nomura DK, Dueber JE (2014) Employing a combinatorial expression approach to characterize xylose utilization in Saccharomyces cerevisiae. Metab Eng 25:20–29. doi:10.1016/j.ymben.2014.06.002, Epub 2014 Jun 13

Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue WF, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, Warren MJ (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS Synth Biol 3(7):454–465. doi:10.1021/sb4001118

Lee H, DeLoache WC, Dueber JE (2012a) Spatial organization of enzymes for metabolic engineering. Metab Eng 14(3):242–251. doi:10.1016/j.ymben.2011.09.003, Epub 2011 Sep 18

Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY (2012b) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol 8:536–546

Lee SY, Mattanovich D, Villaverde A (2012c) Systems metabolic engineering, industrial biotechnology and microbial cell factories. Microb Cell Fact 11:156. doi:10.1186/1475-2859-11-156

Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. Nucleic Acids Res 41(22):10668–10678. doi:10.1093/nar/gkt809, Epub 2013 Sep 12

Lin Z, Xu Z, Li Y, Wang Z, Chen T, Zhao X (2014) Metabolic engineering of Escherichia coli for the production of riboflavin. Microb Cell Fact 13:104. doi:10.1186/s12934-014-0104-5

Liu Y, Zhu Y, Ma W, Shin HD, Li J, Liu L, Du G, Chen J (2014) Spatial modulation of key pathway enzymes by DNA-guided scaffold system and respiration chain engineering for improved N-acetylglucosamine production by Bacillus subtilis. Metab Eng 24:61–69. doi:10.1016/j.ymben.2014.04.004, Epub 2014 May 9

Liu D, Xiao Y, Evans BS, Zhang F (2015) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. ACS Synth Biol 4(2):132–140. doi:10.1021/sb400158w, Epub 2014 Jan 10

Meyer FM, Gerwig J, Hammer E, Herzberg C, Commichau FM, Völker U, Stülke J (2011) Physical interactions between tricarboxylic acid cycle enzymes in Bacillus subtilis: evidence for a metabolon. Metab Eng 13(1):18–27. doi:10.1016/j.ymben.2010.10.001, Epub 2010 Oct 8

Moon TS, Dueber JE, Shiue E, Prather KL (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. Metab Eng 12(3):298–305. doi:10.1016/j.ymben.2010.01.003, Epub 2010 Feb 1

Ng CY, Farasat I, Maranas CD, Salis HM (2015) Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration. Metab Eng 29:86–96. doi:10.1016/j.ymben.2015.03.001

Nielsen J (2003) It is all about metabolic fluxes. J Bacteriol 185(24):7031–7035

Nielsen J, Jewett MC (2008) Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res 8(1):122–131. doi:10.1111/j.1567-1364.2007.00302.x, Epub 2007 Aug 29

Nielsen J, Keasling JD (2011) Synergies between synthetic biology and metabolic engineering. Nat Biotechnol 29(8):693–695. doi:10.1038/nbt.1937

Nielsen J, Fussenegger M, Keasling J, Lee SY, Liao JC, Prather K, Palsson B (2014) Engineering synergy in biotechnology. Nat Chem Biol 10(5):319–322. doi:10.1038/nchembio.1519

Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput Biol 10(2), e1003483. doi:10.1371/journal.pcbi.1003483. eCollection 2014

Pang A, Liang M, Prentice MB, Pickersgill RW (2012) Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. Acta Crystallogr D Biol Crystallogr 68(Pt 12):1642–1652. doi:10.1107/S0907444912039315, Epub 2012 Nov 9

Park M, Sun Q, Liu F, DeLisa MP, Chen W (2014) Positional assembly of enzymes on bacterial outer membrane vesicles for cascade reactions. PLoS One 9(5), e97103. doi:10.1371/journal.pone.0097103. eCollection 2014

Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, Warren MJ (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol Cell 38(2):305–315. doi:10.1016/j.molcel.2010.04.008

Registry of Standard Biological Parts. http://partsregistry.org. Accessed Feb 2015

Sachdeva G, Garg A, Godding D, Way JC, Silver PA (2014) In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Res 42(14):9493–9503. doi:10.1093/nar/gku617, Epub 2014 Jul 17

Saier MH Jr (2013) Microcompartments and protein machines in prokaryotes. J Mol Microbiol Biotechnol 23(4-5):243–269. doi:10.1159/000351625, Epub 2013 Aug 5

Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498:19–42. doi:10.1016/B978-0-12-385120-8.00002-4

Sampson EM, Bobik TA (2008) Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J Bacteriol 190(8):2966–2971. doi:10.1128/JB.01925-07, Epub 2008 Feb 22

Sauer M, Porro D, Mattanovich D, Branduardi P (2010) 16 years research on lactic acid production with yeast – ready for the market? Biotechnol Genet Eng Rev 27:229–256

Schendzielorz G, Dippong M, Grünberger A, Kohlheyer D, Yoshida A, Binder S, Nishiyama C, Nishiyama M, Bott M, Eggeling L (2014) Taking control over control: use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways. ACS Synth Biol 3(1):21–29. doi:10.1021/sb400059y, Epub 2013 Jul 15

Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37(2), e16. doi:10.1093/nar/gkn991, Epub 2008 Dec 12

Solomon KV, Sanders TM, Prather KL (2012) A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng 14(6):661–671. doi:10.1016/j.ymben.2012.08.006, Epub 2012 Sep 28

Soma Y, Tsuruno K, Wada M, Yokota A, Hanai T (2014) Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. Metab Eng 23:175–184. doi:10.1016/j.ymben.2014.02.008, Epub 2014 Feb 24

Tsai SL, DaSilva NA, Chen W (2013) Functional display of complex cellulosomes on the yeast surface via adaptive assembly. ACS Synth Biol 2(1):14–21. doi:10.1021/sb300047u, Epub 2012 Jul 10

Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G (2012) The future of metabolic engineering and synthetic biology: towards a systematic practice. Metab Eng 14(3):233–241

Yamanishi M, Matsuyama T (2012) A modified Cre-lox genetic switch to dynamically control metabolic flow in Saccharomyces cerevisiae. ACS Synth Biol 1(5):172–180. doi:10.1021/sb200017p, Epub 2012 Feb 21

Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat Rev Microbiol 6(9):681–691. doi:10.1038/nrmicro1913

Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat Chem Biol 7(7):445–452. doi:10.1038/nchembio.580

Zhou LB, Zeng AP (2015) Exploring lysine riboswitch for metabolic flux control and improvement of L-lysine synthesis in Corynebacterium glutamicum. ACS Synth Biol [Epub ahead of print]

Molecular Modeling and Its Applications in Protein Engineering

8

Emel Timucin and O. Ugur Sezerman

Contents

8.1 Protein Engineering 282
8.1.1 Rational Design 282
8.1.2 Computer Assistance in Rational Design 282
8.2 Molecular Modeling 283
8.2.1 Protein Structure Prediction 283
8.2.2 Docking 285
8.2.3 Molecular Mechanics 287
8.2.4 Molecular Dynamics 289
8.2.5 Free Energy Calculations 293
8.3 Molecular Modeling Applications in Protein Engineering 296
8.4 Summary and Outlook 300
References 301

Abstract

Protein engineering is the field of study aiming to alter proteins to a desired state that may be useful in applications ranging from medicine to industry. In protein engineering it is of interest to predict amino acid substitutions that are critical to the desired feature. Molecular modeling tools have been extensively used for this purpose. Herein, the focus is on such molecular modeling tools for design of new proteins with improved functions and stabilities. The most widely used molecular modeling tools are described and their applications are exemplified.

E. Timucin
Molecular Biology, Genetics and Biotechnology, Faculty of Natural Sciences and Engineering, Sabanci University, 34956 Istanbul, Turkey

O.U. Sezerman (✉)
Department of Biostatistics and Medical Informatics, School of Medicine, Acibadem University, 34742 Istanbul, Turkey
e-mail: Ugur.Sezerman@acibadem.edu.tr

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_8

8.1 Protein Engineering

Protein engineering has emerged as an indispensable tool for design of new proteins and enzymes with improved properties. Many fundamental questions regarding protein folding, stability, activity, and ligand binding can be answered via protein engineering. Along with its implications in basic research, protein engineering has also a wide variety of medical and industrial applications. Important bioproducts such as vaccines, replacement therapeutics, and industrial enzymes can be produced by the application of protein engineering, marking a unique place for this discipline in today's science and technology.

There has been a growing literature on protein engineering approaches which may rely on different principles, trading off rational approaches based on a design procedure by analysis of high-resolution information against random approaches based on selection of a desired feature by directed evolution methods (Carey 1996). Choosing between random and rational approaches primarily relies on the amount of the starting information, such that random approach is applicable to cases where only little information, i.e., primary sequence, is available. On the other hand, if the resolution of the starting information is high, i.e., three-dimensional structure is available, rational approach becomes the popular choice (Guerois and López de la Paz 2006). Protein engineering approaches also differ in their outcomes in which the rational approach provides insights into molecular mechanisms of protein functions, while outcome of the random approach is dependent on screening methodology.

8.1.1 Rational Design

In rational approach, design of novel protein is the first step which involves selection of critical sites responsible for a given protein character. In the design procedure which is an important test of our understanding of structure and function complementarity, there is an obvious logic. Choosing sites that are close to the active site or binding pocket of an enzyme is critical to design of activity and selectivity, while choosing the sites that would contribute to the structural stability of a protein is crucial for design of stability. To select such sites, one needs to examine the atomistic details thoroughly, limiting the success of design procedure to the resolution of information on the protein of interest. Therefore, the availability of 3D protein structures is imperative for design. Furthermore, the structures in which the protein is captured during an act, i.e., bound to its substrate or analogue, are particularly beneficial for design of enzyme activity and selectivity.

8.1.2 Computer Assistance in Rational Design

Before altering the composition of a protein, it is essential to have a well-defined rationale for the design process which is mostly established through computers.

Computers can provide powerful service for understanding the atomistic details of biological phenomena ruled by proteins and their partners. Technically speaking, proteins are sets of Cartesian coordinates varying at positions (x, y, z) of N, O, C, S, and H atoms of hundreds of amino acids that constitute them. Protein structure is determined by interactions between these atoms and has many possible configurations even for a small fragment which requires use of computers to search through these configurations in an efficient manner. Thus, even if a small portion of protein structure is considered for analysis, computers are the only convenient tool to deal with such large number of datasets. In this respect, computer-assisted approaches become a requisite modeling tool for protein engineering studies since they provide novel insights to complex systems which may be challenging to be understood by experimental studies (Lazakidou 2010; Park and Cochran 2010). Computers can also be used as a convenient viewing interface examining the very details of protein structure. Therefore, the advances in the computer technologies have a direct impact on the success of rational design approaches for prediction of the regulatory regions in proteins mediating stability, function, or selectivity.

8.2 Molecular Modeling

Molecular modeling is the process of analyzing and modeling the atomic level information of biological macromolecules and their interactions through the use of computational methodologies. These studies often combine several in silico methods, ranging from bioinformatics analyses to molecular simulations, to describe structure–function properties and predict beneficial mutations for protein engineering purposes. Here we review the molecular modeling tools that are of particular use in rational protein engineering applications.

8.2.1 Protein Structure Prediction

Many proteins, especially membrane proteins, may not be suitable for structure determination methods such as NMR and X-ray diffraction (Bill et al. 2011). Concurrently, application of these experimental methods would be very labor intensive and time consuming (Floudas et al. 2006). Such limitations faced during protein structure determination procedures accentuate molecular modeling approaches as an exclusive solution for obtaining structural information for these particular “challenging” cases.

Hitherto numerous efforts have been devoted for improvement of computational approaches for structure determination including comparative modeling (Sanchez et al. 2000), ab initio prediction (Bonneau and Baker 2001), and fold assignment (Lemer et al. 1995) methods. Ab initio methods try to construct the structure of a target protein from scratch through establishment of 3D structure and energy relationships (Hardin et al. 2002), while fold assignment methods thread the

query sequence to all the available folds and identify the most suitable fold according to the selected target energy function (Jones 2000). Homology modeling which is an augmented threading method, on the other hand, uses sequence alignment of target protein with a template of known structure (Sanchez et al. 2000). Ab initio prediction methods are classified as de novo approaches, while threading or homology modeling is classified as comparative approaches. Although de novo approaches have predicted structures of small protein domains correctly, they may be computationally demanding especially for proteins formed by more than 150 residues (Sanchez et al. 2000). Therefore, comparative approaches including homology modeling become the most convenient and widely used methods for protein structure determination, particularly for the cases when the ab initio methods are not feasible and threading methods are not effective.

Protein structure prediction methods are among important molecular modeling tools that provide invaluable insights into functional annotations of unknown proteins and atomistic details of critical protein regions for stability, activity, and selectivity (Ceulemans and Russell 2004; Ginalski et al. 2004). Moreover, these methods are also widely used for producing models to assist time-consuming experimental methods such as X-ray crystallography, NMR, and electron microscopy (Teichmann et al. 1999).

8.2.1.1 Homology Modeling

Homology modeling methods use the fact that evolutionary-related proteins share a similar structure (Chandonia and Brenner 2005; Vitkup et al. 2001). The method predicts 3D structure of a target protein by the use of its sequence alignment with an experimental known structure of a related homologous protein which is referred to as "template" (Vyas et al. 2012). Several programs and servers are available for homology modeling that are planned to build a complete model from query sequences. MODELER (Sali and Blundell 1993), SwissModel (Peitsch 1996), I-TASSER (Zhang 2007), and ROBETTA (Kim et al. 2004) are some of the examples.

Homology modeling can generally be performed in six steps. These are listed as follows:

1. Template selection
2. Alignment
3. Backbone generation
4. Loop modeling
5. Side-chain modeling
6. Validation

These steps could be repeated until the most appropriate model is built. Valuable structural insights can be gained from homology modeling studies, yet the quality of the model is directly dependent on the sequence similarity with the template structure (Hilbert et al. 1993) such that the models built over 50 % sequence similarities are accurate enough for sensitive applications including drug discovery

while the identity between 25 % and 50 % can be useful in mutagenesis experiments and lower identity percentages than 25 % are only accepted conditionally (Francoijs et al. 2000; Takeda-Shitaka et al. 2004). After obtaining the alignment with a suitable template, the method proceeds with generation of the protein backbone model. This coarse model of the protein structure is followed by loop modeling and side-chain arrangements which fine-tune the model. Validation of the model is the last step and generally requires assessment of several parameters including solvation potentials testing for the misfolded regions (Holm and Sander 1992) and the Ramachandran plots for the backbone quality (Hooft et al. 1996).

8.2.2 Docking

A biological function is principally realized through a molecular binding event in which one molecule binds to a pocket or surface of another molecule which is usually a protein. The resulting intermolecular complexes exhibit particular features such as geometrical and chemical complementarities which carry paramount implications for design of biological activities (Wang and Kollman 2001; Ju et al. 2009; Werner et al. 2010). Molecular docking, one of the key tools in molecular modeling, predicts three-dimensional (3D) models for intermolecular biological complexes in the bound state. Although there has been a great increase in the protein structures deposited in the protein data bank (PDB) in the last decade (Bernstein et al. 1977), a development which closes the gap between the genomic and the proteomic data, the number of complex molecules that can arise from those protein structures is still very large to handle in experiments. Since the pioneering studies (Kuntz et al. 1982), molecular docking approaches have fulfilled the need to identify key complex molecules that are biologically relevant to us.

Molecular docking approaches predicting 3D models for complex molecules in biological systems are extremely beneficial for our understanding of protein activities. From the protein's perspective, owing to the availability of the receptor–ligand complex, it would be possible to identify the binding modes of the given ligand. Moreover, modes of action and/or reaction mechanisms of enzymes could also be resolved through molecular docking studies. From the ligand's perspective it would be possible to identify the contact and the interactions it establishes in the binding pocket, enabling one to design novel ligand analogues with desired features. Therefore, molecular docking, one of the key tools in molecular modeling, plays an important role in the rational design of inhibitors for receptors as well as substrates for enzymes.

8.2.2.1 Methodology

Docking spots intermolecular complexes through the search of the ligand's all possible geometries and binding energies in a receptor. The protocol starts with the search for accessible protein surfaces to the ligand, the step which is called pose generation, and ends with the ranking of different ligand conformations through a scoring function, the step which is called pose selection. During pose generation,

one should be able to allow ample degrees of freedom of the ligand to sufficiently sample possible conformations within the binding region of protein and to produce true binding mode. Generation of the true bound conformation of the ligand is solely not enough. A successful docking study should also distinguish the true binding mode among all of the other generated modes. This task is performed in the scoring step. During scoring, all of the generated binding modes of the ligand are assessed and ranked quantitatively to unveil the true binding conformation with the highest score. Although pose generation step does not necessarily require information about the binding energy of complex, the scoring step requires this information. Scoring functions used in protein–ligand docking protocols can be force field-based (see Sect. 8.2.3), empirical, and knowledge-based functions. They calculate the ligand–receptor interactions by means of electrostatic interactions, hydrogen bonding, and desolvation energies over a predefined grid including the binding region (Sousa et al. 2006).

A successful docking can be achieved by a combination of the best search algorithm with the best scoring function, and also an unsuccessful docking study may succeed by the use of another docking algorithm with different search and/or scoring functions. The performance of the docking methods has been shown to highly depend on the protein–ligand system used (Clark et al. 2002; Halperin et al. 2002; Schulz-Gasch and Stahl 2003). Owing to the parameterization of the scoring functions, if one docking program performs better with the given ligand, then it would also perform with similar ligands at comparable accuracies.

Versatile docking algorithms are available both with academic and commercial licenses. Among these, AutoDock (Scripps), one of the most widely used small-molecule docking algorithms, predicts dominant intermolecular complexes of proteins and ligands (Goodsell et al. 1996). In this method, the search process starts from a random outside region of the binding site and continues by the ligand's movements until it finds a bound conformation. Although a traditional genetic algorithm (GA) approach is used for the global minimum of the bound complex in the earlier versions, Morris et al. (1998) describe a new approach that employs Lamarckian genetic algorithm for global search and combines it with a local search procedure described previously (Solis 1981). This new version, distributed as AutoDock Vina, can handle ligands with a larger number of degrees of freedom than the routine GA method can. Therefore, AutoDock Vina is better suited to large ligand docking with increased accuracies at improved speeds compared to the older versions of AutoDock (Morris et al. 1998). Other docking algorithms including FlexX (BioSolveIT) (Rarey et al. 1996) and Dock (UCSF) (Ewing et al. 2001) employ a different search method than AutoDock. They use an incremental construction of the ligand in the bound state by placing a rigid anchor in the binding site and using a greedy algorithm to combine the ligand fragments to reconstruct the ligand structure. Although these algorithms differ in the pose generation step, they all use scoring functions based on force fields during selection of the binding mode of the ligand.

Majority of the efforts in docking studies are devoted to predict protein–ligand complexes owing to fundamental roles of small ligand molecules in mediating

protein functions (Muegge and Rarey 2001). Yet another class of docking problems mainly arises from predicting protein–protein complexes in which the ligand turns into a peptide chain decorated with fully flexible side chains. Thus protein–protein docking methods require upgrade of the ligand computations to include flexibility of both backbone and side chains, conformational changes, and also entropic terms (Andrusier et al. 2008). HADDOCK (Dominguez et al. 2003) and Rosetta (Wang et al. 2007) methods are among the widely used protein–protein docking algorithms that are tested in CAPRI[1] challenges (de Vries et al. 2010) which are wide community experiments for prediction of structures of unknown protein–protein complexes.

8.2.2.2 Flexibility Problem of Receptors

Most of the docking methods use a flexible ligand and a rigid protein allowing only the ligand (substrate or drug molecule) to change its orientation depending on its torsional degree of freedom during calculation (Teodoro and Kavraki 2003). As the degrees of freedom, i.e., flexibility, decrease, the computational time required to determine the docked configuration increases. Docking calculations with flexible receptor and flexible ligand docking would suffer from computational costs owing to the large number of degrees of freedom within the flexible receptor, while the use of only flexible ligands whose torsion trees possess fewer degrees of freedom than the receptor costs relatively moderate. In this respect, a trade-off between time/cost and accuracy presents rigid protein docking as the most frequent route for current docking studies. Despite a gain in the computational costs, rigid protein docking might lead to failures in generation of the correct ligand poses and to meaningless scores of the generated poses (Erickson et al. 2004). Therefore, addition of protein flexibility is critical to the prediction of the correct binding mode and also to the reliable scoring of poses (Totrov and Abagyan 2008).

There are some solutions to incorporate tolerable movements of proteins in routine docking methods. Certain degree of flexibility could be maintained in the critical regions of the proteins where the ligand readily interacts, for instance, an active site of an enzyme during substrate docking. In addition to the novel docking approaches incorporating protein flexibility by the use of multiple protein conformations (Totrov and Abagyan 2008; Fischer et al. 2014), other molecular modeling tools such as molecular dynamics and energy minimization algorithms can also be combined by molecular docking approach to overcome the limitations faced in incorporation of the receptor flexibility.

8.2.3 Molecular Mechanics

Molecular mechanics is a computational way of establishing potential energy formulations that describe an atom system. Instead of dealing with the wave

[1] *C*ritical *A*ssessment of *PR*ediction of *I*nteractions (http://www.ebi.ac.uk/msd-srv/capri/)

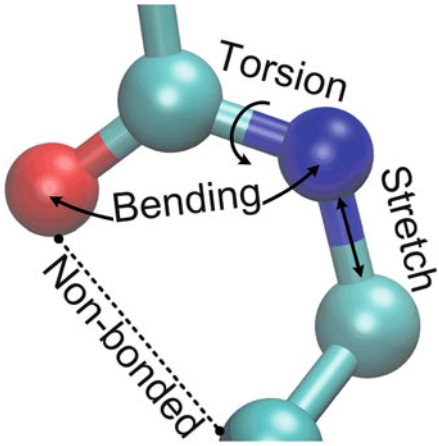


Fig. 8.1 Energy terms for simple molecular mechanics. Bonded interactions include stretching, bending, and torsions, while nonbonded atoms (greater than two bonds apart) interact through van der Waals attraction and electrostatic attraction/repulsion

function that would represent a realistic view of microscopic states, molecular mechanics uses classical mechanical approaches to derive potential energy of atoms. The main reason for this compromise is the fact that the use of classical physics requires much less computational power than a quantum mechanical calculation making the molecular mechanics models very useful for large molecules such as proteins.

Building and/or manipulating in silico protein structure requires knowledge of bond, angle, torsion, and atomic radii of protein atoms. In molecular mechanics, a wire model is built for a biological molecule through definition of potential energy by a number of formulations computing time-dependent changes in bond lengths, bond angles, and torsions and also nonbonding interactions between atoms (Fig. 8.1). In the simplest representation, the potential energy function is given as follows:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} \tag{8.1}$$

where E_{bonded} is the energy of the bonded interactions and $E_{\text{nonbonded}}$ is the sum of the potential energy of nonbonded interactions such as hydrogen bonding and van der Waals interactions. E_{bonded} in which the bonded interactions are treated as simple harmonic potentials can be further decomposed as follows:

$$E_{\text{bonded}} = E_{\text{D}} + E_{\text{B}} + E_{\text{T}} \tag{8.2}$$

where E_{D} is the energy of the bond deformations as in stretches or compressions, E_{B} is the energy of bending in the bond angles, and E_{T} is the energy involved in torsion of the bond. These equations of the potential energy are defined as "force field" and rely on experimental parameters such as force constants and equilibrium values.

The potential energy of a rigid protein was previously described by a number of force fields such as CHARMM (MacKerell et al. 1998), AMBER (Duan et al. 2003), and GROMOS (Oostenbrink et al. 2004). These force fields are

developed for special purposes such that CHARMM and AMBER model protein dynamics while GROMOS works better for modeling lipid behaviors. Although these force fields are developed to model native interactions, they may fail to accurately describe all of the properties possessed by biological systems. Studies assessing performance of the force fields agree that in the cases where the force field parameters can be experimentally determined, different force fields perform all well regardless of the tasks that they have been designed for (Price and Brooks 2002). On the other hand, for the cases where experimental data is insufficient for validation of the parameters, the force fields perform well only at their specialized tasks (Schlick 2010). Another point is that all of these force fields neglect the polarization effect in which the electronic properties of the atoms are omitted from the formulations. Such omission makes these force fields essentially not applicable to the cases where the polarization effect is overwhelming due to the presence of highly charged atoms, e.g., modeling of an ion transport through a membrane (Bastug and Kuyucak 2006, 2007b).

8.2.4 Molecular Dynamics

Molecular dynamics (MD) is a powerful computational method that describes equilibrium and dynamics of a biological system. MD generates configurations of a system by integration of classical mechanics for calculation of the time dependence of a system. This technique permits studies regarding the dynamics of a biological system and thus has broadened our understanding of motions of biological molecules which are often inaccessible for experimentation (Karplus and Petsko 1990; Dodson et al. 2008). Although the first molecular dynamics study dated back to the 1950s when Alder and Wainwright reported the simulations of hard spheres revealing valuable insights to the behavior of simple liquids (Alder and Wainwright 1957, 1959), the first protein simulation was carried out by McCammon et al. by the use of the bovine pancreatic trypsin inhibitor (McCammon et al. 1977). Since then MD simulations have become a routine technique addressing a molecular understanding to several different biological phenomena ranging from molecular recognition, molecular transport to protein folding. Apart from the ability of MD to describe the dynamics of a biomolecular event, the method can also assist experimental procedures for protein structure determination such as X-ray crystallography and NMR techniques.

8.2.4.1 Basics

MD simulations are performed with a set of force field functions that correspond to the potential energy of a static structure. The method relies on classical mechanics in which the equation of motion is described as $F=ma$, where F is the force exerted on the particle, m is the mass, and a is the acceleration. At any temperature above absolute zero, the total energy of a protein molecule is composed of potential and kinetic energy of the thermal motions of the atoms. This motion can be generated via computers by calculating the derivatives of the potential energies, i.e., forces,

through numerical integration. Once the positions are known and the velocities are predicted, it would be possible to determine the state of the system for a future or a past point. The motion trajectories that are the collection of successive frames of the atomic positions and the velocities are obtained for very short integration steps, mostly on the order of femtoseconds (s^{-15}).

In most MD simulations, the molecules are surrounded by boundaries which act as symmetry operators to generate periodic boundary conditions. These periodic boundary conditions enable the simulation to be completed using a relatively small number of particles such that these particles experience forces as they are in bulk solvent. To model a more realistic case, the protein system could be solvated in water. There are usually two options for describing the solvent in MD simulations: the solvent can be treated explicitly by considering the molecular details of each solvent molecule separately, i.e., TIP3P model for water (Jorgensen and Madura 1983), or the solvent can be treated implicitly as a continuous medium.

8.2.4.2 Statistical Ensembles in Molecular Dynamics

A biological phenomenon is characterized through certain macroscopic parameters such as heat, energy, and pressure of biological systems. Using such parameters, one can determine the energetics and the mechanisms of protein activities including ligand binding and folding. However molecular dynamics simulations can provide information at the microscopic level such as atomic coordinates and velocities. Nevertheless, the use of mathematical expressions obtained from statistical mechanics permits a connection between the microscopic information obtained from MD trajectories and the macroscopic properties.

In the classical sense, biology experiments are performed with macroscopic samples that comprise a large number of molecules, e.g., 1 ml of water in a test tube is composed of more than 2×10^{21} molecules. Such a great number of molecules representing the same molecular system would sufficiently sample all of the conformations and energies experienced by this system. Therefore, a biology experiment always represents an average of the conditions realized by a system. Analogous to the macroscopic observations, statistical mechanics uses "ensembles" as the collection satisfying all possible systems with different microscopic properties that can be generated by a single macroscopic condition. In this sense molecular dynamics simulations produce different microscopic states of a unique ensemble in a time-dependent manner. For instance, distinct conformations experienced by an enzyme–substrate complex could be identified through molecular dynamics simulations. Although all of the conformations represent the ground state in which the enzyme and the substrate are physically bound, the availability of different conformations of the same enzyme–substrate complex achieves sufficient sampling and thus enables molecular dynamics simulations to represent an analogous view to macroscopic samples.

8.2.4.3 Running a Molecular Dynamics Simulation

Development of different force fields results in many different molecular dynamics simulation packages (Karplus and McCammon 2002), among which NAMD is one

of the most widely used software packages that uses Charm++ parallel programming model (Phillips et al. 2005).

Generic input files for setting a molecular dynamics simulations include information on the structure and the initial coordinates of the system, e.g., pdb file obtained from NMR, X-ray crystallography, and/or homology modeling studies to build the system with the attributed force field parameters in silico. Before the molecular dynamics simulations, the generated system is energy minimized through use of versatile algorithms such as Newton–Raphson, steepest descent, and conjugate gradient methods. The main purpose behind energy minimization step is to relax the steric clashes caused during generation of the simulation system and also to adjust the system to the applied force field. After energy minimization, the system is ready for initialization of molecular dynamics. During initialization, temperature is increased from 0 K to the desired temperature. Heating is achieved by assigning velocities from a Maxwell–Boltzmann distribution which is renewed for every integration time. In the initialization, kinetic energy might become dominant over potential energy because of heating, emphasizing the necessity of an equilibration of the energetics of the system before data collection. Only after the desired temperature is reached, equilibration simulation permitting an equilibrated kinetic and potential energy can be carried out. During equilibration, a boost in the potential energy with increasing temperature should be observed, confirming an energy transfer from kinetic to potential and equilibration simulations may be carried out until the potential energy remains flat. Finally, production simulations represent data collection step where the protein dynamics including bond fluctuations, domain motions, and conformational rearrangements occur are to be investigated using fully energy-minimized and equilibrated systems.

Different statistical ensembles can be used depending on the purpose of MD simulations. The most common ones used in the simulations of biological systems can be listed follows:

1. NVE ensemble which is characterized by fixed atom number N, volume V, and energy E
2. NVT ensemble which is characterized by fixed atom number N, volume V, and temperature T
3. NPT ensemble which is characterized by fixed atom number N, pressure P, and temperature T

NVE ensemble also called microcanonical ensemble represents a fully isolated system with conserved energy and without any pressure or temperature control. This ensemble with constant energy is usually not recommended for equilibration simulations owing to the need of energy flow for temperature control during equilibration simulations. However, when the desired temperature is achieved after equilibration and NVE, simulations could be useful for investigation of constant energy surface of conformational spaces and also for other cases where any perturbations arising from temperature and/or pressure control to the total energy can be acceptable (Leach 2001).

The simulation systems are usually not large enough to represent microcanonical ensembles where temperature fluctuations are negligible and temperature can be considered as constant. NVT simulations, also referred to as canonical ensembles, are easier to implement than NVE simulations (Leach 2001). NVT ensembles maintain constant temperature due to a temperature-bath coupling system. For example, a thermal denaturation experiment could be replicated when precise temperature control is maintained in an NVT ensemble. Moreover, if one is interested in the given density and temperature, NVT simulations are usually useful for exploring conformational space of molecules under vacuum where pressure is not a significant factor.

NPT ensembles maintain both constant pressure and temperature where the volume, i.e., the unit cell vectors, is allowed to change under constant pressure. This ensemble type, also referred to as isobaric–isothermal ensemble, is particularly useful when the density parameter is unknown. Thus, NPT ensembles can be used during equilibration simulations to achieve the desired temperature and density. Moreover, NPT ensemble could also be applicable to cases when the experimental conditions that include constant pressure such as NMR are to be implemented.

8.2.4.4 Time Scale Limitations in Molecular Dynamics

Many of the protein motions can be probed by molecular dynamics simulations (Fig. 8.2). Among these, fine motions such as bond fluctuations that occur in the femtosecond scale (10^{-15} s) and side chain and loop motions observed in the picosecond (10^{-12} s) and the nanosecond (10^{-9} s) regime, respectively, can be modeled effectively in conventional MD simulations. On the other hand, large conformational changes which are characteristic to the most of the complex protein activities including folding/unfolding and ligand binding occur in microsecond (10^{-6} s) to millisecond (10^{-3} s) time scales. Because slow protein motions require modeling of large free energy jumps (Fig. 8.2), characterization of the critical processes in which proteins fold, bind, and/or react is beyond the reach of conventional molecular dynamics. Nevertheless, computers are becoming faster and cheaper, and still simulations of biological systems up to the millisecond time scale are rarely reported (Shaw 2013). Thus, despite the wide applications of MD method in protein science, this method can be time consuming and computationally expensive especially for identification of complex protein mechanisms that occur much slowly than those that are effectively modeled by MD.

To overcome time scale limitations faced in the current molecular dynamics method, a number of specialized MD approaches that rely on adaptive methods are developed. Among these approaches, those particularly designed for determination of free energy changes of biological processes are the most widely studied ones and will be reviewed next.

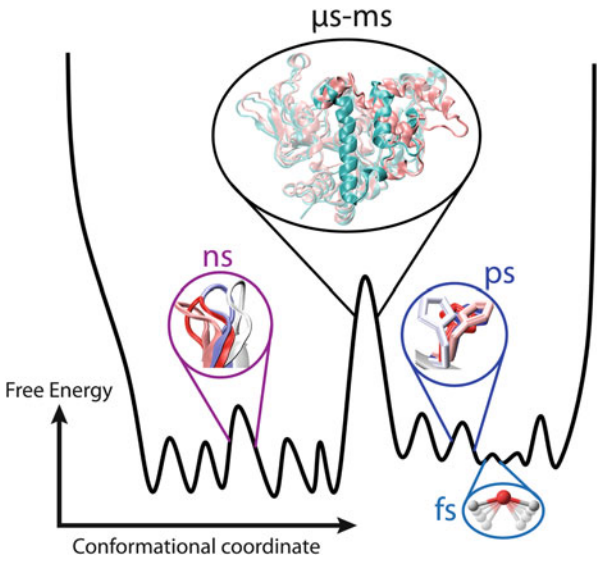


Fig. 8.2 Energy and time barriers faced by protein motions. Different protein motions are shown in free energy and conformational coordinate plot. Fast protein motions including bond vibrations (fs), side chain (ps), and loop (ns) movements can be effectively modeled by molecular dynamics using the current time scales, while slower motions (μs–ms) cannot (Henzler-Wildman and Kern 2007). An example of such slow motion is shown with cartoon representations in which the closed state of a lipase (PDB ID: 1KU0) colored *cyan* changes its conformation by movement of two large helices to the open state (PDB ID: 2W22) colored in *pink*

8.2.5 Free Energy Calculations

Free energy constitutes the most important thermodynamic quantity that describes a chemical process. Free energy change between a reference and a target state ΔG can be used to calculate the equilibrium constant K of biological processes such as folding and ligand binding at a given temperature T according to the Eq. (8.3):

$$\Delta G = G_{\text{final}} - G_{\text{initial}} = -k_{\text{B}} T \ln K \tag{8.3}$$

where k_{B} is the Boltzmann constant.

The theory of free energy calculations/approximations has been developed since the 1930s (Kirkwood and Alder 1968; Kirkwood 1935). Although computational limitations faced at that time narrowed the applications of free energy calculation methods, today continuous increase in the speed of computers along with the current developments in the parallel computing algorithms advances free energy calculations to a new level with widened implications especially in molecular biology (Chipot and Pohorille 2007).

Several biasing methods for determination of free energy changes are available with the use of molecular dynamics theory through coupling of user-defined force (s) in the force field. Compared with the conventional molecular dynamics method,

such approaches would permit to study particularly difficult biological problems such as ligand binding and protein folding at moderate computational costs and thereby overcome energy and time barriers faced during modeling of complex biological processes.

In general there are two types of free energy calculations depending on the path followed during calculation. Path-independent methods include free energy perturbation and thermodynamic integration, while path-dependent methods include steered molecular dynamics and umbrella sampling. Among these, one popular example from each type, free energy perturbation and steered molecular dynamics will be surveyed in the coming sections.

8.2.5.1 Free Energy Perturbation

Free energy perturbation (FEP) is a method used for calculation of free energy changes of a dynamic process shuttling between two equilibrium states (Fig. 8.3). The method is based on running a standard molecular dynamics simulation for an initial state which is perturbed at each time step until formation of a final state. However, if the difference between the final and the initial states is large, the method might suffer from hysteresis effect which might lead to inaccuracies in the computation of free energy changes (Wood et al. 1991; Edholm and Ghosh 1993). FEP-based methods calculate relative free energy changes rather than absolute free energies. Still obtaining relative free energy differences enables one to directly calculate mutational effects by simply ranking the relative changes in the free energies, potentiating FEP methods in protein engineering applications particularly for rational design purposes. Therefore, although FEP method is not practical for applications of ligand binding problem particularly concerned with large ligand models, it is applicable to mutagenesis studies which require perturbation of one amino acid to another.

8.2.5.2 Steered Molecular Dynamics

Steered molecular dynamics (SMD) is a special molecular dynamics application in which a system is driven from one equilibrium state to another along a reaction coordinate by a drag force (Isralewitz et al. 2001; Park et al. 2003). During SMD simulation, one part of the system, i.e., ligand, is coupled with a harmonic oscillator. The spring attached to the ligand is pulled at constant force or velocity (Fig. 8.4).

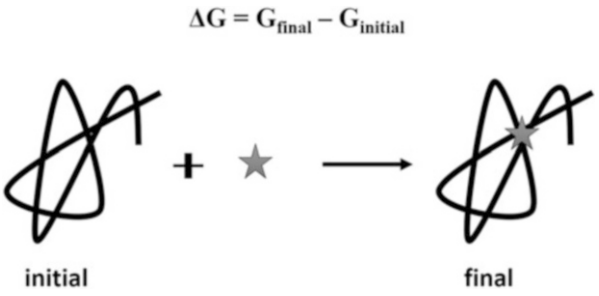


Fig. 8.3 Free energy difference between two states. The free energy change for the transition of system from one state to another is given in Eq. (8.3) central to path-independent methods

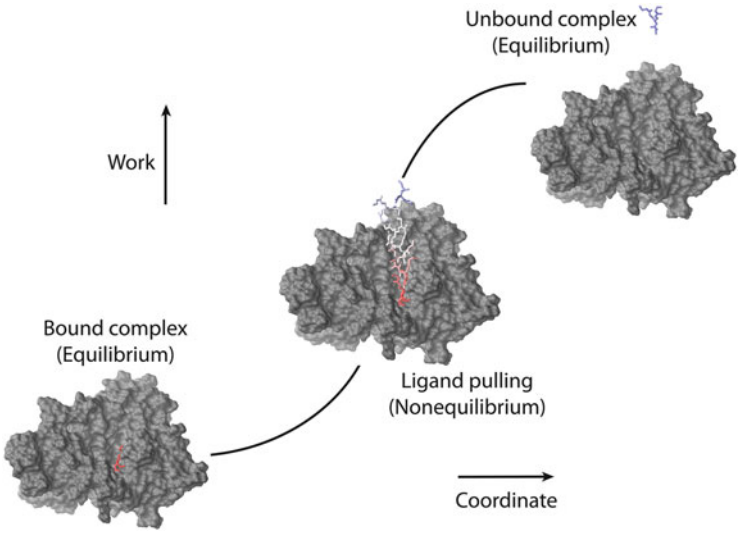


Fig. 8.4 Graphical representation of SMD application: steering the ligand. The protein–ligand complex initially stays in the bound conformation where the ligand is colored *red* in the cleft. After application of an upward drag force to the ligand, the equilibrium state is perturbed and subsequently a series of nonequilibrium states in which ligand conformations are colored according to time step (*red to white*) are generated. Pulling event is lasted until the ligand colored *blue* is completely pulled into the bulk solvent where it does not have any contacts with the protein

The movement of this spring builds up a certain tension which may lead to a molecular transition, e.g., unbinding and unfolding. The force experienced by the pulling event at constant velocity is given as follows:

$$F(t) = \left[vt - z(t)\right] \tag{8.4}$$

where F is the force, v is the velocity, and z is the displacement from the initial time to time t. The line integral of the Eq. (8.4) gives the work done on the system to move it from one state to another. According to Jarzynski's equality, the Boltzmann average of work functions obtained from the SMD simulations provides an estimate of the free energy of the process along the reaction coordinate, which approaches the true value when the system is sampled sufficiently (Jarzynski 1997). This equality is given as follows:

$$e^{-\Delta G/k_{\mathrm{B}}T} = \left\langle e^{-W/k_{\mathrm{B}}T} \right\rangle \tag{8.5}$$

where ΔG is free energy difference, W is the work function, k_{B} is the Boltzmann constant, and T is temperature; the triangle bracket denotes an average of all samplings tried. This particular equality provides a powerful tool for the determination of equilibrium free energy changes in the context of steered molecular

dynamics simulations. Following the Jarzynski's equality, other formulations are also established and can be used for similar purposes (Crooks 1998).

While SMD simulations are less laborious and easier to implement compared to other path-dependent approaches such as umbrella sampling MD simulations, they require relatively more computing time to obtain the same level of accuracy (Bastug et al. 2008; Bastug and Kuyucak 2007a). Nevertheless, this approach may be applied for the cases where comparison of two systems with either different ligands or different macromolecules is considered. The nonequilibrium methods like Jarzynski's equality are obviously useful for the cases where nonequilibrium conditions are dominating and cannot be easily avoided.

8.2.5.3 Other Approaches for Calculating Free Energy Changes

Apart from sophisticated software and formulations for determination of binding free energies, one can also employ a relatively fast and a similarly quantitative estimation of binding free energy changes. FoldX is particularly developed as a protein design tool for quantifying interactions contributing to the stability of proteins and/or protein complexes using a very large set of point mutants that spans most of the structural environments found in proteins (Guerois et al. 2002). Different energy terms taken into account in FoldX have been weighted using empirical data obtained from protein engineering experiments. FoldX created by Guerois et al. is also available as a web server (Schymkowitz et al. 2005).

There are other sequence-based methods that can predict impact of mutations on stability by using several sequence features and machine learning algorithms. I-Mutant (Capriotti et al. 2005), PoPMuSiC (Dehouck et al. 2009), and SDM (Worth et al. 2011) are among examples of such algorithms that are frequently used in prediction of stability-enhancing mutations.

8.3 Molecular Modeling Applications in Protein Engineering

Apart from the current developments in docking algorithms, other molecular modeling tools including molecular dynamics and energy minimization algorithms that incorporate full protein flexibility can be applied to overcome limitations faced during rigid protein docking. Combination of docking and molecular dynamics approaches has been widely encountered in the current literature in prediction of correct ligand and protein poses of biologically important complex structures.

Kua et al. (2002) combined docking and molecular dynamics approaches to investigate the substrate specificity of the enzyme acetylcholinesterase. They used short molecular dynamics simulations without application of any external forces to produce sufficient number of conformations of the enzyme which served as receptors during docking of multiple ligands. Their findings provided insights into not only the binding geometry and the relative binding affinity of acetylcholinesterase substrates but also the interplay between substrate binding and hydrolysis.

In another study, a combined computational methodology is applied to a serine/threonine protein kinase, an attractive target due to its involvement in cancer development (Lv et al. 2015). They combined docking and molecular dynamics approaches with binding free energy calculations to decipher the mechanism of inhibition of this kinase. Their results suggested that the interactions of the kinase with the inhibitor molecules are characterized by hydrophobic residues, providing insights into the rational design of novel inhibitors for serine/threonine protein kinases that can be used as cancer therapeutics.

In steered molecular dynamics, the movement experienced by the initial state destroys the equilibrium and generates a series of nonequilibrium states until the system has reached another equilibrium state, i.e., the final state (Fig. 8.4). Thus, since Jarzynski has postulated an equality between nonequilibrium work functions and equilibrium free energy differences in a system (Eq. 8.5) (Jarzynski 1997), steered molecular dynamics approach has become a popular nonequilibrium method for free energy calculations with numerous applications in molecular biology ranging from ligand binding, enzyme–substrate interactions to protein folding (Patel et al. 2014; Durmaz et al. 2013; Bu et al. 2011; Pappalardo et al. 2007; Pesce et al. 2011; Pongprayoon et al. 2009). Moreover, because the reaction path is a priori defined in SMD method, its applications can be considered as in silico counterparts of single-molecule experiments that utilize atomic force microscopy or laser tweezers (Phillips et al. 2005). Thereof owing to a wealth of information provided by the trajectory data of SMD simulations, SMD method can be utilized, for example, in searching for a ligand's path in the binding cleft (Yang et al. 2009; Perakyla 2009) and also for the interacting residues along a path (Patel et al. 2014). Such information would be very useful in identifying critical residues for rational design of enzyme activity, selectivity, and stability.

Recently a protein engineering study reported modification of the substrate selectivity of the lipase originating from *Bacillus thermocatenulatus* (BTL2) (Durmaz et al. 2013). They introduced a novel approach in which the binding free energies of the enzyme–substrate complexes at the ground (ES) and the transition ($ES^{\ddagger}$) states have been determined from SMD simulations and experiments, respectively. The use of the SMD method provided the prediction of not only the binding affinities of ES complexes but also five candidate mutations from the cleft to enhance the catalytic preference of tributyrin (C4) substrate. Apart from the application of the SMD approach for binding free energy calculation, Durmaz et al. also presented applications of molecular docking method through which the ES complex of the lipase with C4 and C8 substrates and molecular dynamics simulations for testing the stability of the ES complexes under full receptor flexibility were generated. Their rational design strategy was based on generating lipase variants that block the substrates with longer acyl chains than C4 (Fig. 8.5). The mutations were predicted using SMD simulations through tracing the pair distances between the catalytic cleft residues and the ligand atoms, among which the most persistent contacts were selected and mutated to a large bulky amino acid to block the C8 chains but not C4. In the experimental section, the steady-state specific activities toward C4 and C8 substrates corresponding to the

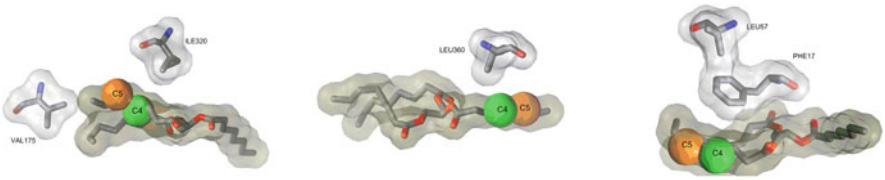


Fig. 8.5 Rational design strategy for modifying chain-length selectivity. Five different residues (V175, I320, L360, L57, and F17) located at the binding pockets of three acyl chains of the C8 substrate are shown. The fourth and the fifth carbons are rendered as van der Waals spheres and colored *green* and *orange,* respectively. Four mutations (L57F, V175F, I320F, and L360F) are designed to generate a steric block for the substrates with longer chains than C4. Three control mutations (V175A, I320A, and F17A) are also performed to test the design strategy. The C8 substrate and the residues are shown in licorice (C, *gray*; N, *blue*; O, *red*)

free energy change of the transition state were measured. The binding free energies of the enzyme–substrate (ES) complexes obtained from SMD simulations and those of the transition state ($ES^{\ddagger}$) complexes measured in experiments were combined in the Gibbs free energy profiles. Durmaz et al. assessed the catalytic efficiencies of eight different lipase variants including the native lipase and achieved the most prominent increase in the catalytic preference of tributyrin (C4) over tricaprylin (C8) in the L360F mutant with -3 kcal mol^{-1} of free energy change. Owing to the high-value products requiring short chain fatty acids in their manufacture (Shetty 2006; Durre 2007), enhancing the catalytic preference of lipases toward short chain substrates as in the L360F variant of BTL2 carries immense potentials for industry.

Modification of catalytic preference of lipases requires knowledge of the kinetic parameters governing fat hydrolysis by lipases (Malcata 1996). However, due to the fact that the enzyme–substrate (ES) complex during this reaction forms at the water–lipid interface (Verger and Haas 1976) unlike most of other enzymes that follow simple Michaelis–Menten kinetics, experimental determination of the binding free energy of the ES complexes formed by lipases is very challenging for lipases owing to physical properties of their fat substrates (Malcata 1996; Hoppe and Theimer 1996). Nevertheless, this particular study has overcome this challenge by the use of SMD simulations with Jarzynski's equality for the calculation of the binding free energy of the ES complex via single-molecule models. Therefore, the described method (Durmaz et al. 2013) can be applied not only to other enzymes to design their substrate selectivity but also to other proteins bearing experimental challenges in identification of their complex reaction mechanisms.

In another example, the product expulsion mechanism of the cellulose-degrading enzyme (Cel7A) has been investigated using multiple free energy calculations methods (Bu et al. 2011). They used FEP/MD and steered MD with Jarzynski's equality to calculate the binding free energy of the cellobiose in the catalytic tunnel of Cel7A and both of the methods converged. With the calculated free energy changes, they were able to verify the experimental finding showing that the glucose expulsion is less favorable than cellobiose expulsion (Gruno et al. 2004), bolstering the reliability of the FEP and the SMD methods for free

energy calculations. Although SMD method becomes less practical if the binding site is deeply buried due to challenges in the description of a simple and clear path for a ligand (Bu et al. 2011), this study adjusted the steering force based on the current conformation of the enzyme and the product. Thus, their approach represents a valid alternative to other enzyme design cases which also bears a buried binding site for the substrates and/or products. Upon analysis of the SMD trajectories, they identified five candidate residues that strongly interacted with the cellobiose in the catalytic tunnel during expulsion. Identification of these sites is of great importance for engineering the protein to the meet the needs of the commercial process. Thus, they proposed that alanine substitutions of these five residues would result in inhibition of cellobiose interaction with the catalytic tunnel and thus hastening of the product expulsion process. In this respect five alanine substitutions were tested and confirmed lowering of the binding free energy for product expulsion.

Cellulosic biomass represents a sustainable and renewable energy resource that can replace with the fossil fuels in the near term. Currently the conversion of cellulose to bioethanol recruits cellulose-degrading enzymes (Ragauskas et al. 2006). Yet owing to the limitations faced during commercialization of the enzymatic degradation, the conversion technology suffers from high costs (Himmel et al. 2007), the fact which emphasizes the necessity for a deeper understanding of the molecular mechanism of the enzymatic degradation of cellulose. The product expulsion step in which the degraded cellobiose or glucose units is expulsed from the catalytic tunnel of the enzyme particularly affects the efficiency of the conversion process (Gusakov et al. 1987) and thus constitutes a major obstacle for commercializing the conversion technology (Andric et al. 2010). However, the reported inhibition constants are not in agreement with each other owing to the variations in experimental conditions such as pH temperature, reinforcing the need for a reliable methodology for determination of the constants for product expulsion. Hence, the particular study of Bu et al. is critical to efforts to develop more efficient cellulose degradation (Bu et al. 2011). Essentially the suggested mutations hold great implications for protein engineering approaches to improve the efficiency of biomass conversion by cellulases. More important is that the free energy calculations have been shown to be robust and reliable alternative to challenging experimental procedures as mentioned for product expulsion.

Another important biotechnological enzyme nitrile hydratase was investigated similarly by SMD (Peplowski et al. 2008). The method particularly allowed identification of the pathways of the substrate and the product from the catalytic cleft. Different from the first two examples, this study applied constant force pulling in the SMD approach and identified a critical residue which acted as a steric barrier along both of the paths. They reported that this critical residue is involved in stereoselectivity and thus the identified residue might base protein engineering study aiming to modify stereoselectivity of the commercially important nitrile hydratase enzyme nitrile hydratase.

In most of the examples, it was shown that the FoldX predicted mutations that altered stability have been confirmed with the experiments (Christensen and Kepp

2012; Timucin and Sezerman 2013; Szczepek et al. 2007) suggesting that the method is applicable for identification of mutations that would affect the stability of proteins. Although the method presents a fast and also reliable assessment of stability in in silico mutagenesis studies, it has yet not been applicable to cases where temperature effect is needed to be investigated (Guerois et al. 2002).

8.4 Summary and Outlook

A decision tree is used to illustrate the possible routes for an engineering approach (Fig. 8.6). When 3D structure of the protein of interest is not available, structure prediction algorithms such as homology modeling can be applied. Stabilizing mutations in the predicted structure could be designed using special tools such as FoldX or a combination of prediction servers which are more reliable than a single predictor. If the desired structure is a complex molecule, along with the predicted protein structure, one also needs to identify the correct binding mode of the ligand. Docking could be used to predict such complex structures which can be further refined by the incorporation of the receptor flexibility by the use of molecular dynamics simulations. Application of free energy calculations to the complex structures might reveal the equilibrium binding constants which would be very beneficial for understanding the activity and the selectivity of enzymes. Moreover, by the virtue of path-dependent free energy calculation methods, one can predict

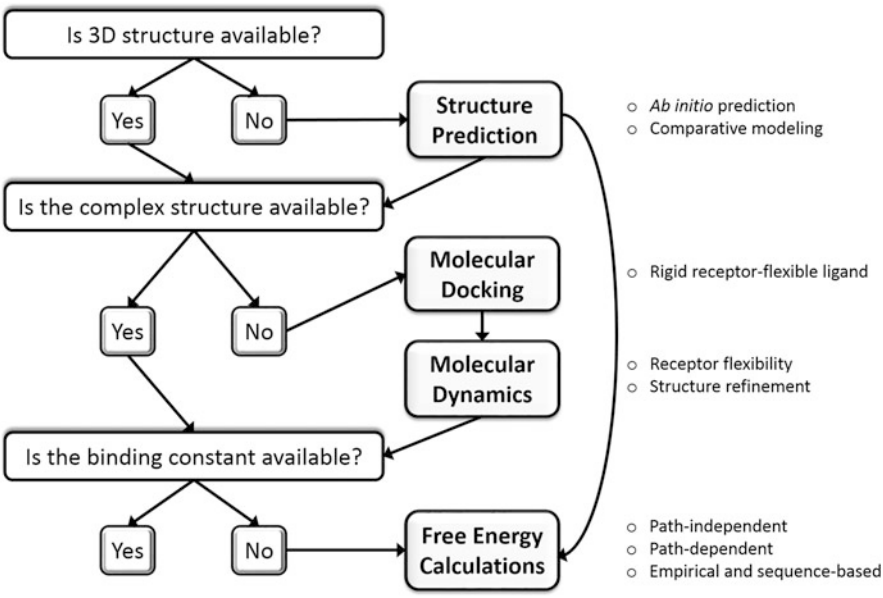


Fig. 8.6 Decision tree for selecting the correct modeling tools

mutations that can alter the equilibrium constants and thus design novel protein/enzyme variants with optimal stabilities, activities, and selectivity.

Owing to the fact that the prediction strength of molecular modeling approaches highly rely on the computational sources, the described methods here would greatly benefit from such efforts ameliorating computer technologies and parallel computing algorithms. By the recent developments in the graphics processing unit accelerated software, researchers in the University of California, San Diego, have performed simulations of biological events on the millisecond time scale. These innovations would obviously produce better outcomes for rational design studies employing molecular modeling approaches to enhance efficiency of industrial biocatalysis and to provide new routes for synthetic biology.

References

Alder BJ, Wainwright TE (1957) Phase transition for a hard sphere system. J Chem Phys 27 (5):1208–1209. doi:10.1063/1.1743957

Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. J Chem Phys 31(2):459–466. doi:10.1063/1.1730376

Andric P, Meyer AS, Jensen PA, Dam-Johansen K (2010) Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis: I. Significance and mechanism of cellobiose and glucose inhibition on cellulolytic enzymes. Biotechnol Adv 28(3):308–324. doi:10.1016/j.biotechadv.2010.01.003

Andrusier N, Mashiach E, Nussinov R, Wolfson HJ (2008) Principles of flexible protein-protein docking. Proteins 73(2):271–289. doi:10.1002/prot.22170

Bastug T, Kuyucak S (2006) Energetics of ion permeation, rejection, binding, and block in gramicidin A from free energy simulations. Biophys J 90(11):3941–3950. doi:10.1529/biophysj.105.074633

Bastug T, Kuyucak S (2007a) Application of Jarzynski's equality in simple versus complex systems. Chem Phys Lett 436:383–388

Bastug T, Kuyucak S (2007b) Free energy simulations of single and double ion occupancy in gramicidin A. J Chem Phys 126(10):105103. doi:10.1063/1.2710267

Bastug T, Chen PC, Patra SM, Kuyucak S (2008) Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. J Chem Phys 128(15):155104. doi:10.1063/1.2904461

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112(3):535–542

Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, Newstead S, Poolman B, Tate CG, Vogel H (2011) Overcoming barriers to membrane protein structure determination. Nat Biotechnol 29(4):335–340. doi:10.1038/nbt.1833

Bonneau R, Baker D (2001) Ab initio protein structure prediction: progress and prospects. Annu Rev Biophys Biomol Struct 30:173–189. doi:10.1146/annurev.biophys.30.1.173

Bu L, Beckham GT, Shirts MR, Nimlos MR, Adney WS, Himmel ME, Crowley MF (2011) Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. J Biol Chem 286(20):18161–18169. doi:10.1074/jbc.M110.212076

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33(Web Server issue):W306–W310. doi:10.1093/nar/gki375

Carey PR (1996) Protein engineering and design. Academic Press, San Diego, CA

Ceulemans H, Russell RB (2004) Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization. J Mol Biol 338(4):783–793. doi:10.1016/j.jmb.2004.02.066

Chandonia JM, Brenner SE (2005) Implications of structural genomics target selection strategies: Pfam5000, whole genome, and random approaches. Proteins 58(1):166–179. doi:10.1002/prot.20298

Chipot C, Pohorille A (2007) Free energy calculations: theory and applications in chemistry and biology. Springer series in chemical physics, vol 86, Study edn. Springer, New York

Christensen NJ, Kepp KP (2012) Accurate stabilities of laccase mutants predicted with a modified FoldX protocol. J Chem Inf Model 52(11):3028–3042. doi:10.1021/ci300398z

Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) Consensus scoring for ligand/protein interactions. J Mol Graph Model 20(4):281–295

Crooks G (1998) Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. J Stat Phys 90(5–6):1481–1487. doi:10.1023/A:1023208217925

de Vries SJ, Melquiond AS, Kastritis PL, Karaca E, Bordogna A, van Dijk M, Rodrigues JP, Bonvin AM (2010) Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions. Proteins 78(15):3242–3249. doi:10.1002/prot.22814

Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. Bioinformatics 25(19):2537–2543. doi:10.1093/bioinformatics/btp445

Dodson GG, Lane DP, Verma CS (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology. EMBO Rep 9(2):144–150. doi:10.1038/sj.embor.7401160

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7):1731–1737. doi:10.1021/ja026939x

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16):1999–2012. doi:10.1002/jcc.10349

Durmaz E, Kuyucak S, Sezerman UO (2013) Modifying the catalytic preference of tributyrin in Bacillus thermocatenulatus lipase through in-silico modeling of enzyme-substrate complex. Protein Eng Des Sel 26(5):325–333. doi:10.1093/protein/gzt004

Durre P (2007) Biobutanol: an attractive biofuel. Biotechnol J 2(12):1525–1534. doi:10.1002/biot.200700168

Edholm O, Ghosh I (1993) Hysteresis and statistical errors in free energy perturbation L to D amino acid conversion. Mol Simul 10(2–6):241–253. doi:10.1080/08927029308022167

Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 47(1):45–55. doi:10.1021/jm030209y

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aid Mol Des 15(5):411–428

Fischer M, Coleman RG, Fraser JS, Shoichet BK (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. Nat Chem 6(7):575–583. doi:10.1038/nchem.1954

Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R (2006) Advances in protein structure prediction and de novo protein design: a review. Chem Eng Sci 61(3):966–988. doi:10.1016/j.ces.2005.04.009

Francoijs CJ, Klomp JP, Knegtel RM (2000) Sequence annotation of nuclear receptor ligand-binding domains by automated homology modeling. Protein Eng 13(6):391–394

Ginalski K, Rychlewski L, Baker D, Grishin NV (2004) Protein structure prediction for the male-specific region of the human Y chromosome. Proc Natl Acad Sci USA 101(8):2305–2310

Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9(1):1–5. doi:10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6

Gruno M, Valjamae P, Pettersson G, Johansson G (2004) Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate. Biotechnol Bioeng 86(5):503–511. doi:10.1002/bit.10838

Guerois R, López de la Paz M (2006) Protein design: methods and applications. In: Methods in molecular biology, vol 340. Humana Press, Totowa, NJ

Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320(2):369–387. doi:10.1016/S0022-2836(02)00442-4

Gusakov AV, Sinitsyn AP, Klyosov AA (1987) Factors affecting the enzymatic hydrolysis of cellulose in batch and continuous reactors: computer simulation and experiment. Biotechnol Bioeng 29(7):906–910. doi:10.1002/bit.260290715

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47(4):409–443. doi:10.1002/prot.10115

Hardin C, Pogorelov TV, Luthey-Schulten Z (2002) Ab initio protein structure prediction. Curr Opin Struct Biol 12(2):176–181

Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. Nature 450 (7172):964–972. doi:10.1038/nature06522

Hilbert M, Böhm G, Jaenicke R (1993) Structural relationships of homologous proteins as a fundamental principle in homology modeling. Proteins 17(2):138–151. doi:10.1002/prot.340170204

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315(5813):804–807. doi:10.1126/science.1137016

Holm L, Sander C (1992) Evaluation of protein models by atomic solvation preference. J Mol Biol 225(1):93–105

Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381 (6580):272. doi:10.1038/381272a0

Hoppe A, Theimer RR (1996) Titrimetric test for lipase activity using stabilized triolein emulsions. Phytochemistry 42(4):973–978. doi:10.1016/0031-9422(96)00051-9

Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K (2001) Steered molecular dynamics investigations of protein function. J Mol Graph Model 19(1):13–25. doi:10.1016/S1093-3263(00)00133-9

Jarzynski C (1997) Nonequilibrium equality for free energy differences. Phys Rev Lett 78:2690–2693

Jones DT (2000) Protein structure prediction in the postgenomic era. Curr Opin Struct Biol 10 (3):371–379

Jorgensen WL, Madura JD (1983) Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. J Am Chem Soc 105(6):1407–1413. doi:10.1021/ja00344a001

Ju P, Pages G, Riek RP, Chen PC, Torres AM, Bansal PS, Kuyucak S, Kuchel PW, Vandenberg JI (2009) The pore domain outer helix contributes to both activation and inactivation of the HERG K+ channel. J Biol Chem 284(2):1000–1008. doi:10.1074/jbc.M806400200

Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652. doi:10.1038/nsb0902-646

Karplus M, Petsko GA (1990) Molecular dynamics simulations in biology. Nature 347 (6294):631–639. doi:10.1038/347631a0

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web Server issue):W526–W531. doi:10.1093/nar/gkh468

Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3(5):300–313. doi:10.1063/1.1749657

Kirkwood JG, Alder BJ (1968) Theory of liquids. His collected works. Gordon and Breach, New York

Kua J, Zhang Y, McCammon JA (2002) Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. J Am Chem Soc 124(28):8260–8267. doi:10.1021/ja020429l

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288

Lazakidou AA (2010) Biocomputation and biomedical informatics: case studies and applications. Medical Information Science Reference, Hershey

Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Prentice Hall, Harlow

Lemer CM, Rooman MJ, Wodak SJ (1995) Protein structure prediction by threading methods: evaluation of current techniques. Proteins 23(3):337–355. doi:10.1002/prot.340230308

Lv M, Ma S, Tian Y, Zhang X, Lv W, Zhai H (2015) Computational studies on the binding mechanism between triazolone inhibitors and Chk1 by molecular docking and molecular dynamics. Mol Biosyst 11(1):275–286. doi:10.1039/c4mb00449c

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102 (18):3586–3616. doi:10.1021/jp973084f

Malcata FX (1996) Engineering of/with lipases. NATO ASI series. Series E, Applied sciences, vol 317. Kluwer Academic, Dordrecht

McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267 (5612):585–590

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 19(14):1639–1662

Muegge I, Rarey M (2001) Small molecule docking and scoring, Reviews in computational chemistry. Wiley, New York, pp 1–60. doi:10.1002/0471224413.ch1

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25(13):1656–1676. doi:10.1002/jcc.20090

Pappalardo M, Milardi D, Grasso D, La Rosa C (2007) Steered molecular dynamics studies reveal different unfolding pathways of prions from mammalian and non-mammalian species. New J Chem 31(6):901–905. doi:10.1039/B700764G

Park SJ, Cochran JR (2010) Protein engineering and design. CRC, Boca Raton, FL

Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys 119:3559–3566

Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, Cavalli A (2014) Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. J Chem Inf Model 54(2):470–480. doi:10.1021/ci4003574

Peitsch MC (1996) ProMod and Swiss-Model: internet-based tools for automated comparative protein modelling. Biochem Soc Trans 24(1):274–279

Peplowski L, Kubiak K, Nowak W (2008) Mechanical aspects of nitrile hydratase enzymatic activity. Steered molecular dynamics simulations of Pseudonocardia thermophila JCM 3095. Chem Phys Lett 467(1–3):144–149. doi:10.1016/j.cplett.2008.10.072

Perakyla M (2009) Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. Eur Biophys J 38(2):185–198. doi:10.1007/s00249-008-0369-x

Pesce A, Nardini M, Dewilde S, Capece L, Marti MA, Congia S, Salter MD, Blouin GC, Estrin DA, Ascenzi P, Moens L, Bolognesi M, Olson JS (2011) Ligand migration in the apolar tunnel of Cerebratulus lacteus mini-hemoglobin. J Biol Chem 286(7):5347–5358. doi:10.1074/jbc.M110.169045

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26 (16):1781–1802. doi:10.1002/jcc.20289

Pongprayoon P, Beckstein O, Wee CL, Sansom MS (2009) Simulations of anion transport through OprP reveal the molecular basis for high affinity and selectivity for phosphate. Proc Natl Acad Sci USA 106(51):21614–21618. doi:10.1073/pnas.0907315106

Price DJ, Brooks CL 3rd (2002) Modern protein force fields behave comparably in molecular dynamics simulations. J Comput Chem 23(11):1045–1057. doi:10.1002/jcc.10083

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311(5760):484–489. doi:10.1126/science.1114736

Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261(3):470–489. doi:10.1006/jmbi.1996.0477

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815. doi:10.1006/jmbi.1993.1626

Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Sali A (2000) Protein structure modeling for structural genomics. Nat Struct Biol 7(Suppl):986–990. doi:10.1038/80776

Schlick T (2010) Molecular modeling and simulation: an interdisciplinary guide. Interdisciplinary applied mathematics, vol 21, 2nd edn. Springer, New York

Schulz-Gasch T, Stahl M (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. J Mol Model 9(1):47–57. doi:10.1007/s00894-002-0112-y

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. Nucleic Acids Res 33(Web Server issue):W382–W388. doi:10.1093/nar/gki387

Shaw DE (2013) Millisecond-long molecular dynamics simulations of proteins on a special-purpose machine. Biophys J 104(2, Suppl 1):45a. doi:10.1016/j.bpj.2012.11.289

Shetty K (2006) Food biotechnology, 2nd edn. CRC Press, Taylor & Francis, New York

Solis FJW (1981) Minimization by random search techniques. Math Oper Res 6(1):19–30. doi:10.1287/moor.6.1.19

Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. Proteins 65(1):15–26. doi:10.1002/prot.21082

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25(7):786–793. doi:10.1038/nbt1317

Takeda-Shitaka M, Takaya D, Chiba C, Tanaka H, Umeyama H (2004) Protein structure prediction in structure based drug design. Curr Med Chem 11(5):551–558

Teichmann SA, Chothia C, Gerstein M (1999) Advances in structural genomics. Curr Opin Struct Biol 9(3):390–399. doi:10.1016/S0959-440X(99)80053-0

Teodoro ML, Kavraki LE (2003) Conformational flexibility models for the receptor in structure based drug design. Curr Pharm Des 9(20):1635–1648

Timucin E, Sezerman OU (2013) The conserved lid tryptophan, W211, potentiates thermostability and thermoactivity in bacterial thermoalkalophilic lipases. PLoS One 8(12):e85186. doi:10.1371/journal.pone.0085186, PONE-D-13-32526 [pii]

Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 18(2):178–184. doi:10.1016/j.sbi.2008.01.004

Verger R, Haas GHD (1976) Interfacial enzyme kinetics of lipolysis. Annu Rev Biophys Bioeng 5:77–117

Vitkup D, Melamud E, Moult J, Sander C (2001) Completeness in structural genomics. Nat Struct Biol 8(6):559–566. doi:10.1038/88640

Vyas VK, Ukawala RD, Ghate M, Chintha C (2012) Homology modeling a fast tool for drug discovery: current perspectives. Indian J Pharm Sci 74(1):1–17. doi:10.4103/0250-474X.102537

Wang W, Kollman PA (2001) Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. Proc Natl Acad Sci USA 98(26):14937–14942. doi:10.1073/pnas.251265598

Wang C, Bradley P, Baker D (2007) Protein-protein docking with backbone flexibility. J Mol Biol 373(2):503–519. doi:10.1016/j.jmb.2007.07.050

Werner T, Sander K, Tanrikulu Y, Kottke T, Proschak E, Stark H, Schneider G (2010) In silico characterization of ligand binding modes in the human histamine H4 receptor and their impact on receptor activation. Chembiochem 11(13):1850–1855. doi:10.1002/cbic.201000180

Wood RH, Muhlbauer WCF, Thompson PT (1991) Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis. J Phys Chem 95(17):6670–6675. doi:10.1021/j100170a054

Worth CL, Preissner R, Blundell TL (2011) SDM--a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39(Web Server issue):W215–W222. doi:10.1093/nar/gkr363

Yang LJ, Zou J, Xie HZ, Li LL, Wei YQ, Yang SY (2009) Steered molecular dynamics simulations reveal the likelier dissociation pathway of imatinib from its targeting kinases c-Kit and Abl. PLoS One 4(12):e8470. doi:10.1371/journal.pone.0008470

Zhang Y (2007) Template-based modeling and free modeling by I-TASSER in CASP7. Proteins 69(Suppl 8):108–117. doi:10.1002/prot.21702

Synthetic Biopolymers

9

Christian P. Kubicek

Contents

9.1 Introduction 308
9.2 Biomedical Areas of Application of Synthetic Biopolymers 309
9.2.1 Medical Devices 309
9.2.2 Tissue Engineering 310
9.2.3 Drug Delivery and Release Control 311
9.2.4 Gene Delivery 312
9.2.5 Optical Tracing and Bioimaging 312
9.3 Synthetic Biopolymers 312
9.3.1 Aliphatic Polyesters 312
9.3.2 Poly(trimethylene carbonate) 319
9.3.3 Poly(ester amides) 320
9.3.4 Poly(orthoesters) 321
9.3.5 Polyanhydrides 322
9.3.6 Other Polymers 324
9.3.7 Polyphosphoester 326
9.3.8 Poly(amino acids) 327
9.4 Conclusions 329
References 329

Abstract

Matters, surfaces, or constructs that interact with biological systems and evaluate, treat, augment, or replace any part of the body are termed biomaterials. They can be produced from a variety of chemical compounds, yet biopolymers have always been the main and most favored type of biomaterials. The main types of biopolymers are biosynthetically derived polysaccharides and proteins and

C.P. Kubicek (✉)
Institute of Chemical Engineering, Research Area Biotechnology and Microbiology, Vienna University of Technology, Wien, Austria
e-mail: ckubicek@mail.zserv.tuwien.ac.at

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_9

aliphatic polyesters and polyphosphoester that are obtained by chemical synthesis. The latter have become attractive alternatives to biosynthetic biopolymers for biomedical applications because they have less risk to trigger immune responses in the body and can more easily be modified. This review summarizes the main polymers applied in this area, outlines their routes of synthesis, and illustrates the advances in their biomedical applications.

9.1 Introduction

Biomaterials are defined as a matters, surfaces, or constructs that interact with biological systems and evaluate, treat, augment, or replace any tissue, organ, or function of the body (Williams 2007). The study of biomaterials has a 50-year-long history and plays an important role in human health.

Biomaterials can be produced from a variety of chemical compounds, yet biopolymers have always been the main and most favored type of biomaterials. According to their degradation properties, biopolymers can be further classified into biodegradable and nonbiodegradable biopolymers. In the beginning, biopolymers were mainly developed for purposes—such as bone substitution or fixing materials and dental materials—that required a long-term stability of the material in the body. In recent years, however, the developments in tissue engineering, controlled drug delivery, and also nonmedical applications have raised the need of biomaterials with predictable and tunable degradation properties and rates.

Biopolymers can be derived either by biosynthesis (which implies synthesis by a living cell, mostly a microorganism) or chemical synthesis. Polysaccharides and proteins are typical biosynthetically derived biopolymers, while aliphatic polyesters and polyphosphoester (PPE) are typical synthetic biopolymers. The latter have become attractive alternatives to biosynthetic biopolymers for biomedical applications for the following reasons: they are less risky in triggering an immune response in the human body; chemical modifications, frequently needed to tune the properties according to the application in mind, are more easy to perform than with biopolymers; and a variety of properties can be obtained with properly designed synthetic biopolymers and further modifications are possible without altering the bulk properties.

The molecular biology revolution of the 1970s and advances in genomics and proteomics in the 1990s and 2000s further hand an enormous impact on the design and use of biomaterials. Specific molecules that have been identified to be important for some clinical processes can now be incorporated into materials as bioactive components (Ratner and Bryant 2004). Such combination products—such as the drug-eluting vascular stent INFUSE bone graft device (from Medtronic)—represent today the state of the art of commercial products and represent a billion-dollar business (Huebsch and Mooney 2009).

I will here review the current state of production and application of synthetic biopolymers. To this end, I will first briefly introduce into the areas where these polymers are to be applied and highlight the requirements that arise from them.

9.2 Biomedical Areas of Application of Synthetic Biopolymers

9.2.1 Medical Devices

Synthetic biodegradable polymers have attracted considerable attention for the design and function of medical devices. The general requirements of materials used for medical devices are suitable mechanical properties and an appropriate degradation time that allows fulfilling its task but guarantees its disappearance thereafter. Of course the materials should also not cause toxic effects or elicit an immune response, and the body should be able to fully metabolize the degradation products.

9.2.1.1 Drug-Eluting Stents (DES)

The introduction of percutaneous coronary intervention in the late 1970s revolutionized the management of coronary artery disease, providing an effective, quick, safe, and increasingly widely available method for coronary revascularization for many patients (Sheth and Giugliano 1995). The rapid development in this field led to the introduction of a number of new technologies, including intracoronary stents that have proven to provide improved efficacy and long-term safety. The two major available materials for stent manufacture are either bare-metal stents or drug-eluting stents. The latter have the advantages to achieve the lowest rates of restenosis, stent thrombosis, and recurrent myocardial infarction and are now used as default treatment (Sheth and Giugliano 1995; Kukreja et al. 2008).

9.2.1.2 Orthopedic Devices

This term summarizes various pieces of equipment that are used for preventing and treating deformities and injuries of the musculoskeletal system in man. They include bandages, splints, prostheses, and special apparatus. The use of synthetic biopolymers for their manufacture has been warranted by several requirements: the major one is the need of a device, which can be used as an implant but will not necessitate a second surgical event for removal, and thus implies biodegradability. In addition, biodegradation may offer further advantages. As an example, a fractured bone that has been fixated with a rigid, nonbiodegradable stainless steel implant has a tendency for refracture upon removal of the implant, because the bone has not been carrying a sufficient load during the healing process. In contrast, an implant prepared from a biodegradable polymer can be engineered to degrade at a rate that will slowly transfer load to the healing bone (Athanasiou et al. 1998). Furthermore, a second application with tremendous potential is the use of these materials for drug delivery. This allows the application of pharmaceutical compounds such as proteins that accelerate bone morphogenesis or antibiotics at

a slow rate and would thus speed up the healing process after a fracture or help prevent bone infections (osteomyelitis) (Middleton and Tipton 2000). The specific requirements for polymers suitable for an application as an orthopedic device are that they can be easily processed into the final product form, have an acceptable shelf life, can be easily sterilized, and have mechanical properties that will remain intact until the surrounding tissue has healed. Of course, the general requirements such as modifiable biodegradation, biocompatibility, and strength are also essential in this area.

9.2.1.3 Disposable Medical Devices

Apart from the abovenamed major areas of applications, many disposable medical devices, such as syringes, injection pipes, surgical gloves, pads, etc., are usually made of nondegradable plastics, resulting in serious environmental and economic issues. Therefore, there is an increasing interest in their manufacture from biodegradable materials.

9.2.2 Tissue Engineering

Tissue engineering is an interdisciplinary field that applies the principles of engineering and life sciences toward the development of biological substitutes used to restore, maintain, or improve tissue functions (Place et al. 2009; Vacanti 2012). In tissue engineering, cells are cultured on a scaffold to form a natural tissue, and the formed tissue is then implanted in the defect part in the patient's body. Also, a scaffold or a scaffold with cells can be implanted in vivo, and the host's body uses this implant to synthesize new tissues. The success of tissue engineering thereby largely depends on the role played by three-dimensional porous scaffolds. An ideal scaffold should be biodegradable and its degradation products metabolizable, but it also must not cause inflammation or be rejected by the immune system. Finally, it must be characterized by proper mechanical properties to support the growth of new tissues.

Examples for the successful use of synthetic biopolymers in tissue engineering include bone tissue (Rezwan et al. 2006; Ren et al. 2005), cartilage tissue (Liu and Ma 2004), cardiovascular (Venkatraman et al. 2008), arterial replacement (Isenberg et al. 2006), heart valve (Neuenschwander and Hoerstrup 2004), nerve regeneration (Bini et al. 2005; Ciardelli and Chiono 2006), and engineering of dermal substitutes for skin regeneration (Venugopal and Ramakrishna 2005).

A limitation that still limits the application of synthetic polymers in tissue engineering is that they lack molecular properties that promote cell adhesion, proliferation, and tissue recovery. A solution to overcome these deficiencies is the inclusion or copolymerization of components such as collagen or hyaluronic acid required for these processes. Thus, synthetic biodegradable polymers with which this is possible are clearly preferred (Yoo et al. 2005; He et al. 2005; Sahoo et al. 2007).

9.2.3 Drug Delivery and Release Control

The success of many pharmaceuticals depends on whether they can arrive at their destined cellular target. Various approaches, formulations, technologies, and systems for transporting a pharmaceutical compound in the body have been developed to guarantee its desired therapeutic effect. In most cases, this either involves targeting the drug to a particular place within the body or facilitating systemic pharmacokinetics. While this can in some cases also be achieved by the chemical formulation of the drug, the more common and successful way is the use of medical devices that are summarized under the term drug delivery.

Biomaterials are ideally suited for this purpose since they make both targeted delivery (i.e., the drug will only be active in the target area of the body, e.g., in the cancerous tissues) and sustained release (i.e., formulations in which the drug is released over a period of time in a controlled manner from a formulation) possible (Nair and Laurencin 2006a, b; Andreopoulos 2003). Thereby, synthetic biopolymers are mainly used as biodegradable microspheres that can be loaded with the drug and as drug polymer conjugates. Biodegradable polymers with reactive groups (carboxyl groups have been proven particularly useful) can conjugate drugs via ester or amide bonds to form a biodegradable macromolecular prodrug and thereby prevent the side effects of addition of the free drugs. During the degradation of biodegradable polymers, the drug will then be released. Similarly, targeting the drug to the desired place of action can be achieved by linking targeting ligands such as antibodies, peptides, nucleic acid aptamers (see below), carbohydrates, and small molecules to the surface of long-circulating nanoparticles, to deliver the drug-encapsulated nanoparticles to specially identified sites to minimize undesired effects (Brannon-Peppas and Blanchette 2004). Peptides with short sequences of 5–10 amino acids (such as the cRGD peptide [cyclic (Arg–Gly–Asp–D-Phe–Lys)] which targets the v3 integrin; Nasongkla et al. 2004) can be used in binding assays to target tumors.

Nucleic acid ligands such as aptamers (small oligodeoxynucleotides that specifically bind to a target) and spiegelmers (aptamers consisting of RNA molecules) represent novel classes of target agents. They can be encapsulated and so injected into the body and provide a much high ratio of tumor reduction than with other techniques (Chen et al. 2014).

Stimuli-responsive polymers can respond to small changes in environmental stimuli with distinct changes in their conformation, polarity, or chemical composition and are therefore very attractive for biomedical applications (Alarcon et al. 2005). The most convenient stimuli thereby include temperature, pH, light, and electric pulses. Various such biodegradable synthetic materials have therefore been designed and applied in drug delivery, tissue engineering, bioseparation, and biosensor designing (for review, see Tian et al. 2012). Polymers that react to electric stimuli have thereby shown to also regulate cell attachment, proliferation, and differentiation and are of major interest in this (Schmidt et al. 1997; Kotwal and Schmidt 2001).

9.2.4 Gene Delivery

A high number of human diseases are genetically encoded. Gene therapy could be a way to fix this problem at its source. By adding a corrected copy of a defective gene to the cell, gene therapy promises to help the diseased tissues and organs work properly. To introduce the gene into the cell, a vector (typically a virus) is needed to deliver the gene to the cells that require it. This can be done in a patient's body in two ways (Wang et al. 2014): in vivo, i.e., to inject the vector directly into the patient, aiming to target the affected cells, and ex vivo by delivering the gene to cells that have been removed from the body and are growing in culture. After introduction of the gene and confirmation of integration and production of the correct expression product have been confirmed, the cells are put back into the patient. The latter method is currently preferred because it will less likely to trigger an immune response.

Synthetic biopolymers have recently been proposed as safer alternatives to viral vectors for gene delivery (Cook et al. 2005; Wong et al. 2007). In this regard, an ideal gene carrier should be able to deliver the target gene to specific cells with high efficacy, be biodegradable, and be excreted from the body after a given time period. For this reason, synthetic biopolymers are increasingly used in this field.

9.2.5 Optical Tracing and Bioimaging

Optical tracing and bioimaging using fluorescent probes with low toxicity, high sensitivity, and ability to recognize molecules or proteins are another recent and important technology in medicine. Near-infrared fluorescent imaging probes are particularly useful because components of the surface tissue only weakly absorb it, and it can thus penetrate the tissue (Funovics et al. 2003).

9.3 Synthetic Biopolymers

9.3.1 Aliphatic Polyesters

Aliphatic esters comprise the earliest and also most extensively investigated class of biodegradable polymers for biomedical applications (Tian et al. 2012). This uniqueness is the result of an immense diversity and synthetic versatility. Aliphatic polyesters can be developed from a variety of monomers by rather inexpensive and straightforward synthetic strategies such as ring-opening and condensation polymerization routes depending on the monomeric units. In addition, they are biodegradable, bioresorbable, and biocompatible. Their structure also allows the incorporation of reactive groups, which not only allow the introduction of additional moieties appropriate for tuning properties such as hydrophilicity, biodegradation rates, and bioadhesion but also—as already emphasized above—attachment of components for drug targeting and delivery (Nair and Laurencin 2006b; Ulery

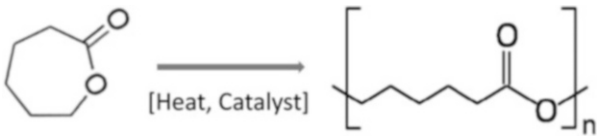


Fig. 9.1 Synthesis of polycaprolactone by ring-opening polymerization (ROP)

et al. 2011). Particularly interesting are functional groups containing carboxyl, hydroxyl, amino, ketal, bromo, chloro, and carbon–carbon double bonds or triple bonds.

Aliphatic polyesters with reactive groups can be prepared by the homopolymerization or copolymerization of cyclic monomers bearing protected functional groups (Fig. 9.1). The resulting aliphatic polyesters are thermoplastic polymers with hydrolytically labile aliphatic ester linkages in their backbone. Some of them can also be produced by fermentation with bacteria: synthesis of polyhydroxyalkanoates containing two hydroxyalkanoate monomers, lactate, glycolate, and 2-hydroxybutyrate is possible with bacteria in which the broad substrate specificities of polyhydroxyalkanoate synthase and propionyl-CoA transferase are engineered to incorporate the above monomers into the polymer chain (for review, see Matsumoto and Taguchi 2013).

Various synthetic routes for developing polyesters have been recently reviewed by Tian et al. (2012). A potential drawback of aliphatic polyesters, however, is that their degradation mainly proceeds via bulk erosion, i.e., the polymeric matrices depolymerize all over their cross section and have nonlinear and discontinuous degradation kinetics (Albertsson 2002).

9.3.1.1 Poly-α-hydroxy Acids

Among the class of aliphatic polyesters, the most extensively investigated polymers are the poly-α-hydroxy acids, which include poly(glycolic acid) and the stereoisomeric forms of poly(lactic acid). Poly-α-hydroxy esters can be synthesized by the polycondensation of bifunctional monomers such as hydroxy acids, diacids with diols, and diacid chlorides with diols or by the ester interchange reaction of diesters and diols (Nair and Laurencin 2006a, b). However, this method has a number of drawbacks, such as the necessity for high temperatures and long reaction times which can lead to racemization or other side reactions (Jérôme and Lecomte 2008). Also the molecular weight of the polymers prepared by this method is usually $<$30 kDa, because of kinetic problems of the reaction. The polycondensation route has therefore not been extensively investigated for developing biomaterials (Edlund and Albertsson 2003). Selected examples of synthesis are reviewed by Seyednejad et al. (2011).

Ring-opening polymerization (ROP) of cyclic lactones (and occasionally also cyclic anhydrides) of varying ring sizes has proven as the most effective one-step polymerization route to develop high molecular weight homo- and co-polyesters. It can be performed either in melt or solution and with or without (protected) functional groups on the lactone (Xudong et al. 2003; Williams 2007). The advantages of ROP over polycondensation are the milder reaction conditions,

shorter reaction times, and absence of by-products (Lofgren et al. 1995). Ring-opening polymerization can proceed via radical, anionic, or cationic polymerization as described below in more details (for review, see Nuyken and Pask 2013). During ROP, specific initiator molecules such as hydroxyl-containing molecules can control the molecular weight of the polymers. The rate of polymerization can be controlled by the application of a wide range of biocompatible catalytic systems, such as antimony trioxide or antimony trihalides, zinc lactate, or stannous (II) 2--ethylhexanoate. The latter is approved by the US Food and Drug Administration (FDA) as a food stabilizer and thus most commonly used. The ring-opening polymerization occurs under a nitrogen atmosphere at two subsequent temperatures (195 ˚C and 230 ˚C) (Stridsberg et al. 2002). To further improve their biocompatibility, several polymerization routes have been developed that use no or very little organic solvents (Nair and Laurencin 2006b). Biocatalytic synthesis of polyesters by bacterial lipases is yet another elegant method for developing polyesters (Piotrowska and Sobczak 2014; Hernandez et al. 2014), but beyond the scope of this review.

The most extensively studied monomers for aliphatic polyester synthesis for biomedical applications are glycolides, lactides, and caprolactones (Middleton and Tipton 2000).

Polyglycolides

Polyglycolides are the simplest aliphatic linear polyesters and were the first biodegradable synthetic polymer investigated for biomedical applications. It is highly crystalline (45–55 % crystallinity) and therefore exhibits a high tensile modulus (i.e., a measure of stiffness of an elastic material used to describe the elastic properties of objects when they are stretched or compressed; it is defined as ratio of force per unit area along an axis to ratio of deformation over initial length along that axis). It also has a very low solubility in organic solvents. The glass transition temperature of the polymer (i.e., the reversible transition from a hard and relatively brittle state into a molten or rubber-like state) ranges from 35 to 40 ˚C, and the melting point is greater than 200 ˚C. In spite of its low solubility, this polymer can be fabricated into a variety of forms and structures by extrusion, injection, and compression molding or solvent casting (Gunatillake et al. 2006).

Due to their excellent fiber-forming ability, polyglycolides were initially investigated for developing resorbable sutures. The first biodegradable synthetic suture on polyglycolide basis (manufactured as DEXON) was approved by the FDA in 1969. In addition, nonwoven polyglycolide fabrics have been extensively used as scaffolding matrices for tissue regeneration due to their excellent degradability, good initial mechanical properties, and cell viability on the matrices. A polyglycolide nonwoven fabric–fibrin glue composite matrix is currently undergoing clinical trials as a biocompatible dural substitute because of its excellent skin-closing ability without requiring sutures and its ability to help regenerate biological tissue (Terasaka et al. 2006).

Polyglycolides also show excellent mechanical properties due to their high crystallinity and higher stiffness than any other degradable polymeric system

used clinically and have been shown to exhibit a tensile modulus of approximately 12.5 GPa. Because of this, they are considered as internal bone fixation devices. However, polyglycolide homopolymers lose their strength and mass within 1–2 and 6–12 months, respectively, which are rather high degradation rates that limit some of the biomedical applications. This can be overcome by synthesis of copolymers.

The final degradation product is glycine which can be excreted in the urine or catabolized into carbon dioxide and water via the citric acid cycle.

Polylactides

L-Lactic acid is a nontoxic constituent of human metabolism, and its polymer is therefore absolutely biocompatible. It is also fully hydrolyzable, and its biodegradation yields lactic acid, which can be metabolized by the body. Finally, the monomer can be produced by fermentation with microorganisms from a variety of carbon sources including carbohydrates from renewable plant biomass (Wang et al. 2015). Polylactides have therefore received increasing interest during the last 15 years.

Unlike glycolides, however, lactic acid is a chiral molecule and exists in two optically active forms: L-lactic and D-lactic acid. The polymerization of either of these monomers leads to the formation of semicrystalline polymers, whereas polymerization of a racemic mixture results in the formation of amorphous polymers. Because the biotechnological production yields L-lactic acid, poly(L-lactide) (PLLA) is mostly produced. It has 37 % crystallinity, a glass transition temperature of 60–65 °C, and a melting temperature of approximately 175 °C (Middleton and Tipton 2000). In comparison to polyglycolides, it is a slow-degrading polymer due to its higher hydrophobicity, and total resorption of high molecular weight PLLA can take between 2 and 5.6 years in vivo (Bergsma et al. 1995). In addition, it has a good tensile strength and a high tensile modulus (approximately 4.8 GPa) and, hence, has been considered an ideal biomaterial for load-bearing applications, such as orthopedic fixation devices. PLLA can also form high-strength fibers and was FDA approved in 1971 for the development of an improved suture over DEXON (vide supra). PLLA has also been investigated as scaffolding material for developing ligament replacement or augmentation devices for the replacement of nondegradable fibers, e.g., for long-term blood vessel conduits (Lu et al. 2005; Cooper et al. 2005; Silberman et al. 2005; cf. Nair and Laurencin 2006b).

However, the long persistence of PLLA prevents its use in medical applications where faster degradation rates are needed, such as implant material. A solution for this is the synthesis of copolymers (e.g., with glycolides or DL-lactides). As an example, a 70:30 amorphous copolymer of poly(L-lactide-*co*-DL-lactide) has a significantly increased degradation rate and constitutes a bioresorbable implant material (marketed as Resomer LR708) (Leinonen et al. 2002).

In contrast to PLLA, a polymer with randomly distributed L- and D-lactide units (poly(DL-lactide), PDLLA) is amorphous and thus exhibits much lower strength than PLLA, loses its strength within 1–2 months, and undergoes a loss in mass within 12–16 months while still displaying the same glass transition temperature

(55–60 °C). These properties render it a preferred candidate for developing drug delivery vehicles and as low-strength scaffolding material for tissue regeneration.

Toward having a variety of polymers with a broad spectrum of degradation rates, poly(lactide-*co*-glycolide) (PLGA) polymers were synthesized from both PLLA and PDLLA: the resistance to hydrolytic degradation was found to be lowest at a 1:1 ratio, but became more pronounced at either end of the copolymer composition range (Miller et al. 1977; Middleton and Tipton 1998; Gunatillake et al. 2006). The rate of degradation of PLGA has been shown to depend on a variety of parameters including the lactide/glycolide ratio, molecular weight, and the shape and structure of the matrix. PLGA demonstrates good cell adhesion and proliferation, thus making it a potential candidate for tissue engineering applications.

Various studies have been performed so far using micro- and nanofabrication techniques to form three-dimensional scaffolds based on PLGA that can be used in a wide range of biomedical applications (Pan and Ding 2012; Romagnoli et al. 2013; Asti and Gioglio 2014). PuraSorb PLG, a semicrystalline bioresorbable copolymer of L-lactide and glycolide with a monomer ratio of 80L:20G (Tiainen et al. 2002), is the most popular of these materials and used in a large number of applications. In addition, a copolymer containing 90 % glycolic acid (GA) and 10 % L-lactic acid (LA) was used for the development of the multifilament suture Vicryl. Fabrication of monofilament sutures was subsequently achieved by copolymerization with other polymers such as poly(glycolide-*co*-trimethylene carbonate) and poly(glycolide-*co*-caprolactone) (see below for more details) that contain hard and soft segments along the polymer backbone and can therefore be used for modulation of the property. Other applications of PLGA are in the form of meshes (Vicryl meshes, e.g., the tissue engineering skin graft Dermagraft), suture reinforcements, skin replacement materials, and dura mater substitutes (Nair and Laurencin 2006b).

Another application of biodegradable PLGA is for use in guided tissue regeneration by providing a permeable material for space preservation (Castillo-Dali et al. 2014). To this end, composites with biologically interacting polymers such as alginate, chitosan, collagen, or gelatin have been investigated. A composite PLGA–collagen matrix membrane (marketed as CYTOPLAST Resorbs) is available for guided tissue regeneration.

PLGA has also been attempted to be used as material for controlled drug delivery vehicles (microspheres, microcapsules, nanospheres, or nanofibers). The advantage of the use of PLGA would have been that the drugs or proteins to be delivered were undergoing varying extents of interactions with the base polymer which resulted in rapid or prolonged release profiles (Ueda and Tabata 2003). However, for the same reason, the achievement of zero-order release kinetics from these polymer matrices appeared to be difficult and the acidic products produced during biodegradation were capable of denaturing the proteins. Thus, surface-eroding polymers, whose degradation would not affect the encapsulated material (see below), would be more suitable candidates for developing drug delivery vehicles.

PLGA microspheres have been shown to also be safe and promising for drug and gene delivery into the brain. Due to their size, these microspheres can be easily

implanted in discrete, precise, and functional areas of the brain without causing damage to the surrounding tissue (Li et al. 2013). Potential applications in neurodegenerative diseases by delivery of the neurotrophic factor (Gujral et al. 2013) or treatment of Parkinson's disease (Garbayo et al. 2013) have also been investigated. PLGA nanoparticles are also options for local cancer therapy and may also enhance response to combination anticancer regimens (Dinarvand et al. 2011; Sadat Tabatabaei Mirakabad et al. 2014). Among these, OncoGel—a controlled-release depot formulation of paclitaxel—has been demonstrated to physically target the drug to the tumor site with only very little entering blood circulation, resulting in an acceptable safety profile with dose-limiting toxicities. In addition, OncoGel is effective both in stand-alone treatments as well as in combination therapies, displaying a synergistic effect in the latter (Elstad and Fowers 2009).

9.3.1.2 Polycaprolactone

Polycaprolactone (PCL) is a semicrystalline polyester that is soluble in a wide range of organic solvents and has a low melting point (55–60 ˚C) and glass transition temperature of 60 ˚C. Importantly, it forms miscible blends with a wide range of polymers. PCL also displays low tensile strength (approximately 23 MPa) but an extremely high elongation at breakage (47-fold; Gunatillake et al. 2006). Its degradation rate is very low (2–3 years). It is commercially prepared by ROP of ε-caprolactone (Fig. 9.1).

The broad spectrum of compatibility with a wide range of other polymers, its versatile nature, and ease of fabrication have established PCL to be the major polymer of interest worldwide for drug delivery and tissue engineering applications. Yet to this end, several of its properties need to be corrected to fit the applications, e.g., its hydrophobic nature does not allow facile release of hydrophobic drugs and micelle formation, and the long degradation time slows down tissue replacement. As for scaffolds, PCL's mechanical property limits its application to hard tissue engineering only (Woodruff and Hutmacher 2010). PCL has therefore become a model system for testing the effect of modifications toward overcoming these undesirable properties and thus allowing successful application in pharmaceutical formulations (Dash and Konkimalla 2012). A summary of modifications and their consequences is given in Table 9.1. The copolymerization of morpholine-2,5-dione derivatives with lactide or lactones is also a convenient way to prepare aliphatic biopolymers bearing reactive groups. Veld in't et al. (1992) reported ROP of either ε-caprolactone or DL-lactide with morpholine-2,5-dione derivative functional substituents such as benzyl-protected carboxylic acid, benzyloxycarbonyl-protected amine, and *p*-methoxy-protected thiol groups. The resulting copolymers, when deprotected, yielded poly(ester amides) with the respective functional pendant groups.

Due to the slow degradation, high permeability to many drugs, and nontoxicity, PCL was initially investigated as a long-term drug/vaccine delivery vehicle, such as for the long-term zero-order release of levonorgestrel (Ma et al. 2006). Also, polysaccharide-coated PCL fiber mats remarkably combine the mechanical resistance of PCL with the surface properties of chitosan or hyaluronate. The control of

Table 9.1 PCL modifications and their consequences[a]

Polymer used	Properties altered (major)	Feature achieved				
		Crystallinity	Tensile strength	Elongation at break value	Amphiphilic	Others
Starch	Mechanical property	D	D	D		
Polyethylene glycol	Ionic and mechanical property		I	D		Thermoresponsive
Methoxypolyethylene glycol	Ionic and mechanical property				x	
Chitosan	Ionic and mechanical property				x	
Polyethylene oxide	Ionic and mechanical property				x	
Polyethylene amine	Ionic property				x	
Poly-lactic and poly-glycolic acid	Mechanical and ionic property		I	I		
Polyurethane	Mechanical property					Thermoresponsive
Acrylate	Ionic property					pH- and thermoresponsive
Hydroxyapatite	Mechanical property	I	I			
Polyvinyl alcohol/ pyrrolidone	Mechanical property	D	D			

[a]Data compiled from Dash and Konkimalla (2012). D indicates "decreased," I "increased." x specifies the achievement of this property

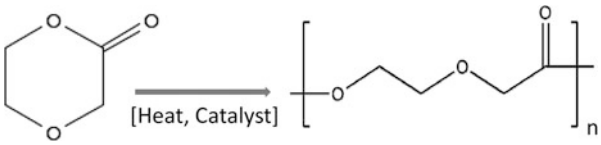


Fig. 9.2 Synthesis of polydioxanone by ROP

the nanofiber structure can thereby be controlled by the electrospinning technology, thus precisely designing biomaterials for tissue engineering, e.g., in wound healing applications (Croisier et al. 2014). PCL has also been proposed as a material for root canal filling in dentistry (as a composite named Resilon). It thereby can replace gutta-percha, has similar handling properties, but is not biodegradable (Lotfi et al. 2013). A comprehensive review of the various approaches to modify the properties of PCL has been presented by Dash and Konkimalla (2012).

9.3.1.3 Polydioxanone

Although biodegradable polylactides and glycolides have allowed for the development of versatile resorbable multifilament sutures for biomedical applications, materials were also looked for that would facilitate the formation of monofilament sutures. Multifilament sutures have a higher risk of infection associated with their use and cause a greater amount of friction when penetrating tissues. Polydioxanone (PDS) is one of the few materials from which currently available antimicrobial sutures (PDS [polydioxanone] Plus) have been manufactured (Anonymous 2014). In addition to sutures, PDS has also been investigated for several orthopedic applications as fixation screws for small bone and osteochondral fragments (Orthosorb absorbable pins) (Prior et al. 1997).

PDS is a colorless, semicrystalline polymer prepared by the ROP of *p*-dioxanone (Fig. 9.2). The polymer exhibits a very low glass transition temperature ranging from −10 to 0 °C. Due to its high crystallinity and hydrophobicity, it is degraded only at a moderate rate (6–12 months), and the resulting glyoxylate is excreted in the urine or converted into glycine and subsequently catabolized to carbon dioxide and water. The tensile strength is very low.

9.3.2 Poly(trimethylene carbonate)

Aliphatic polycarbonates are particularly good materials to synthesize materials with appropriate properties because they possess side chains (hydroxyl, amino, and carboxyl groups and others) that can easily meet to be used for functionalization of biomaterials (Tian et al. 2012). Moreover, aliphatic polycarbonates have good biocompatibility, low toxicity, and good biodegradability (Cameron and Shaver 2011).

Trimethylene carbonate (=1,3-propylene carbonate) is a 6-membered cyclic carbonate ester. It is a colorless solid that upon heating converts to the polytrimethyl carbonate. This compound may be prepared from 1,3-propanediol

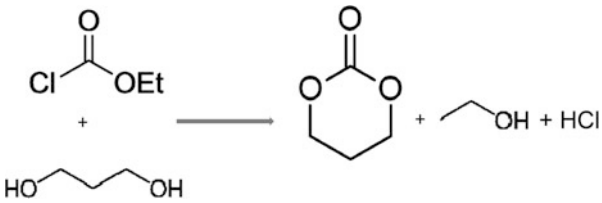


Fig. 9.3 Synthesis of trimethylene carbonate from ethyl chloroformate (*top*) and 1,3-dibutanediol (*bottom*)

and ethyl chloroformate or oxetane and carbon dioxide with an appropriate catalyst (Pyo et al. 2012; Fig. 9.3). High molecular weight aliphatic polycarbonates can be prepared by the ROP of trimethylene carbonate (Chen et al. 1997). Six-membered cyclic carbonates are also easily polymerized and copolymerized with various heterocyclic monomers to form polycarbonates without any decarboxylation under proper conditions (Tian et al. 2012).

PTC is an elastomeric polyester with excellent flexibility but poor mechanical strength and thus a potential attractive material for soft tissue regeneration. In addition, low molecular weight PTCs are tested for use in drug delivery vehicles. Unlike the previously described polyesters, PTC undergoes surface degradation.

Its low mechanical stability has been increased by developing copolymers with glycolides that are used as flexible suture materials (marketed as Maxon) and orthopedic tacks and screws (marketed as Acufex). In addition, a tripolymer consisting of Biosyn, a terpolymer composed of PTC, glycolide, and dioxane, has been prepared that has reduced stiffness and is also used as suture material.

9.3.3 Poly(ester amides)

Poly(ester amide)s are polymers with ester and amide groups on their chemical structure which allow excellent biodegradation and better mechanical and thermal endurance than the other aliphatic polyesters because of the strong hydrogen bonds between amide groups.

Poly(ester amide)s can be prepared from different monomers such as ε-caprolactam, adipic acid, and 1,4-butanediol (Fig. 9.4). Using radiolabeled poly(ester amides) in rats showed that they become degraded by hydrolysis of their ester linkages, while their amide blocks remained stable (Shi et al. 2009). To overcome this limited biodegradability, amino acid units have been incorporated into the backbone of poly(ester amide)s (Katsarava et al. 1993, 1999; Arabuli et al. 1994). The resulting polymers not only display good biodegradability but also show excellent thermal stability, tensile strength, and elastic modulus. CAMEO is a poly(ester amide) blend based on leucine or phenylalanine that has been developed for the site-specific delivery of small hydrophobic drugs and peptides (Dhandayuthapani et al. 2011).

Another class of elastomeric poly(ester amide)s constitute of cross-linked networks based on the amino alcohol 1,3-diamino-2-hydroxypropane (Bettinger et al. 2008). They feature a tensile modulus of around 1 MPa and reversible

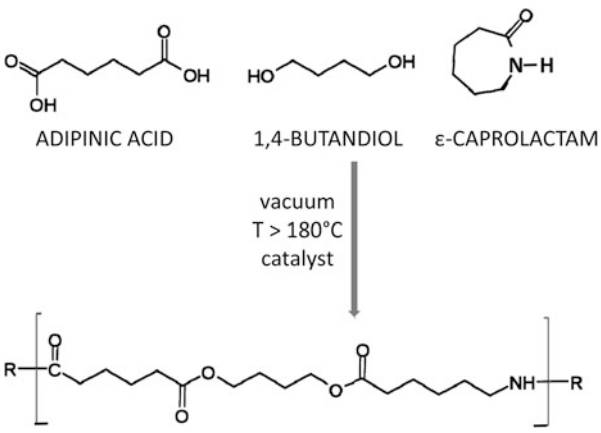


Fig. 9.4 Synthesis of random copoly(ester amide) based on adipic acid, ε-caprolactam, and 1,4-butanediol

elongations up to 92 % and are in vitro and in vivo biocompatible. They display half-lives up to 20 months in vivo.

9.3.4 Poly(orthoesters)

The polymers described so far all undergo bulk erosion, i.e., degradation occurs throughout the whole material equally and both the surface and the inside of the material. However, drug delivery requires a constant or controllable release of the drug. This is only possible with surface erosion if a very thin material is used or if surface area is kept constant. In addition, surface erosion is useful for protecting water-soluble drugs until the time of desired drug release (Lyu and Untereker 2009). This has consequently led to the search for more hydrophobic polymers with hydrolytically sensitive backbones that would favor surface erosion.

Over the last 30 years, poly(orthoesters) have evolved through four families, designated as POE I, POE II, POE III, and POE IV (Heller et al. 2002). Although the orthoester linkages are hydrolytically labile, the polymer is hydrophobic enough such that its erosion in aqueous environments is very slow. Figure 9.5 shows the structures of different types of poly(orthoesters). Poly(orthoester) I (POE I) is synthesized by the transesterification between a diol and diethoxytetrahydrofuran. One of its hydrolysis products, γ-hydroxybutyric acid, autocatalyzes further degradation of the polymer. Therefore, poly(orthoester) II (POE II) was synthesized to overcome the autocatalytic effect of POE I. Poly(orthoester) III (POE III) is synthesized by the direct polymerization of a triol with an orthoester. In this case, the polymer chains are highly flexible making the polymer a gel-like material at room temperature. The viscous nature allows for the incorporation of therapeutic agents into the polymer matrix without the need for solvents. However, the gel-like consistency and the technical difficulties in scaling up the synthetic procedure limited the production of poly(orthoester) III. Poly(orthoester) IV was therefore developed as a modification of poly(orthoester) II to allow for an appreciable

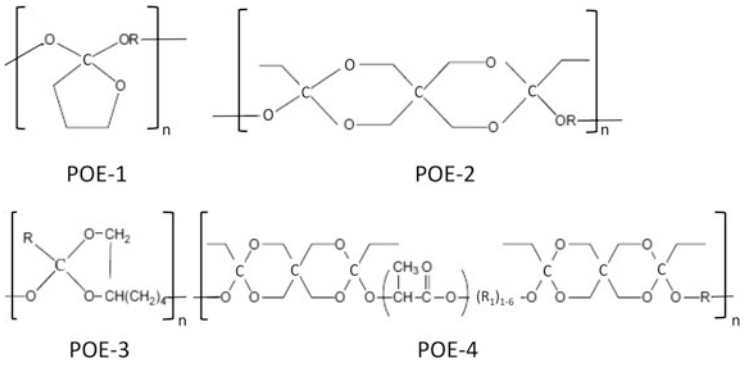


Fig. 9.5 Structure of different types of poly orthoesters

degradation rate without the addition of an acid excipient. This was achieved by incorporating short segments based on lactic or glycolic acid into the polymer backbone. Upon exposure to aqueous environments, the latent acid will undergo hydrolysis, and the liberated lactic or glycolic acid will catalyze the further hydrolysis of the polymer. Consequently, the pH in the interior of the matrix will remain neutral during erosion. Due to the versatility of its synthesis that allows the production of polymers with appropriate mechanical and thermal properties and drug release rates varying from days to months, POE IV has been considered to be the biomaterial with greatest potential (Heller et al. 2002).

Two physical forms of such polymers are under development. One form, solid materials, can be fabricated into shapes such as wafers, strands, or microspheres. The other form is injectable semisolid materials that allow drug incorporation by a simple mixing at room temperature and without the use of solvents. Important applications under development are treatment of postsurgical pain, osteoarthritis, and ophthalmic diseases as well as the delivery of proteins, including DNA. Block copolymers of poly(orthoester) and polyethylene glycol have been prepared and tested as a matrix for drug delivery and as micelles, primarily for tumor targeting (Nair and Laurencin 2006b).

9.3.5 Polyanhydrides

Another group of biopolymers with surface-eroding properties that have been extensively investigated for drug delivery applications are polyanhydrides. Since their aliphatic anhydride bond is highly sensitive to hydrolysis, they are one of the most hydrolytically labile polymers, yet the hydrophobicity of the polyanhydrides prevents penetration of water into the matrix thus rendering them only accessible to surface erosion. It was originally developed in 1932 as a fiber-forming polymer for textile applications (Palmer 2001). However, in view of the hydrolytic instability and surface-eroding nature, as described above, this class of polymers was tested as

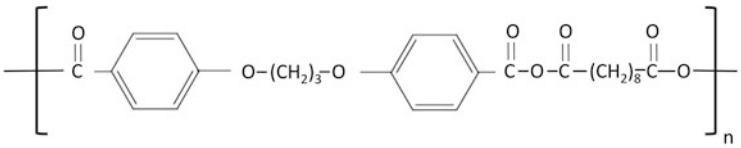


Fig. 9.6 Structure of poly[(carboxyphenoxy propane)-sebacic acid]

candidate materials for controlled drug delivery applications already in the 1980s (Leong et al. 1985). Polyanhydrides with particularly slow biodegradation rates can be prepared with aliphatic linear fatty acid dimers as polymerization partners (Kumar et al. 2002). In 1996, it was approved by the US FDA, as a drug delivery vehicle following extensive in vitro and in vivo drug release and biocompatibility evaluations (Nair and Laurencin 2006b).

Polyanhydrides are synthesized via melt condensation, a polymerization that involves the reaction of dicarboxylic acid monomers with excess acetic anhydride at a high temperature and under a vacuum. Alternatively, ROP of anhydrides, interfacial condensation (i.e., when a water-soluble monomer and an organic-soluble monomer are brought together at the interface between water and a water-immiscible organic solvent), or reaction of diacyl chlorides with coupling agents such as phosgene or diphosgene has been described (Tian et al. 2012).

The most extensively investigated polyanhydride is 1,3-bis (*p*-carboxyphenoxy propane sebacic acid (PCPP-SA; Fig. 9.6), which was approved as the chemotherapeutic agent bis-chloroethylnitrosourea (BCNU) for treatment of brain cancer. To this end, it is made into disks (marketed as Gliadel), which are infused with BCNU, and implanted under the skull of the patient (Nagpal 2012). In a similar way, a copolymer of sebacic acid and an erucic acid dimer has been found to serve as a potential delivery vehicle for gentamicin (marketed as Septacin) in the treatment of osteomyelitis and skin abscess (Li et al. 2002).

Polyanhydrides have also been investigated as an attractive material for orthopedic applications which involve controlled drug delivery. As an example, treatment of bone fractures through fixation requires the use of materials with sufficient strength to aid the patient by fixation with a tissue-compatible material and also facilitate molding into a desired shape for ensuring regional mechanical properties. While metallic implants are widely used in many treatments, they have a number of disadvantages such as stress shielding during healing and chronic inflammation. These consequences can be avoided when degradable polymeric materials are used which can simultaneously deliver therapeutic drugs to treat infections or provide growth factors to accelerate the growth of the new bone. Yet the limited mechanical strength of polyanhydrides poses a potential problem in this application. To overcome this bottleneck, several solutions have been offered: one is the use of poly (anhydride-*co*-imides), particularly such based on succinic acid trimellitylimidoglycine and trimellitylimidoalanine, which display an unusual strength due to the imide segments in the polymer backbone (Uhrich et al. 1995; Attawia et al. 1996). Alternatively, the incorporation of acrylic groups in the monomeric

units to form polyanhydrides that could be cross-linked by photoactivation has been used (Muggli et al. 1998; Burkoth and Anseth 2000). They would be injectable and could thus be used for filling irregularly shaped bone defects or for soft tissue repairs that require materials with a liquid or putty-like consistency, which can set and be molded into a desired shape under physiological conditions. Different types of initiator–accelerator systems and energy sources have been investigated to develop cross-linkable matrices with appropriate thickness for this purpose. The most effective composition for the photopolymerization of these polymers was found to be 1.0 % (w/v) camphorquinone (CQ) and 1.0 % (w/v) ethyl-4-*N*,*N*-dimethyl aminobenzoate (EDMAB) with 150 mW/cm^2 of blue light. In addition, the inclusion of tricalcium phosphate or calcium sulfate created high-strength matrices (Gunatillake and Adhikari 2003).

9.3.6 Other Polymers

9.3.6.1 Poly(alkyl cyanoacrylates)

Poly(alkyl cyanoacrylates) (PCA; Fig. 9.7) represent the major class of biodegradable acrylate polymers that are used in biomedical applications. The monomer, cyanoacrylate, was discovered in 1942 as an unwanted by-product of the synthesis of clear plastic gun sights during World War II, which stuck to everything that it came in contact with. It was later rediscovered by Eastman Kodak researchers and sold to Loctite in the 1960s (Butterworth 1988). The observation that cyanoacrylate rapidly polymerizes in the presence of moisture led to the discovery of PCAs, which are consequently prepared by the anionic polymerization of alkyl cyanoacrylic monomers with a trace amount of moisture as the initiator. They proved as excellent synthetic surgical glue, skin adhesive, and an embolic material (ASGE Technology Committee et al. 2013).

The neighboring CN and keto groups withdraw electrons from the methylene hydrogen atoms, thus rendering the backbone of this polymer sensitive to hydrolysis. The degradation is therefore very fast (between a few hours and days), depending on the length of the alkyl side groups.

Dermabond (2-octyl cyanoacrylate) has been approved by the US FDA as a tissue adhesive for topical skin application, although allergic reactions have been observed in some cases (Ricci et al. 2014). Poly(alkyl cyanoacrylates) have also been extensively investigated for use as gene delivery vehicles and are considered to be unique matrices for the delivery of oligodeoxynucleotides because of the hydrophobic interactions of the matrix with the oligonucleotides and have been

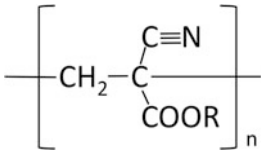


Fig. 9.7 Structure of poly (alkyl cyanoacrylate)

employed for the preparation of nanoparticles as drug delivery vehicles (Zhang et al. 2007; Duan et al. 2009; Chung et al. 2013). It has several advantages over other polymeric nanoparticles such as easy preparation, high utility size ranges, absence of solvent residues, ability to form stealth nanoparticles, and the ability of PCA to absorb or encapsulate a wide range of drug or protein molecules, particularly for cancer therapy (Lai et al. 2014).

9.3.6.2 Polyphosphazenes

In addition to organic polymers, also inorganic or inorganic–organic hybrid polymers have been investigated for their suitability as biodegradable biomaterials. Polyphosphazenes are hybrid polymers with a backbone of alternating phosphorus and nitrogen atoms containing two hydrolytically stable organic side groups attached to each phosphorus atom, which provides the molecule with unusual flexibility (Fig. 9.8). To render the polymer backbone susceptible to hydrolysis, amino acid esters with glucose, glycerol, glycolate, lactate, or imidazole are incorporated, which creates degradation profiles ranging from few hours to years (Baillargeon and Mequanint 2014). Among the amino acid esters investigated, polyphosphazenes substituted with glycine ethyl ester show the fastest degradation (Nair and Laurencin 2006b).

The common route for polyphosphazene synthesis involves two steps: first, ROP of the cyclic trimer hexachlorocyclotriphosphazene, which yields the reactive macromolecular intermediate poly(dichlorophosphazene). Subsequently, the chlorine side units are replaced by one of a broad range of organic side groups (Allcock 2003). Alternatively, they can be produced by a living cationic polymerization process (Aoshima and Kanaoka 2009), which involves condensation of the monomer, trichloro(trimethylsilyl)phosphoranimine under PCl_5-catalyzed release of trimethylsilylchloride. The advantage of this synthetic route is that it can be performed at room temperature and yields polymers with controllable chain lengths and narrow molecular weight distribution (Allcock 2003).

The in vitro and in vivo biocompatibility of biodegradable polyphosphazenes have been extensively investigated by Laurencin and coworkers (Nair et al. 2004, 2006a, b; Laurencin and Nair 2004; Kumbar et al. 2006). They assessed in a rat model that most of the amino acid esters of polyphosphazenes elicited only a minimal to mild response by the tissue when subcutaneously implanted, thus recommending them as excellent materials for drug delivery. They are excellently compatible with bone tissue and would thus make good matrices for bone tissue

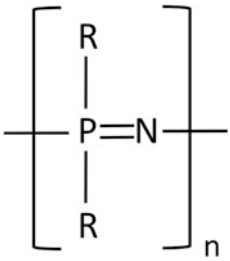


Fig. 9.8 General structure of polyphosphazene

engineering (Nair et al. 2006a, b), yet their soft elastomeric properties limit this use for load-bearing applications.

9.3.7 Polyphosphoester

Polyphosphoesters represent a wide range of biodegradable polymers with repeating phosphoester linkages in the backbone, where the phosphorus atom is pentavalent and it thereby allows introduction of bioactive molecules and extensive modification of the physical and chemical properties of the polymers (Wang et al. 2009). Depending on the nature of the side group connected to the phosphorous atom, polyphosphate, polyphosphonate, polyphosphite, or polyphosphoramidate polymers can be distinguished. In the frame of this review, I will only focus on polyphosphates with P–O side linkages. These polyphosphoesters have been investigated as biomaterials for about two decades in drug/gene delivery and tissue engineering. Polyphosphoesters display good in vitro cytocompatibility and good in vivo tissue compatibility and similarity to biomacromolecules, such as nucleic acid, which makes them attractive for special applications.

Figure 9.9 shows the general structure of polyphosphoester, where R and R´ can be varied to develop polymers with a wide range of physicochemical properties, including biodegradability. Polyphosphate esters can be synthesized by polycondensation, polyaddition, transesterification, and ROP, and this has been extensively reviewed by Wang et al. (2009). The display excellent in vitro cytocompatibility and good in vivo tissue compatibility.

Copolymers with other monomers, such as DL-lactide—thus forming poly (lactide-*co*-ethyl phosphate)—offer further special advantages in drug delivery applications because they display near zero-order release profile of chemotherapeutic drugs. One prominent example is the delivery of paclitaxel, a medication used to treat ovarian, breast, lung, pancreatic, and other cancers (Viúdez et al. 2014), which is marketed as PACLIMER microspheres (Harper et al. 1999; Lapidus et al. 2004).

Water-soluble cationic polyphosphoesters such as polyphosphoramidates have been investigated as gene carriers (Wang et al. 2004), because their chemical

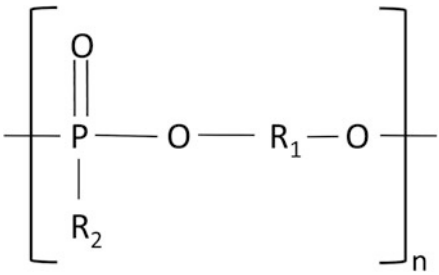


Fig. 9.9 General structure of polyphosphoesters

reactivity allows the conjugation of specific ligands to target the carrier to desired compartments of the cell.

9.3.8 Poly(amino acids)

Poly(amino acid)s as synthetic biomaterials have been explored for various applications (Anderson et al. 1985). However, their application is limited because of their poor or pH-dependent solubility and lack of functional groups and because they can cause immunogenic reactions (Nair and Laurencin 2006b). To overcome these limitations, pseudo-amino acid polymers that are composed of amino acids but which are not linked by peptide bonds but by esters, imino carbonates, and carbonates have been designed.

9.3.8.1 Pseudo-poly(amino acid)s

One of the most extensively investigated systems is derived from tyrosine using desaminotyrosyl-tyrosine alkyl esters as the building blocks (Fig. 9.10). The advantage of tyrosine is its aromatic ring, and the resulting polymers thus show good engineering properties. Several tyrosine-based pseudo-poly(amino acid)s with different nonpeptide linkages have been developed for biomaterial applications, such as polycarbonate (Pulapura and Kohn 1992), polyiminocarbonate (Li and Kohn 1989), polyacrylate (polymers made by reacting aromatic dicarboxylic acids with diphenols; they represent tough, durable, and heat-resistant thermoplastic polymers originally used in aerospace applications and for electronics) (Fiordeliso et al. 1994), and polyphosphate (Gupta and Lopina 2004). Also, tyrosine-based copolymers, e.g., with polyethylene glycol (PEG), have been developed (Bourke and Kohn 2003). The materials show excellent biocompatibility and are therefore suitable for tissue engineering application (Ertel and Kohn 1994). L-Tyrosine-based diphenolic dipeptides were also used to synthesize polyurethanes, using two different macrodiols, polyethylene glycol and PCL (Sarkar et al. 2009).

The most versatile tyrosine-derived polymers are the polycarbonates. They have the advantage that the glass transition temperatures (50–90 °C) and the mechanical properties (strength 50–70 Mpa, stiffness 1–2 Gpa) can be easily tailored by varying the pendant alkyl chain (Ertel and Kohn 1994). The good in vitro and in vivo osteocompatibility and mechanical properties of carbonate polymers make them suitable candidates for developing long-term orthopedic implants (James et al. 1999). A unique property of the tyrosine-derived polycarbonates is that they

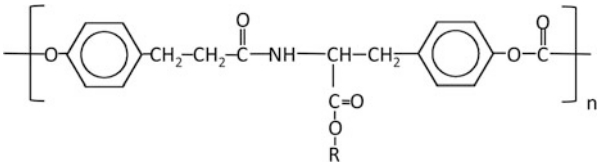


Fig. 9.10 Chemical structure of desaminotyrosyl-tyrosine alkyl esters

do not take up more than 5 % water during degradation and therefore maintain their shape for a longer period of time (Bourke and Kohn 2003).

Tyrosine-derived polyarylates are soft elastomeric alternatives that have faster degradation rates than the carbonates. They are unique as several of their properties (e.g., glass transition temperature, hydrophobicity/hydrophilicity, or crystallinity) are well predictable, thus allowing the development of polymers with a wide spectrum of properties from soft and elastomeric to strong materials, as well as from amorphous to liquid crystalline materials (Huang et al. 2009).

9.3.8.2 Poly(L-glutamic acid)

Poly(L-glutamic acid), an artificial protein, is consequently highly susceptible to degradation. A biodistribution study demonstrated that at a molecular weight of 11,000, the polymer can be largely recovered in the kidneys and urine with minimal retention in other tissues (Li 2002). Interestingly, it has been demonstrated not to be immunogenic (Li 2002). The pendant-free γ-carboxyl group in each repeating unit of L-glutamic acid is negatively charged at a neutral pH, which renders the polymer water soluble. The carboxyl groups also provide functionality for drug attachment. As will be seen later, PG is biodegradable and nontoxic. These features make PG a promising candidate as a carrier of polymer–drug conjugates for selective delivery of chemotherapeutic agents (Nicol et al. 2002).

L-PGA is usually prepared from poly(γ-benzyl-L-glutamate) by removing the benzyl protecting group with the use of hydrogen bromide (Idelson and Blout 1958). Alternatively, it can be synthesized by alkaline hydrolysis of poly(L-methyl glutamate), but since alkaline hydrolysis leads also to racemization, this method is less favored (Hanby et al. 1950).

L-PGA has also been extensively investigated for developing polymeric drugs by conjugating anticancer drugs to the polymer backbone. The conjugation has been shown to significantly increase the aqueous solubility, plasma distribution time, and tumor distribution of the drugs (for a detailed review, see Li 2002). Additionally, L-PGA has been cross-linked to gelatin or porcine collagen to achieve a biodegradable biological adhesive and hemostat with improved soft tissue binding and hemostatic properties (Otani et al. 1998; Sekine et al. 2001).

A related, but structurally different, polymer comprised of glutamic acid is poly (γ-glutamic acid). In this polymer, L-glutamic acid monomers are linked via amide bonds between γ-carboxyl and α-amino groups of adjacent monomers. Since it is a naturally occurring, a water-soluble polymer can be produced by *Bacillus subtilis* (Xu et al. 2014). It has the advantage over L-PGA that it cannot be degraded by proteolytic enzymes, although it can be hydrolyzed by mild acid conditions, where poly(α-glutamic acid) remains mostly intact. Its manifold applications have been reviewed by Ogunleye et al. (2014), but since it is not a synthetic biopolymer, it shall not be treated further here.

9.3.8.3 Poly(aspartic acid)

Aspartate is the other acidic natural amino acid and almost identical to glutamic acid except for one missing methylene group. Consequently, poly-L-aspartate

(PAA) shares many of the properties and advantages in application of poly-L-glutamate. Several block copolymers with aspartic acid and other synthetic biodegradable polymeric moieties have been developed to form core-forming micellar nanostructures for use as smart drug delivery vehicles (Matsumura et al. 2004; Oh et al. 2013). A chemically modified form of PAA—α, β-poly(*N*-2-hydroxyethyl)-D, L aspartamide—is a synthetic water-soluble and biocompatible polymer that undergoes extensive investigation as a plasma expander. PAA can also be converted into a hydrogel by high-energy radiations and is currently being investigated for various biomedical applications (Zahuriaan-Mehr et al. 2009; Thombre and Sarwade 2005).

Poly(aspartic acid) (PAA) can be easily synthesized from aspartic acid by thermal polymerization through a succinimide intermediate (Kovács et al. 1953).

9.4 Conclusions

The advances in synthetic organic chemistry have enabled the synthesis of a plethora of synthetic biopolymers and allowed their modification to serve manifold different requirements. All of them are biodegradable, but the enzymes and enzymatic mechanisms have in most cases not been studied in detail. Further progress in this area may thus be obtained by a more thorough analysis of their depolymerization, which may not only aid to the design of polymers with predictable degradation characteristics but also lead to the discovery of enzymes which may be used in the biocatalytic synthesis of these polymers from their monomers. In addition, the current progress in metabolic engineering (see chapter 7; this volume) offers the biotechnological preparation of many of the monomers and even some of the polymers without the involvement of hazardous chemicals and high temperatures. I anticipate that further success of biodegradable implants will strongly be influenced by these two lines of research and expand our ability to custom design or modify existing biomaterials toward appropriate biocompatibility, degradation, and physical properties.

References

Alarcon CDH, Pennadam S, Alexander C (2005) Stimuli responsive polymers for biomedical applications. Chem Soc Rev 34:276–285

Albertsson AC (ed) (2002) Degradable aliphatic polyesters. In: Advances in polymer sciences, Vol 157. Springer, Berlin

Allcock HR (2003) Chemistry and applications of polyphosphazenes. Wiley, New York

Anderson JM, Spilizewski KJ, Hiltner A (1985) Poly-alpha-amino acids as biomedical polymers. In: Williams DE (ed) Biocompatibility of tissue analogs. CRC Press Inc, Boca Raton, pp 67–88

Andreopoulos AG (2003) Classification and characteristics of polymeric controlled-release systems. Med Device Technol 14:20–21

Anonymous (2014) Antibacterial sutures for wound closure after surgery: a review of clinical and cost-effectiveness and guidelines for use [Internet]. Canadian Agency for Drugs and

Technologies in Health, Ottawa (ON). 21 Nov 2014. http://www.ncbi.nlm.nih.gov/books/NBK263224/

Aoshima S, Kanaoka S (2009) A renaissance in living cationic polymerization. Chem Rev 109:5245–5287

Arabuli N, Tsitlanadze G, Edilashvili L, Kharadze D, Goguadze T, Beridze V (1994) Heterochain polymers based on natural amino acids - synthesis and enzymatic-hydrolysis of regular poly (ester amide)s based on bis(L-phenylalanine) alpha, omega-alkylene diesters and adipic acid. Macromol Chem Phys 195:2279–2289

ASGE Technology Committee, Bhat YM, Banerjee S, Barth BA, Chauhan SS, Gottlieb KT, Konda V, Maple JT, Murad FM, Pfau PR, Pleskow DK, Siddiqui UD, Tokar JL, Wang A, Rodriguez SA (2013) Tissue adhesives: cyanoacrylate glue and fibrin sealant. Gastrointest Endosci 78:209–215

Asti A, Gioglio L (2014) Natural and synthetic biodegradable polymers: different scaffolds for cell expansion and tissue formation. Int J Artif Organs 37:187–205

Athanasiou KA, Agrawal CE, Barber FA, Burkhart SS (1998) Orthopaedic applications for PLA-PGA biodegradable polymers. J Arthrosc Relat Surg 14:726–737

Attawia MA, Uhrich KE, Botchwey E, Langer R, Laurencin CT (1996) *In vitro* bone biocompatibility of poly (anhydride-co-imides) containing pyromellitylimidoalanine. J Orthop Res 14:445–454

Baillargeon AL, Mequanint K (2014) Biodegradable polyphosphazene biomaterials for tissue engineering and delivery of therapeutics. Biomed Res Int 2014:761373

Bergsma JE, Rozema FR, Bos RR, Boering G, de Bruijn WC, Pennings AJ (1995) *In vivo* degradation and biocompatibility study of in vitro pre-degraded as-polymerized polylactide particles. Biomaterials 16:267–274

Bettinger CJ, Bruggeman JP, Borenstein JT, Langer RS (2008) Amino alcohol–based degradable poly(ester amide) elastomers. Biomaterials 29:2315–2325

Bini TB, Gao SJ, Wang S, Ramakrishna S (2005) Development of fibrous biodegradable polymer conduits for guided nerve regeneration. J Mater Sci Mater Med 16:367–375

Bourke SL, Kohn J (2003) Polymers derived from the amino acid L-tyrosine: polycarbonates, polyarylates and copolymers with poly(ethylene glycol). Adv Drug Deliv Rev 55:447–466

Brannon-Peppas L, Blanchette JO (2004) Nanoparticle and targeted systems for cancer therapy. Adv Drug Deliv Rev 56:1649–1659

Burkoth AK, Anseth KS (2000) A review of photocrosslinked polyanhydrides: *in situ* forming degradable networks. Biomaterials 21:2395–2404

Butterworth KW (1988) The loctite story. Newcomen Society of the United States, New York

Cameron DJ, Shaver MP (2011) Aliphatic polyester polymer stars: synthesis, properties and applications in biomedicine and nanotechnology. Chem Soc Rev 40:1761–1776

Castillo-Dali G, Velazquez-Cayon R, Serrera-Figallo MA, Rodriguez-Gonzalez-Elipe A, Gutierrez-Perez JL, Torres-Lagares D (2014) Importance of PLGA in scaffolds for guided bone regeneration: a focused review. J Oral Implantol. Feb 19 [Epub ahead of print]

Chen XH, McCarthy SP, Gross RA (1997) Synthesis, characterization, and epoxidation of an aliphatic polycarbonate from 2,2-(2-pentene-1,5-diyl)trimethylene carbonate ((c)HTC) ring-opening polymerization. Macromolecules 30:3470–3476

Chen Y, Zhu Z, Yu Y (2014) Novel methodologies in analysis of small molecule biomarkers and living cells. Tumour Biol 35:9469–9477

Chung CY, Yang JT, Kuo YC (2013) Polybutylcyanoacrylate nanoparticle-mediated neurotrophin-3 gene delivery for differentiating iPS cells into neurons. Biomaterials 34:5562–5570

Ciardelli G, Chiono V (2006) Materials for peripheral nerve regeneration. Macromol Biosci 6:13–26

Cook SE, Park IK, Kim EM, Jeong HJ, Park TG, Choi YJ, Akaike T, Cho CS (2005) Galactosylated polyethylenimine-graft-poly (vinyl pyrrolidone) as a hepatocyte-targeting gene carrier. J Control Release 105:151–163

Cooper JA, Lu HH, Ko FK, Freeman JW, Laurencin CT (2005) Fiber-based tissue-engineered scaffold for ligament replacement: design considerations and *in vitro* evaluation. Biomaterials 26:1523–1532

Croisier F, Atanasova G, Poumay Y, Jérôme C (2014) Polysaccharide-coated PCL nanofibers for wound dressing applications. Adv Health Mater 3:2032–2039

Dash TK, Konkimalla VB (2012) Polymeric modification and its implication in drug delivery: poly-ε-caprolactone (PCL) as a model polymer. Mol Pharm 9:2365–2379

Dhandayuthapani B, Yoshida J, Maekawa T, Sakthi Kumar D (2011) Polymeric scaffolds in tissue engineering application: a review. Int J Polymer Sci Article ID 290602. doi:10.1155/2011/290602

Dinarvand R, Sepehri N, Manoochehri S, Rouhani H, Atyabi F (2011) Polylactide-co-glycolide nanoparticles for controlled delivery of anticancer agents. Int J Nanomed 6:877–895

Duan J, Zhang Y, Chen W, Shen C, Liao M, Pan Y, Wang J, Deng X, Zhao J (2009) Cationic polybutyl cyanoacrylate nanoparticles for DNA delivery. J Biomed Biotechnol 2009:149254

Edlund U, Albertsson AC (2003) Polyesters based on diacid monomers. Adv Drug Deliv Rev 55:585–609

Elstad NL, Fowers KD (2009) OncoGel (ReGel/paclitaxel)–clinical applications for a novel paclitaxel delivery system. Adv Drug Deliv Rev 61:785–794

Ertel SI, Kohn J (1994) Evaluation of a series of tyrosine-derived polycarbonates for biomaterial applications. J Biomed Mater Res 28:919–30

Fiordeliso J, Bron S, Kohn J (1994) Design, synthesis and preliminary characterization of tyrosine-containing polyarylates: new biomaterials for medical applications. J Biomater Sci Polymer Ed 5:497–510

Funovics M, Weissleder R, Tung CH (2003) Protease sensors for bioimaging. Anal Bioanal Chem 377:956–963

Garbayo E, Ansorena E, Blanco-Prieto MJ (2013) Drug development in Parkinson's disease: from emerging molecules to innovative drug delivery systems. Maturitas 76:272–278

Gujral C, Minagawa Y, Fujimoto K, Kitano H, Nakaji-Hirabayashi T (2013) Biodegradable microparticles for strictly regulating the release of neurotrophic factors. J Control Release 168:307–316

Gunatillake PA, Adhikari R (2003) Biodegradable synthetic polymers for tissue engineering. Eur Cells Mater 5:1–16

Gunatillake P, Mayadunne R, Adhikari R (2006) Recent developments in biodegradable synthetic polymers. Biotechnol Ann Rev 12:301–347

Gupta AS, Lopina ST (2004) Synthesis and characterization of L-tyrosine based novel polyphosphates for potential biomaterial applications. Polymer 45:4653–4662

Hanby WE, Waley SG, Watson J (1950) Synthetic polypeptides. Part II. Polyglutamic acid. J Chem Soc 1950:3239–3249

Harper E, Dang W, Lapidus RG, Garver RI Jr (1999) Enhanced efficacy of a novel controlled release paclitaxel formulation (PACLIMER delivery system) for local-regional therapy of lung cancer tumor nodules in mice. Clin Cancer Res 5:4242–4248

He W, Yong T, Teo WE, Ma ZW, Ramakrishna S (2005) Fabrication and endothelialization of collagen-blended biodegradable polymer nanofibers: potential vascular graft for blood vessel tissue engineering. Tissue Eng 11:1574–1588

Heller J, Barr J, Ng SY, Abdellauoi KS, Gurny R (2002) Poly(ortho esters): synthesis, characterization, properties and uses. Adv Drug Deliv Rev 54:1015–1039

Hernandez N, Williams RC, Cochrane EW (2014) The battle for the "green" polymer. Different approaches for biopolymer synthesis: bioadvantaged vs. Bioreplacement. Org Biomol Chem 12:2834–2837

Huang X, Shen CY, Chen JC, Li Q (2009) A library of L-tyrosine-derived biodegradable polyarylates for potential biomaterial applications. Part I: Synthesis, characterization and accelerated hydrolytic degradation. J Biomater Sci Polym Ed 20:935–955

Huebsch N, Mooney DJ (2009) Inspiration and application in the evolution of biomaterials. Nature 462:426–432

Idelson M, Blout ER (1958) High molecular weight poly(α-l-glutamic acid): preparation and physical rotation changes. J Am Chem Soc 80:4631–4634

Isenberg BC, Williams C, Tranquillo RT (2006) Small-diameter artificial arteries engineered *in vitro*. Circ Res 98:25–35

James K, Levene H, Parsons JR, Kohn J (1999) Small changes in polymer chemistry has a large effect on the bone-implant interface: evaluation of a series of degradable tyrosine-derived polycarbonates in bone defects. Biomaterials 20:2203–2212

Jérôme C, Lecomte P (2008) Recent advances in the synthesis of aliphatic polyesters by ring-opening polymerization. Adv Drug Deliver Rev 60:1056–1076

Katsarava R, Kharadze D, Kirmelashvili L, Medzmariashvili N, Goguadze T, Tsitlanadze G (1993) Polyamides from 2,2´-P-phenylenebis (Δ2-5-oxazolone)s and N,N´-bistrimethylsilylated diamines-synthesis of polyamides containing dipeptide links in the main chains. Makromol Chem Phys 194:143–150

Katsarava R, Beridze V, Arabuli N, Kharadze D, Chu CC, Won CY (1999) Amino acid-based bioanalogous polymers. Synthesis, and study of regular poly(ester amide)s based on bis(alpha-amino acid) alpha, omega-alkylene diesters, and aliphatic dicarboxylic acids. J Polym Sci Polym Chem 37:391–407

Kotwal A, Schmidt CE (2001) Electrical stimulation alters protein adsorption and nerve cell interactions with electrically conducting biomaterials. Biomaterials 22:1055–1064

Kovács J, Könyves I, Pusztai Á (1953) Darstellung von Polyasparaginsäuren (Polyaspartsäuren) aus dem thermischen Autokondensationsprodukt der Asparaginsäure. Experientia 9:459–460

Kukreja N, Onuma Y, Daemen J, Serruys PW (2008) The future of drug-eluting stents. Pharmacol Res 57:171–180

Kumar N, Langer RS, Domb AJ (2002) Polyanhydrides: an overview. Adv Drug Deliv Rev 54:889–910

Kumbar SG, Bhattacharyya S, Nukavarapu SP, Khan YM, Nair LS, Laurencin CT (2006) *In vitro* and *in vivo* characterization of biodegradable poly (organo phosphazenes) for biomedical applications. J Inorg Organomet Chem 16:365–385

Lai P, Daear W, Löbenberg R, Prenner EJ (2014) Overview of the preparation of organic polymeric nanoparticles for drug delivery based on gelatine, chitosan, poly(D, L-lactide-co-glycolic acid) and polyalkylcyanoacrylate. Colloids Surf B Biointerfaces 118:154–163

Lapidus RG, Dang W, Rosen DM, Gady AM, Zabelinka Y, O'Meally R, DeWeese TL, Denmeade SR (2004) Anti-tumor effect of combination therapy with intratumoral controlled-release paclitaxel (PACLIMER microspheres) and radiation. Prostate 58:291–298

Laurencin CT, Nair LS (2004) Polyphosphazene nanofibers for biomedical applications: preliminary studies. In: Guceri S, Gogotsi YG, Kuznetsov V (eds) Nanoengineered nanofibrous materials, NATO-ASI proceedings. Kluwer, Dordrecht, pp 281–300

Leinonen S, Suokas E, Veiranto M, Tormala P, Waris T, Ashammakhi N (2002) Healing power of bioadsorbable ciprofloxacin-containing self reinforced poly(L/DL-lactide 70/30 bioactive glass 13 miniscrews in human cadaver bone. J Cranifac Surg 13:212–218

Leong KW, Brott BC, Langer R (1985) Biodegradable polyanhydrides as drug carrier matrices. Characterization, degradation and release characteristics. J Biomed Mater Res 19:941–55

Li C (2002) Poly(L-glutamic acid) - anticancer drug conjugates. Adv Drug Deliv Rev 54:695–713

Li C, Kohn J (1989) Synthesis of poly(iminocarbonates): degradable polymers with potential application as disposable plastics and as biomaterials. Macromolecules 22:2036–2039

Li LC, Deng J, Stephens D (2002) Polyanhydride implant for antibiotic delivery – from the bench to the clinic. Adv Drug Deliv Rev 54:963–986

Li J, Wang Y, Zhu Y, Oupický D (2013) Recent advances in delivery of drug-nucleic acid combinations for cancer treatment. J Control Release 172:589–600

Liu XH, Ma PX (2004) Polymeric scaffolds for bone tissue engineering. Ann Biomed Eng 32:477–486

Lofgren A, Albertsson AC, Dubois P, Herome R (1995) Recent advances in ring opening polymerization of lactones and related compounds. J Macromol Sci Rev Macromol Chem Phys C35:379–418
Lotfi M, Ghasemi N, Rahimi S, Vosoughhosseini S, Saghiri MA, Shahidi A (2013) Resilon: a comprehensive literature review. J Dent Res Dent Clin Dent Prospects 7:119–130
Lu HH, Cooper JA, Manuel S, Freeman JW, Attawia MA, Ko FK et al (2005) Anterior cruciate ligament regeneration using braided biodegradable scaffolds: *in vitro* optimization studies. Biomaterials 26:4805–4816
Lyu S, Untereker D (2009) Degradability of polymers for implantable biomedical devices. Int J Mol Sci 10:4033–4065
Ma G, Song C, Sun H, Yang J, Leng X (2006) A biodegradable levonorgestrel-releasing implant made of PCL/F68 compound as tested in rats and dogs. Contraception 74:141–147
Matsumoto K, Taguchi S (2013) Biosynthetic polyesters consisting of 2-hydroxyalkanoic acids: current challenges and unresolved questions. Appl Microbiol Biotechnol 97:8011–8021
Matsumura Y, Hamaguchi T, Ura T, Muro K, Yamada Y, Shimada Y et al (2004) Phase I clinical trial and pharmacokinetic evaluation of NK911, micelle-encapsulated doxorubicin. Br J Cancer 91:1775–1781
Middleton JC, Tipton AJ (1998) Synthetic biodegradable polymers as medical devices. Med Plast Biomater 1998:31–38
Middleton JC, Tipton AJ (2000) Synthetic biodegradable polymers as orthopaedic devices. Biomaterials 21:2335–46
Miller RA, Brady JM, Cutright DE (1977) Degradation rates of oral resorbable implants (polylactates and polyglycolates): rate modification with changes in PLA/PGA copolymer ratios. J Biomed Mater Res 11:711–719
Muggli DS, Burkoth AK, Keyser SA, Lee HR, Anseth KS (1998) Reaction behaviour of biodegradable, photocross-linkable polyanhydrides. Macromolecules 31:4120–4125
Nagpal S (2012) The role of BCNU polymer wafers (Gliadel) in the treatment of malignant glioma. Neurosurg Clin N Am 23:289–95
Nair LS, Laurencin CT (2006a) Biodegradable polymers as biomaterials. Prog Polym Sci 32:762–798
Nair LS, Laurencin CT (2006b) Polymers as biomaterials for tissue engineering and controlled drug delivery. Adv Biochem Eng Biotechnol 102:47–90
Nair LS, Lee D, Laurencin CT (2004) Polyphosphazenes as novel biomaterials. In: Narasimhan A, Mallapragada A (eds) Handbook of biodegradable polymeric materials and applications. American Scientific Publication, Los Angeles, CA, pp 277–306
Nair LS, Khan YM, Laurencin CT (2006a) Polyphosphazenes. In: Hollinger H (ed) An introduction to biomaterials. CRC, Boca Raton, pp 273–290
Nair LS, Lee DA, Bender JD, Barrett EW, Greigh YE, Brown PW et al (2006b) Synthesis, characterization and osteocompatibility evaluations of novel alanine based polyphosphazenes. J Biomed Mater Res 76A:206–213
Nasongkla N, Shuai X, Ai H, Weinberg BD, Pink J, Boothman DA, Gao JM (2004) cRGD-functionalized polymer micelles for targeted doxorubicin delivery. Angew Chem Int Ed 43:6323–6327
Neuenschwander S, Hoerstrup SP (2004) Heart valve tissue engineering. Transplant Immunol 12:359–365
Nicol F, Wong M, MacLaughlin FC, Perrard J, Wilson E, Nordstrom JL et al (2002) L-glutamate, an anionic polymer, enhances transgene expression for plasmids delivered by intramuscular injection with *in vivo* electroporation. Gene Ther 9:1351–1358
Nuyken O, Pask SD (2013) Ring-opening polymerization—an introductory review. Polymers 5:361–403
Ogunleye A, Bhat A, Irorere VU, Hill D, Williams C (2014) Radecka I (2014) Poly-γ-glutamic acid - production, properties and applications. Microbiology. doi:10.1099/mic.0.081448-0
Oh NM, Oh KT, Youn YS, Lee DK, Cha KH, Lee DH, Lee ES (2013) Poly(L-aspartic acid) nanogels for lysosome-selective antitumor drug delivery. Colloids Surf B Biointerfaces 101:298–306

Otani Y, Tabata Y, Ikada Y (1998) Hemostatic capability of rapidly curable from gelatin, poly (L-glutamic acid) and carbodiimide. Biomaterials 19:2091–2098
Palmer RJ (2001) Polyamides, plastics. Encycl Polymer Sci Technol. doi:10.1002/0471440264.pst251
Pan Z, Ding J (2012) Poly(lactide-co-glycolide) porous scaffolds for tissue engineering and regenerative medicine. Interface Focus 2:366–377
Piotrowska U, Sobczak M (2014) Enzymatic polymerization of cyclic monomers in ionic liquids as a prospective synthesis method for polyesters used in drug delivery systems. Molecules 20:1–23
Place ES, George JH, Williams CK, Stevens MM (2009) Synthetic polymer scaffolds for tissue engineering. Chem Soc Rev 38:1139–1151
Prior TD, Grace DL, MacLean JB, Allen PW, Chapman PG, Day A (1997) Correction of hallux abductus valgus by Mitchell's metatarsal osteotomy: comparing standard fixation methods with absorbable polydioxanone pins. Foot 7:121–125
Pulapura S, Kohn J (1992) Tyrosine-derived polycarbonates: backbone-modified "pseudo"-poly (amino acids) designed for biomedical applications. Biopolymers 32:411–417
Pyo SH, Persson P, Mollaahmad MA, Sörensen K, Lundmark S, Hatti-Kaul R (2012) Cyclic carbonates as monomers for phosgene- and isocyanate-free polyurethanes and polycarbonates. Pure Appl Chem 84:637
Ratner BD, Bryant SJ (2004) Biomaterials: where we have been and where we are going. Annu Rev Biomed Eng 6:41–75
Ren TB, Ren J, Jia XZ, Pan KF (2005) The bone formation *in vitro* and mandibular defect repair using PLGA porous scaffolds. J Biomed Mater Res A 74:562–569
Rezwan K, Chen QZ, Blaker JJ, Boccaccini AR (2006) Biodegradable and bioactive porous polymer/inorganic composite scaffolds for bone tissue engineering. Biomaterials 27:3413–3431
Ricci JA, Parekh NN, Desai NS (2014) Diffuse cutaneous allergic reaction to dermabond. Prehosp Disaster Med 29:546–548
Romagnoli C, D'Asta F, Brandi ML (2013) Drug delivery using composite scaffolds in the context of bone tissue engineering. Clin Cases Miner Bone Metab 10:155–161
Sadat Tabatabaei Mirakabad F, Nejati-Koshki K, Akbarzadeh A, Yamchi MR, Milani M, Zarghami N, Zeighamian V, Rahimzadeh A, Alimohammadi S, Hanifehpour Y, Joo SW (2014) PLGA-based nanoparticles as cancer drug delivery systems. Asian Pac J Cancer Prev 15:517–535
Sahoo S, Cho-Hong JG, Siew-Lok T (2007) Development of hybrid polymer scaffolds for potential applications in ligament and tendon tissue engineering. Biomed Mater 2:169–173
Sarkar D, Yang JC, Gupta AS, Lopina ST (2009) Synthesis and characterization of L-tyrosine based polyurethanes for biomaterial applications. J Biomed Mater Res A 90:263–71
Schmidt CE, Shastri VR, Vacanti JP, Langer R (1997) Stimulation of neurite outgrowth using an electrically conducting polymer. Proc Natl Acad Sci USA 94:8948–8953
Sekine T, Nakamura T, Shimizu Y, Ueda H, Matsumoto K, Takimoto Y et al (2001) A new type of surgical adhesive made from porcine collagen and polyglutamic acid. J Biomed Mater Res 54:305–310
Seyednejad H, Ghassemi AH, van Nostrum CF, Vermonden T, Hennink WE (2011) Functional aliphatic polyesters for biomedical and pharmaceutical applications. J Control Release 152:168–176
Sheth SD, Giugliano RP (1995) Coronary artery stents: advances in technology. Hosp Pract 42:83–91
Shi R, Chen D, Liu Q, Wu Y, Xu X, Zhang L, Tian W (2009) Recent advances in synthetic bioelastomers. Int J Mol Sci 10:4223–4256
Silberman M, Nelson KD, Eberhart RC (2005) Mechanical properties and in vitro degradation of bioresorbable fibers and expandable fiber-based stents. J Biomed Mater Res B Appl Biomater 74:792–799
Stridsberg KM, Ryner M, Albertsson AN (2002) Controlled ring-opening polymerization: polymers with designed macromolecular architecture. Adv Polymer Sci 157:41–65

Terasaka S, Iwasaki Y, Shinya N, Uchida T (2006) Fibrin glue and polyglycolic acid nonwoven fabric as a biocompatible dural substitute. Neurosurgery 58:134–139
Thombre SM, Sarwade BD (2005) Synthesis and biodegradability of polyaspartic acid: a critical review. J Macromol Sci A 42:1299–1315
Tiainen J, Veiranto M, Suokas E, Tormala P, Waris T, Ninkoviv M et al (2002) Bioabsorbable ciprofloxacin-containing and plain self reinforced polylactide-polyglycolide 80/20 screws: pullout strength properties in human cadaver parietal bones. J Craniofac Surg 13:427–433
Tian H, Tang Z, Zhuang X, Chen X, Jing X (2012) Biodegradable synthetic polymers: preparation, functionalization and biomedical application. Prog Polym Sci 37:237–280
Ueda H, Tabata Y (2003) Polyhydroxyalkanonate derivatives in current clinical applications and trials. Adv Drug Deliv Rev 55:501–518
Uhrich KE, Gupta A, Thomas TT, Laurencin C, Langer R (1995) Synthesis and characterization of degradable polyanhydrides. Macromolecule 28:2148–2193
Ulery BD, Nair LS, Laurencin CT (2011) Biomedical applications of biodegradable polymers. J Polym Sci B Polym Phys 49:832–864
Vacanti JP (2012) Tissue engineering and the road to whole organs. Br J Surg 99:451–453
Veld in't PJA, Dijkstra PJ, Feijen J (1992) Synthesis of biodegradable polyesteramides with pendant functional groups. Makro Chem 193:2713–2730
Venkatraman S, Boey F, Lao LL (2008) Implanted cardiovascular polymers: natural, synthetic and bio-inspired. Prog Polym Sci 33:853–874
Venugopal J, Ramakrishna S (2005) Biocompatible nanofiber matrices for the engineering of a dermal substitute for skin regeneration. Tissue Eng 11:847–854
Viúdez A, Ramírez N, Hernández-García I, Carvalho FL, Vera R, Hidalgo M (2014) Nab-paclitaxel: a flattering facelift. Crit Rev Oncol Hematol 92:166–180
Wang J, Gao SJ, Zhang PC, Wang S, Mao HQ, Leong KW (2004) Polyphosphoramidate gene carriers: effect of charge group on gene transfer efficiency. Gene Ther 11:1001–1010
Wang YC, Yuan YY, Du JZ, Yang XZ, Wang J (2009) Recent progress in polyphosphoesters: from controlled synthesis to biomedical applications. Macromol Biosci 9:1154–1164
Wang X, Tai Z, Zhang W, Gao S (2014) Current status on gene therapy for hepatocellular carcinoma, with particular focus on the gene delivery approaches. Curr Gene Ther. Dec 23 [Epub ahead of print]
Wang Y, Tashiro Y, Sonomoto K (2015) Fermentative production of lactic acid from renewable materials: recent achievements, prospects, and limits. J Biosci Bioeng 119:10–18
Williams CK (2007) Synthesis of functionalized biodegradable polyesters. Chem Soc Rev 36:1573–1580
Wong SY, Pelet JM, Putnam D (2007) Polymer systems for gene delivery—past, present, and future. Prog Polym Sci 32:799–837
Woodruff MA, Hutmacher DW (2010) The return of a forgotten polymer-polycaprolactone in the 21st century. Prog Polym Sci 35:1217–1256
Xu Z, Feng X, Zhang D, Tang B, Lei P, Liang J, Xu H (2014) Enhanced poly(γ-glutamic acid) fermentation by Bacillus subtilis NX-2 immobilized in an aerobic plant fibrous-bed bioreactor. Bioresour Technol 155:8–14
Xudong L, Christophe D, Robert J (2003) Novel aliphatic polyesters based on functional cyclic (di)esters. Macromol Rapid Commun 24:161–172
Yoo HS, Lee EA, Yoon JJ, Park TG (2005) Hyaluronic acid modified biodegradable scaffolds for cartilage tissue engineering. Biomaterials 26:1925–1933
Zahuriaan-Mehr MJ, Pourjavadi A, Salimi H, Kurdtabar M (2009) Protein- and homo poly(amino acid)-based hydrogels with super-swelling properties. Polymer Adv Technol 20:655–671
Zhang Y, Zhang Y, Chen J, Zhang B, Pan Y, Ren L, Zhao J, Luo Y, Zhai D, Wang S, Wang J (2007) Polybutylcyanoacrylate nanoparticles as novel vectors in cancer gene therapy. Nanomedicine 3:144–153

Xenobiotic Life

10

Dario Cecchi and Sheref S. Mansy

Contents

10.1 Controlling Cellular Activity with Artificial RNA Molecules 339
10.2 Controlling Cellular Activity with Engineered Proteins 341
10.3 Xeno-nucleic Acids 343
10.4 Xeno-amino Acids 346
10.5 Artificial Cells 348
10.6 Conclusion 351
References 351

Abstract

The ability to generate living systems that exploit unnatural molecules has expanded greatly. There are a series of unnatural nucleotides, amino acids, and membranes to choose from. However, at this early stage, many challenges remain in implementing the newly built parts inside of living cells. For example, a library of unnatural nucleotides is available with altered heterocycle or sugar units, but aside from one impressive example, no genomes have been stably altered in a way that incorporates these artificial nucleotides. Similar problems exist for proteins in that current technology allows for one or two unnatural amino acids to be incorporated into a single protein inside of a cell but not more. Many of the encountered difficulties arise from interference with multiple cellular pathways that have evolved over billions of years of evolution to be codependent. That is why some success has been found in exploiting the newly made artificial systems in vitro. Just as one example, artificial, nonliving systems can be built with a single gene that can be used to direct the synthesis of two entirely different protein sequences with no limit on the number of unnatural

D. Cecchi • S.S. Mansy (✉)
CIBIO, University of Trento, via Sommarive 9, 38123 Povo, Italy
e-mail: mansy@science.unitn.it

A. Glieder et al. (eds.), *Synthetic Biology*, DOI 10.1007/978-3-319-22708-5_10

amino acids. However, a different set of challenges exists for nonliving xenobiotic systems that severely limit their utility. Nevertheless, remarkable progress has been made within and outside of living systems.

The term xenobiotic could be interpreted to mean many things, from the presence of foreign, i.e., xeno, molecules within a living organism that are not normally found in biology to the existence of natural molecules within a new host cell, or even altered levels of molecules normally found within the organism. Although from some vantage points, all technology based on recombinant gene expression could be categorized as xenobiotic, we will focus here on work that either pushes the boundaries of living technologies or has the potential to reveal answers to fundamental biological questions.

Nucleic acids are the most tractable facet of biology. Therefore, much of the work in xenobiology is built upon advances in nucleic acid chemistry that began over half a century ago. Oligonucleotides have been synthesized since the 1950s (Michelson and Todd 1955; Gilham and Khorana 1958) and used to direct the cell-free synthesis of peptides in the 1960s (Khorana 1965). Synthetic genes encoding functional tRNAs and peptides were made in the 1970s (Khorana et al. 1972; Itakura et al. 1977), and proteins were produced by synthetic genes in the 1980s (Edge et al. 1981). However, more than 20 years passed before DNA synthesis technology advanced sufficiently to allow for the construction of synthetic genomes. Eventually, synthetic viruses (Cello et al. 2002; Smith et al. 2003), a bacterial genome (Gibson et al. 2008), and a eukaryotic chromosome (Annaluru et al. 2014) were all synthesized and found to be functional. Genomes can also be engineered without resynthesizing the entire genome. The cleavage of specific regions of genomic DNA can be guided through protein–DNA interactions (TALENs) (Christian et al. 2010), by manipulations of the bacterial defense machinery (CRISPR/Cas9) (Marraffini and Sontheimer 2010; Mali et al. 2013), and genomic changes can be made at multiple sites contemporarily through multiplex automated genome engineering (MAGE) (Wang et al. 2009). Supporting technology continues to advance today. In 2003 the J. Craig Venter Institute took 5 years to synthesize the Phi-X174 bacteriophage genome, whereas approximately 2 weeks were only needed for Autodesk, a software company in California, to repeat the feat in 2014. But the ability to assemble genomes is not the same as understanding how to build a xenobiotic organism or even a simple, free-living, artificial cell. When a hybrid organism was created by inserting the *Synechocystis* genome into the *Bacillus subtilis* genome, the resulting organism was only able to grow on media that supported the growth of *B. subtilis* (Itaya et al. 2005). That is, the organism may have been a genomic hybrid, but the organism was not a functional hybrid that retained equivalently the properties encoded within both parent genomes. Problems are even observed when manipulating the much smaller genome of a bacteriophage. Some of the overlapping genes of T7 bacteriophage were pulled apart to facilitate subsequent modeling and engineering work.

However, this seemingly small change to the genome, small in that none of the protein sequences were knowingly modified, resulted in bacteriophages with altered infectivity (Chan et al. 2005). After over 60 years of advancing DNA technology culminating in the synthesis of genomes, we still know relatively little about what we are synthesizing (Torino et al. 2013). We do not know the function of many of the genes that are essential for cell survival (Glass et al. 2006). We cannot accurately predict expression levels, and we cannot anticipate interfering interactions between newly inserted genetic elements with existing cell architecture.

So far, technological advances in nucleic acid synthesis and manipulation have only been used to synthesize genomes that are essentially equivalent to those of extant cellular life. Progress in making more fundamental changes to the mechanics of cellular function has been much slower. Nevertheless, a steady path forward has been paved by several laboratories that design and implement mechanisms to control gene expression, expanded the chemical building blocks of life, and synthesize organelle-like entities. There is still much left to be done before truly different forms of life are synthesized, either from scratch or by modifying existing organisms, but the process has begun.

10.1 Controlling Cellular Activity with Artificial RNA Molecules

The easiest biological molecules to build are nucleic acids, and the functional versatility of RNA lends itself to a broad spectrum of possibilities in engineering new cellular behavior. Through simple base-pairing rules, RNA molecules that hybridize with target sequences can be easily designed to regulate gene expression (Liang et al. 2011; Isaacs et al. 2006). When more complex designs are needed, several computational methods are available to calculate RNA molecules with desired function (Beisel and Smolke 2009; Sparkman-Yager et al. 2015; Rodrigo et al. 2012; Salis et al. 2009). Importantly, functional RNA molecules can also be generated through selection techniques (Ellington and Szostak 1990; Tuerk and Gold 1990; Robertson and Joyce 1990), thereby facilitating the construction of molecules beyond our computational capabilities. In vitro selections have generated aptamers (RNA sequences that bind ligands), ribozymes, and aptazymes (allosteric ribozymes built from aptamer and ribozyme domains) (Breaker and Joyce 2014; Famulok and Mayer 2014). After decades of such largely in vitro studies, many RNA-based biological parts are now available for use inside of living cells.

The three main ways that gene expression can be engineered through the activity of artificial RNA sequences are by intervening within processes that regulate mRNA stability, transcription, and translation (Fig. 10.1). Since structured RNAs are more resistant to the activity of nucleases, the incorporation of self-cleaving ribozyme sequences at the termini of mRNA can greatly affect mRNA half-life and thus ultimately protein levels (Liang et al. 2011). If aptazymes are used, then mRNA half-life can be controlled by the presence or absence of a small molecule (Carothers et al. 2011) in a manner similar to that seen with the glmS riboswitch

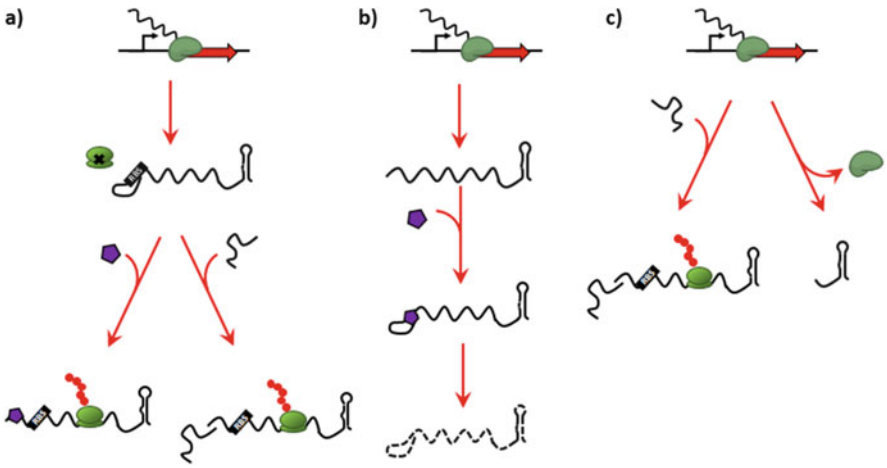


Fig. 10.1 Controlling gene expression with RNA. Due to the modularity and simple base-pairing rules of RNA, it is possible to control gene expression through the activity of RNA in at least three ways. (**a**) Access to the ribosome can be blocked by the folding of the 5′-UTR, which can be opened either through the binding of small-molecule ligands or through hybridization with a complementary RNA strand. (**b**) An aptazyme at the 5′-UTR can selectively control the degradation of the mRNA in response to the binding of the cognate ligand. (**c**) RNA molecules can prevent premature termination of transcription

(Winkler et al. 2004). Control of transcription is typically operated through modulation of stem–loop termination structures. For example, RNA molecules can be designed to work in trans by hybridizing to and thus altering a termination structure within the 5'-UTR of mRNA (Lucks et al. 2011; Brantl and Wagner 2000). Similar trans-acting strand exchange reactions are exploitable at the translation level. Riboregulators can be engineered to expose mRNA–ribosome binding sites that are otherwise inactive for translation due to internal base pairing (Isaacs et al. 2004). An updated version of this system exploits mRNA structures with a blocked start codon that is liberated upon interaction with a trigger RNA strand (Green et al. 2014). The strength of these systems is the ease of engineering new pathways that are responsive to changing expression levels of the cell. For example, the upregulation of one pathway could be engineered to lead to a cascade of events that activate or suppress other pathways through mRNA–mRNA interactions mediated by sequences placed within the untranslated regions of the mRNA molecules. However, these systems do not typically exploit RNA to sense environmental changes. Instead, protein activity is relied upon to initiate the engineered pathways.

Cis-acting RNA-based regulatory systems typically depend on riboswitch sequences. Artificial riboswitches can be generated by selecting for a short linker sequence between a previously identified aptamer and an output protein-coding sequence. Although natural riboswitches usually turn off protein expression upon ligand binding, artificial riboswitches tend to activate gene expression in response

to ligand binding (Topp and Gallivan 2007; Nomura and Yokobayashi 2007; Suess et al. 2004; Ogawa 2011). In theory, since aptamers can be generated to bind a variety of ligands, it should be possible to build new artificial riboswitches that respond to a series of molecules for which no biological equivalent protein sensor exists. However, in practice, riboswitches have shown remarkably little variety in what can be sensed. Nearly all artificial riboswitches regulate gene expression in response to theophylline, thiamine pyrophosphate, or tetracycline. Until robust methods are developed that can generate riboswitches with new sensing capabilities, riboswitches will remain primarily a tool for proof-of-concept studies and likely will not be used to engineer organisms with exploitable features.

RNA can be used for more than controlling gene expression. Since nucleic acids can be made to fold into desired shapes (Seeman 1982; Rothemund 2006), and since there are available RNA aptamers that bind specific protein domains, genetically encoded RNA scaffolds that bring together cellular components can be built. This is not much different than the strategies used by natural cells that exploit localization to organelles, proteinaceous microcompartments, or metabolons to regulate metabolic flux. One advantage of localization is the ability to more efficiently funnel molecules down desired synthetic paths, thereby avoiding the loss of intermediates (that could be toxic) through diffusion to other cellular processes (Chen and Silver 2012). Thus far, RNA scaffolds have been used in vivo to increase the efficiency of hydrogen, pentadecane, and succinate production (Delebecque et al. 2011; Sachdeva et al. 2014).

10.2 Controlling Cellular Activity with Engineered Proteins

The situation with engineered protein-based systems is similar to that of RNA in that a variety of tools are available to build proteins that inhibit or activate gene expression (Fig. 10.2). The rules for building zinc-finger and transcription activator-like (TAL) effector domains to bind to specific DNA sequences are known (Moscou and Bogdanove 2009; Boch et al. 2009; Joung and Sander 2013; Copeland et al. 2014). Each Cys2His2 zinc finger recognizes three nucleotides and can be fused together to build longer recognition motifs (Copeland et al. 2014). TAL effector proteins contain repeating units approximately 34 amino acids in length of which two consecutive amino acids dictate single-nucleotide specificity (Moscou and Bogdanove 2009; Boch et al. 2009). More recently, the CRISPR system has gained popularity since this system uses an RNA sequence as a guide to bind to DNA. In other words, by simply changing the sequence of the guide RNA, Cas9 can be targeted to different sites on the genome (Qi et al. 2013). Since Cas9 has nuclease activity, a nuclease-inactivated mutant form of the protein is used. All of these programmable DNA-binding systems can be targeted to promoter sequences and thus can block transcription through a simple steric-based mechanism. Additionally, activators rather than repressors can be built if these proteins are fused to a domain that recruits RNA polymerase (natural TAL effectors do not need such a domain in plants). While these techniques have been more widely

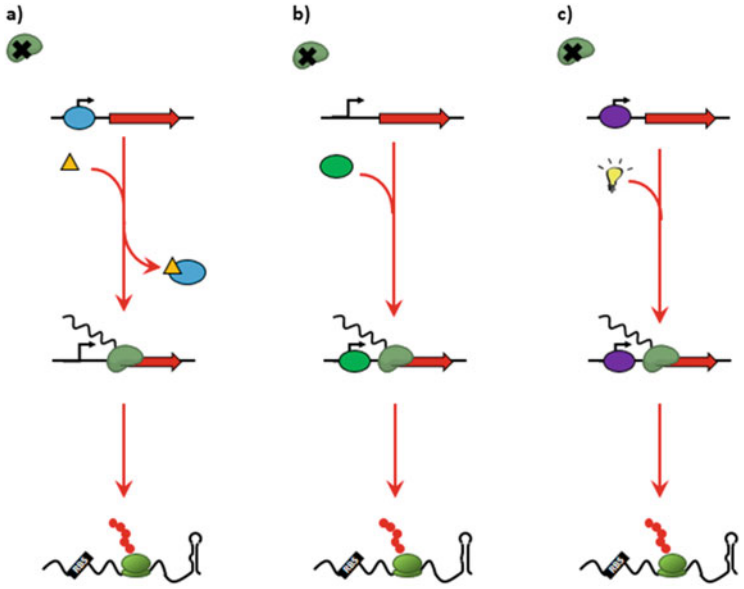


Fig. 10.2 Controlling gene expression through proteins. Proteins can regulate gene expression by (**a**) preventing the binding of the RNA polymerases or by (**b**) specifically recruiting the RNA polymerase upstream of the gene of interest. Protein-mediated transcriptional activation can be in response to light (**c**)

adopted in eukaryotic systems, implementation in bacteria has been demonstrated (Politz et al. 2013; Nielsen and Voigt 2014; Bikard et al. 2013).

None of the programmable DNA targeting systems just described are ligand responsive. Natural transcription factors, such as those of the well-studied LacI and TetR family of repressors, regulate transcription in response to ligand availability. The way to get some measure of programmability out of such transcription factors would be to mix and match DNA-binding and ligand-binding domains from different transcription factors, that is, to generate functional chimera proteins. Progress in doing just that has been made with the LacI family of repressors. By maintaining the same DNA-binding domain of LacI and substituting in the ligand-binding domain of other family members, new chimera proteins capable of sensing different sugars and a nucleoside were produced (Meinhardt et al. 2012). DNA specificity could be changed by making point mutations (Nielsen and Voigt 2014); however, substitutions of the DNA-binding domain with those of other family members were not reported. Assuming that DNA-binding domain substitutions would yield functionally active chimera, only about 10 different promoters would be available from the set of LacI family members investigated thus far for the construction of chimera. The number of DNA-binding sequences could be greatly expanded through genomic mining, as has recently been reported for TetR homologues (Stanton et al. 2014). Similar strategies to uncover the ligand-binding specificities of transcription factors represent a much bigger challenge. Although

the protein-based small-molecule sensing mechanisms have outperformed the RNA-based mechanisms, the number of available biological parts shown to consistently work is, nevertheless, small.

Other means of sensing are possible. For example, bacterial two-component systems essentially separate the sensing and DNA-binding functionalities on two different proteins. Signals are transduced between the two through phosphorylation. Chimera of the sensor-kinase component can be made to alter the input signal that activates the response regulator, i.e., the DNA-binding component of the two-component system. One of the more well-known synthetic biology accomplishments, that of producing photographic *E. coli* (Levskaya et al. 2005), was dependent upon such an engineered two-component system. Further advances in engineering light regulation of gene expression have led to the generation of promising tools. In mammalian cells light-responsive proteins were fused to transactivation domains and zinc finger (Polstein and Gersbach 2012, 2014), TAL effector (Konermann et al. 2013), or nuclease-inactivated Cas9 (Polstein and Gersbach 2015; Nihongaki et al. 2015) to control transcription. Similar optogenetic tools were exploited to regulate heart muscles (Bruegmann et al. 2010), neural activity in the brain (Chow et al. 2010), and gene expression (Ye et al. 2011). Remarkably, the switching of the light itself can be controlled by brain activity so that light-responsive gene expression can be activated by the brain (Folcher et al. 2014). The possibility to convert brain activity into gene expression may allow one day for therapies that are responsive to the mental state of the patient.

Proteins can also be engineered to serve as scaffolds to organize enzymes for more efficient synthesis, as was previously described for RNA above (Chen and Silver 2012; Good et al. 2011). Engineered scaffolds often make use of protein–protein interaction domains, such as SH3, PDZ, and leucine zippers, and have been used to improve the yield of mevalonate and glucaric acid synthesis and as a means of controlling signal transduction (Dueber et al. 2009; Whitaker et al. 2012). An alternative route toward localization is to entrap the enzymes inside of cage-like structures composed of proteins, that is, inside of bacterial microcompartments (Chen and Silver 2012; Kerfeld et al. 2010). These protein-defined organelles are naturally found in bacteria, e.g., the carboxysome, and are just now being explored as a platform for the engineered localization of enzymes (Cai et al. 2014; Frank et al. 2013). Similar strategies are being pursued with capsid and designed protein complexes (Wörsdörfer et al. 2011; Lai et al. 2014; Ni and Chau 2014).

10.3 Xeno-nucleic Acids

Instead of changing the order in which natural building blocks are put together to make an artificial polymer, the building blocks themselves can be changed. With nucleic acids, the changes can be made with the nucleobase or ribose sugar units (Fig. 10.3). Several changes have been attempted and have given rise to both polymers that are incapable of supporting heredity and those that show genetic activity in vitro and in vivo. One of the better known differences between G–C and

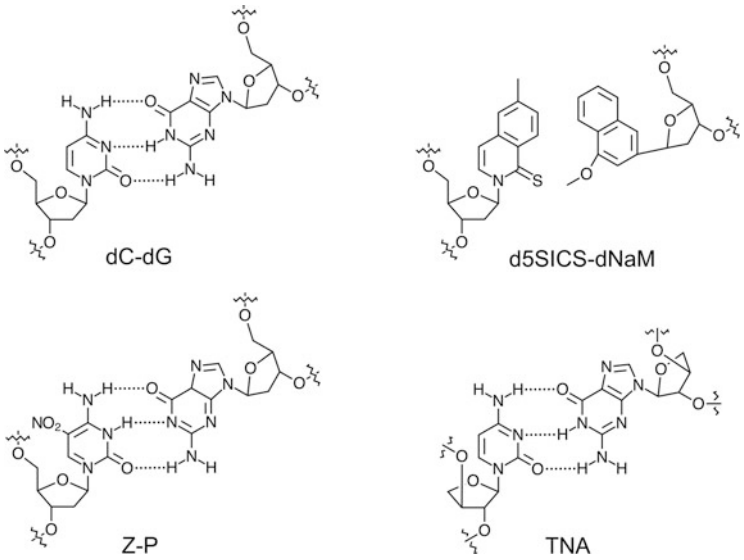


Fig. 10.3 Xeno-nucleic acids. The dG–dC base pair is shown for comparison. The remaining base pairs illustrate interactions not naturally found in cellular life, including an alternate hydrogen-bonding pattern (Z-P), a base pair lacking hydrogen bonds (d5SICS-dNaM), and sugar-substituted nucleotides (TNA)

A–T base pairs is the difference in hydrogen-bonding pattern between the two pairing interactions. New artificial heterocycles have been developed that retain the purine–pyrimidine geometry but with altered hydrogen-bonding patterns so that the existing molecular components of nucleic acid polymers can be expanded (Benner 2004a). For example, high-fidelity replication via PCR can be accomplished with nucleic acids containing six different base pairs exploiting twelve different nucleotides (Yang et al. 2010). In addition to polymerases, this expanded set of nucleotides are substrates for ligases, kinases, and remarkably even the ribosome (Benner 2004a; Bain et al. 1992; Leal et al. 2014). Since the artificial nucleotides do not base pair with natural nucleotides, new codons can be generated without altering the existing genetic code. Although the expanded set of nucleotides based on altered hydrogen-bonding patterns have been quite successful for technological applications, including the generation of aptamers (Sefah et al. 2014; Kimoto et al. 2013), these systems have not been implemented inside of living cells.

Xeno-nucleotides with heterocycles that completely lack hydrogen bond donors or acceptors are still capable of pairing via shape complementarity and packing interactions (Krueger and Kool 2009; Dhami et al. 2014). These xeno-nucleotides with hydrophobic heterocycles do not typically base pair with natural nucleotides and can be enzymatically replicated and transcribed into RNA. Interestingly, variants synthesized by Romesberg's group have shown activity in vivo. Much engineering was needed to get an organism to exploit and depend upon the artificial, hydrophobic base pair. First, to ensure the uptake of the xeno-nucleotides, *E. coli*

was made to express an algal nucleotide triphosphate transporter that showed sufficiently broad substrate specificity. Previous work showed that the unnatural base pair could be replicated by DNA polymerase I and not by DNA polymerase III, the enzyme that replicates the majority of the genome (Seo et al. 2009; Li et al. 2014). Nevertheless, by clever placement of the artificial base pair near the origin of replication of a plasmid prepared in vitro, the plasmid was replicated in *E. coli* in a manner dependent upon the xeno-nucleotides (Malyshev et al. 2014). The placement of the unnatural base pair near the origin of replication increased the chances of being replicated by DNA polymerase I in both the leading and lagging strands. This first example of in vivo-replicated DNA containing an unnatural base pair is an impressive step forward toward the synthesis of organisms fundamentally different from contemporary life. However, these xeno-nucleotides so far have served no function other than to support the replication of a plasmid. To make deeper changes to cell architecture, polymerases capable of robustly copying genomes containing a series of unnatural base pairs would seem to be needed, as would the ability to serve some type of function in a transcript or by encoding an unnatural amino acid for protein synthesis.

Recently natural and engineered polymerases were isolated that can faithfully replicate a class of xeno-nucleotides that contain a sugar molecule other than ribose. However, replication is not through a type of replicase that directly copies both strands of the duplex. Instead, the XNA (xeno-nucleic acid) polymer is transcribed from a DNA template, and the XNA can be reverse transcribed back to DNA. In this way, a replicative cycle is achieved through an intermediate DNA step. This is possible because several XNAs with substituted sugar groups retain the ability to hybridize with DNA and RNA through normal Watson–Crick base pairing. The first example of such a polymer was TNA (threose nucleic acid) in which the five-carbon ribose was substituted with the four-carbon threose (Schöning et al. 2000). Not only can TNA be robustly produced and replicated enzymatically (Chaput et al. 2003; Ichida et al. 2005; Pinheiro et al. 2012), the molecule can also fold into functionally active aptamers capable of specifically recognizing ligands (Yu et al. 2012). Another XNA with a sugar-modified backbone, that of HNA (1,5-anhydrohexitol nucleic acid), can similarly give rise to in vitro evolved heredity and function (Pinheiro et al. 2012). Since several sugar-substituted XNAs are able to base pair with DNA and RNA and are resistant to cellular nucleases, XNAs may find use in biotechnology. Further, since XNAs can also possess catalytic activity (Taylor et al. 2014), it is conceivable that genes encoding XNA polymers could confer a measurable phenotype to living cells. It is presently unclear if XNA could direct the synthesis of proteins. It should be noted that not all sugar-substituted XNAs are capable of base pairing with natural nucleic acids. For example, XNAs with α-arabinopyranosyl units form highly stable duplex helical structures through base-pairing interactions but do not hybridize with DNA or RNA (Beier et al. 1999).

Much progress has been made in building varying types of XNA molecules. Alternate hydrogen-bonding heterocycles and even those completely lacking the ability to hydrogen bond can be exploited for the construction of XNAs. Similarly,

the sugar unit can be exchanged while still retaining the ability to function as a genetic polymer. It will be interesting to see how far such strategies can be pushed. Presumably the further from extant nucleic acids the XNAs are made to be, the harder it will be to identify or engineer polymerases capable of copying these XNAs. One region of the nucleic acid structure that cannot be easily manipulated is the phosphate of the phosphodiester backbone. If nucleic acid analogues are expected to function in water, then the charge density provided by phosphate appears to be ideally suited to maintain the overall structure of the molecule (Benner 2004b). Further, polymerases able to synthesize polymers of XNAs with a non-phosphate-containing backbone would likely require drastically different active sites.

10.4 Xeno-amino Acids

The simplest way to make proteins inside of a living cell with unnatural amino acids is to use an auxotrophic strain and feed in a close structural homologue of the amino acid that the organism is incapable of synthesizing (Liu and Schultz 2010). In this way, the natural amino acid can be substituted with the unnatural amino acid throughout the entire organism. To expand the amino acids available for protein synthesis rather than to substitute an unnatural amino acid for a natural amino acid, some form of genetic reprogramming is necessary. This is done most frequently by the recoding of a stop codon, typically the amber UAG, in a manner that is decoded by an orthogonal aminoacyl-tRNA synthetase–tRNA pair. In order to have the unnatural amino acid incorporated in response to only the desired codon, it is important that the tRNA is not recognized by the host translation machinery and that the orthogonal aminoacyl-tRNA synthetase does not aminoacylate natural host tRNA. This is usually accomplished by taking an aminoacyl-tRNA synthetase–tRNA pair from an organism from a different domain of life and then changing the anticodon sequence to match the desired codon. As described for the auxotrophic strain method, the charging of the tRNA with the unnatural amino acid can sometimes be achieved through the promiscuous activity of the aminoacyl-tRNA synthetase. However, oftentimes rounds of positive and negative selections applied to both the tRNA and the aminoacyl-tRNA synthetase are necessary to build an efficient and orthogonal system (Liu and Schultz 2010). Although there are many factors that influence the overall efficiency of genetically encoded unnatural amino acid incorporation, this methodology has a long track record of success. Difficulties persist, however, in incorporating several different unnatural amino acids in a protein at once inside of a living cell. In fact, the upper limit thus far is two different unnatural amino acids in one protein (Wan et al. 2010; Chatterjee et al. 2013). Nevertheless, this work has laid the foundation for genome-wide reprogramming. For example, it is now possible to substitute all of the UAG stop codons in the *E. coli* genome with UAA and thus to delete the genes encoding release factor 1 (Lajoie et al. 2013). By then placing UAG codons at specific positions and adding an orthogonal aminoacyl-tRNA synthetase–tRNA pair into this recoded *E. coli*, the

recoded organism can specifically incorporate unnatural amino acids at precise locations. More extensive recoding could conceivably yield an organism resistant to bacteriophage infection. Since the genetic code of the bacteriophage would be different from the recoded bacterium, the bacteriophage genome would not properly express in the intended host (Lajoie et al. 2013).

To more greatly expand the space in which unnatural amino acids could be incorporated, a quadruplet decoding system has been developed. Three-nucleotide codons allow for 64 different compositions of which only three (the three stop codons) have been exploited for the introduction of unnatural amino acids. A quadruplet code allows for 256 possibilities thereby greatly increasing the opportunities for expanding the amino acid pool from which to support protein function. Translation of the quadruplet codon exploits a tRNA designed to contain an expanded anticodon similar to that seen for natural frameshift suppressors (Anderson et al. 2004). The same needs for orthogonality of tRNA and aminoacyl-tRNA synthetases apply here as for the introduction of unnatural amino acids through the suppression of stop codons. However, the efficiency of the quadruplet system is decreased in comparison to translation with normal three-nucleotide codons, which may, in part, be due to competition with tRNA molecules that recognize the first three nucleotides of the quadruplet codon. In fact, the efficiency of UAGN quadruplet codons was increased when used in the recoded *E. coli* described above, because the release factor that recognizes UAG was deleted (Chatterjee et al. 2014). That is, there was less competition with natural cellular components for the processing of the UAGN sequence. An alternative strategy to increase the efficiency of quadruplet codons is to exploit mutant, orthogonal ribosomes that were previously selected for improved processivity (Neumann et al. 2010).

There are two general types of orthogonal ribosomes. The first simply contains an altered anti-Shine–Dalgarno sequence at the 3′ end of the 16S rRNA and thus recognizes a different sequence within the mRNA (Rackham and Chin 2005). In other words, these orthogonal ribosomes do not bind to or translate normal, cellular mRNA containing the Shine–Dalgarno sequence (ribosome binding site) but rather specifically recognize and translate mRNA with a different sequence. This is quite useful because now unnatural amino acids can be processed by orthogonal mRNA–ribosome pairs and thus suffer less from competition with natural resources. In fact, the original orthogonal ribosome was later used to generate versions optimized for the incorporation of unnatural amino acids using three-nucleotide (Wang et al. 2007) and four-nucleotide codons (Neumann et al. 2010). The other type of orthogonal ribosome is based on interactions between tRNA molecules and the ribosome as opposed to mRNA–ribosome interactions (Terasaka et al. 2014). Specific base-pairing interactions occur between the conserved 3′ end of each tRNA and 23S rRNA within the peptidyl transferase center of the ribosome. Mutations that disrupt this interaction block translation but compensatory mutations that reestablish base pairing can restore protein synthesis (Terasaka et al. 2014; Kim and Green 1999). In this case, the orthogonal ribosome cannot use natural tRNAs and instead can only exploit orthogonal tRNA molecules. Thus far, the orthogonal

ribosome–tRNA pairs have only been exploited in vitro where they were coupled with another useful innovation. Rather than go through the difficult process of identifying compatible aminoacyl-tRNA synthetases and tRNA molecules for the incorporation of unnatural amino acids, Suga and colleagues exploit a series of ribozymes that they previously selected that are capable of aminoacylating tRNAs with activated amino acids in a manner largely independent of the tRNA and the unnatural amino acid (Passioura and Suga 2013). Impressively, a single mRNA can be used to produce multiple proteins simply by controlling the aminoacylation of the tRNA molecules (Terasaka et al. 2014). However, it should be noted that the possibilities are much greater in vitro where competition with cellular machinery is nonexistent. Peptides with much greater degrees of unnatural amino acids can be generated in vitro (Schlippe et al. 2012).

10.5 Artificial Cells

Truly xenobiotic life would be built from the ground up entirely with components not found in contemporary biology. Although our understanding of cellular life is too incomplete to succeed in such a task now (Forlin et al. 2012), the process of building increasingly improved cellular mimics with biological and abiological components will likely reveal much about the foundations of cellular function that are missed by studies that depend upon the manipulation of living cells. Many synthetic cell studies are geared toward gaining insight into the origins of life and exploit molecules deemed to be prebiotically plausible (Del Bianco and Mansy 2012; Blain and Szostak 2014; Deamer 1997; Luisi 2003). In this section we consider artificial cells that are meant to mimic extant rather than primitive cellular life. Considering the difficulties there are for building genetically encoded artificial cells, it is clearly easier to build technologies within living cells. However, there are some distinct advantages for using artificial cells. Technologies built with artificial cells can incorporate the features of life that are desirable while intentionally leaving out the features of life that can cause complications. For example, if an engineered cell were to be used to diagnose or eradicate a disease state (Kotula et al. 2014; Gupta et al. 2013a; Hwang et al. 2014), then artificial cells incapable of replicating would be more desirable than a living cell which would evolve and proliferate (Lentini et al. 2014). Artificial cells are also easier to engineer since there are less chances for complications arising from interfering interactions between the stripped down components of an artificial cell. Similarly, it could be easier to build artificial cells with orthogonal molecules since competitions with host molecules are removed. For example, the incorporation of unnatural amino acids into proteins faces obstacles associated with the effect of the unnatural amino acid on all of the protein-coding sequences throughout the organism. Although such complications can be ameliorated by modifying the entire genome of the organism (Lajoie et al. 2013) or by exploiting orthogonal ribosomes (Rackham and Chin 2005), the problem largely does not exist with artificial cells since there are no leftover “host” molecules to deal with.

The expression of genes in an artificial cell is mediated by transcription–translation machinery that is not genetically encoded within the artificial cell. Typically, such systems are either based on *E. coli* cell extracts (Spirin and Swartz 2007; Jewett et al. 2008; Shin and Noireaux 2012) or reconstituted translation machinery (i.e., the PURE system) (Shimizu et al. 2001). The RNA polymerase can either be from bacteriophage, such as T7, or bacterial (e.g., *E. coli*). T7 RNA polymerase has the advantage of being much simpler and capable of generating large amounts of RNA and consequently protein. However, transcription–translation is not coupled, since T7 RNA polymerase synthesizes RNA much faster than *E. coli* ribosomes can synthesize proteins (Spirin and Swartz 2007). Although inherently more difficult to work with, the complexity of *E. coli* RNA polymerase does allow for further means of gene regulation through the activity of sigma factors (Shin and Noireaux 2012). For the most part, the behavior of genetic constructs in vitro is the same as that observed in vivo with the available transcription–translation systems (Chappell et al. 2013). The spacing requirements and the influences of sequence context on open reading frames are similar (Lentini et al. 2013; Chizzolini et al. 2014; Karig et al. 2012), even if differences in transcriptional terminator strength (Du et al. 2009) are observed. Interestingly, synthetic, codon-optimized genes can behave differently from natural coding sequences (Lentini et al. 2014; Whitaker et al. 2014). The main problem faced with gene expression in artificial cells is the longevity of the system (Chizzolini et al. 2014; Stögbauer et al. 2012). Protein production typically halts after a couple of hours, although artificial cells containing an *E. coli* cell extract can be made to last longer by the feeding in of nutrients through pore proteins (Noireaux and Libchaber 2004). Nevertheless, as the complexity of genetic circuits grow, it will be important to build better energy-regenerating systems.

To make an artificial cell useful, it must be able to integrate itself within a population of cells. This would seem to require that the artificial cells are able to control gene expression in response to chemical stimuli. Unfortunately, very few examples of gene expression controlled by small molecules have been demonstrated in vitro (Shin and Noireaux 2012; Chappell et al. 2013; Karig et al. 2012; Martini and Mansy 2011; Noireaux et al. 2003) and even fewer in artificial cells. Thus far, artificial cells have only been engineered to respond to theophylline (Lentini et al. 2014; Martini and Mansy 2011), IPTG (Kobori et al. 2013), and arabinose (Shin and Noireaux 2012). But even with this small set of available parts, potential applications have emerged. Lentini et al. built an artificial cell that was able to chemically communicate with *E. coli* (Lentini et al. 2014). More importantly, the artificial cell manipulated the behavior of *E. coli* in response to a chemical message that *E. coli* cannot naturally sense. This was accomplished with artificial cells that expressed a pore protein in response to theophylline, a membrane-permeable molecule that does not induce a response in *E. coli* (Fig. 10.4). Since the artificial cell contained an impermeable signaling molecule (IPTG), the chemical message was only sent to *E. coli* once theophylline was present (Lentini et al. 2014). To improve the utility of the artificial cell, it will

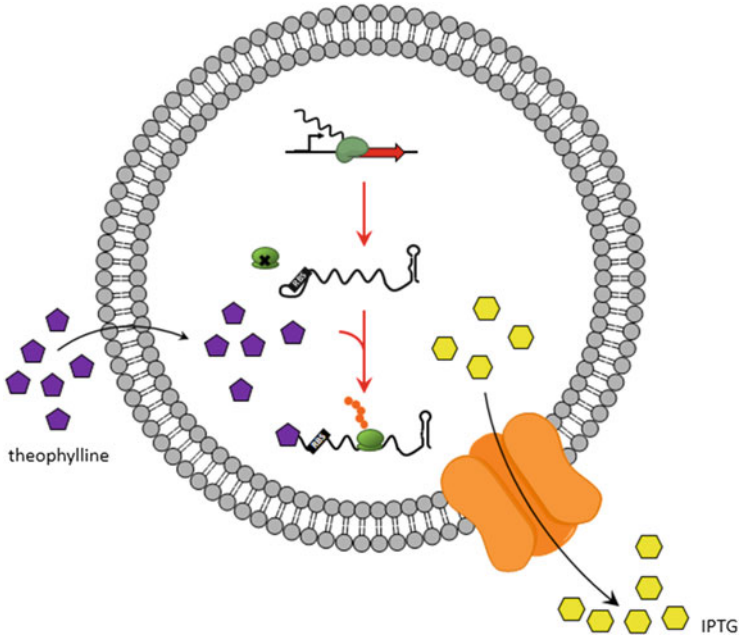


Fig. 10.4 A schematic showing an artificial cell that senses theophylline through the activity of a riboswitch and in response releases entrapped IPTG through an expressed pore protein

be necessary to incorporate sensing mechanisms that regulate gene expression in response to small molecules secreted by natural cells.

To get life-like function it is not necessary to make everything look like a living cell. Although living cells are enclosed within a membrane largely composed of lipids and proteins, nonprotein components, such as DNA origami pores, can be used to facilitate the diffusion of solutes (Langecker et al. 2012). Further, it is not even necessary to use a lipid membrane at all. Chitosan–alginate compartments can encapsulate purified enzymes that synthesize and release small molecules to bacteria (Gupta et al. 2013b). Although transcription–translation was not carried out in this study, gene expression is possible within calcium alginate particles (Kwon et al. 2008). Others have made non-lipid compartments composed of proteins that are selectively permeable and support gene expression with an *E. coli* cell extract (Huang et al. 2013). Aside from the molecules that define the compartment, internal organization of artificial cells can be made to mimic living cells by incorporating mixtures of aqueous polymer solutions (Keating 2012). Protein partitioning between the different aqueous phases is dependent upon the structural properties of the protein so that a protein can be made to switch phases through conformation transitions. In other words, localization does not require structural RNA or protein scaffolds but rather can rely on aqueous phase-separated systems. Toward this end it has been shown that aqueous two-phase systems inside of artificial cells support gene expression and that the expressed fluorescent protein preferentially partitioned

to one of the two aqueous phases (Torre et al. 2014). More work will be needed to decipher how to encode such protein partitioning so that the metabolic flux of the artificial cell could be controlled, in part, by where the proteins reside within the synthetic cytosol.

10.6 Conclusion

Xenobiotic life has progressed greatly. Abiotic nucleic acid and protein polymers are being produced, and some engineered organisms depend on unnatural component parts for survival. However, the field is still very much in its infancy. An organism that depends on a single unnatural base pair that does not encode any particular function leaves much room for further developments. Similarly, despite the fact that over 100 unnatural amino acids can be incorporated into living cells, only two different unnatural amino acids at a time can be put into a single protein. Again, that is far from a fully unnatural polymer or from a greatly expanded code. Artificial cell research has only now begun to try to incorporate function beyond the simple expression of cascades of fluorescent proteins. Artificial cells can now communicate with living cells and sense the environment, but their utility will be greatly decreased unless artificial cells can be made to chemically communicate in response to the presence of natural cells. Perhaps the most interesting advance would be if some of these different xenobiotic avenues of research were combined so that unnatural nucleic acids were used to direct the synthesis of unnatural proteins inside of an artificial cell housed within a compartment defined by a non-lipid membrane.

References

Anderson JC et al (2004) An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci USA 101:7566–7571. doi:10.1073/pnas.0401517101

Annaluru N et al (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344:55–58. doi:10.1126/science.1249252

Bain JD, Switzer C, Chamberlin AR, Benner SA (1992) Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. Nature 356:537–539. doi:10.1038/356537a0

Beier M, Reck F, Wagner T, Krishnamurthy R, Eschenmoser A (1999) Chemical etiology of nucleic acid structure: comparing pentopyranosyl-(2′→4′) oligonucleotides with RNA. Science 283:699–703

Beisel CL, Smolke CD (2009) Design principles for riboswitch function. PLoS Comput Biol 5: e1000363. doi:10.1371/journal.pcbi.1000363

Benner SA (2004a) Understanding nucleic acids using synthetic chemistry. Acc Chem Res 37:784–797. doi:10.1021/ar040004z

Benner SA (2004b) Chemistry. Redesigning genetics. Science 306:625–626. doi:10.1126/science.1101104

Bikard D et al (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437. doi:10.1093/nar/gkt520

Blain JC, Szostak JW (2014) Progress toward synthetic cells. Annu Rev Biochem 83:615–640. doi:10.1146/annurev-biochem-080411-124036

Boch J et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512. doi:10.1126/science.1178811

Brantl S, Wagner EG (2000) Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. Mol Microbiol 35:1469–1482

Breaker RR, Joyce GF (2014) The expanding view of RNA and DNA function. Chem Biol 21:1059–1065. doi:10.1016/j.chembiol.2014.07.008

Bruegmann T et al (2010) Optogenetic control of heart muscle in vitro and in vivo. Nat Methods 7:897–900. doi:10.1038/nmeth.1512

Cai F, Sutter M, Bernstein SL, Kinney JN, Kerfeld CA (2014) Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. ACS Synth Biol. doi:10.1021/sb500226j

Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. Science 334:1716–1719. doi:10.1126/science.1212209

Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297:1016–1018. doi:10.1126/science.1072266

Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. Mol Syst Biol 1:2005.0018. doi:10.1038/msb4100025

Chappell J, Jensen K, Freemont PS (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. Nucleic Acids Res 41:3471–3481. doi:10.1093/nar/gkt052

Chaput JC, Ichida JK, Szostak JW (2003) DNA polymerase-mediated DNA synthesis on a TNA template. J Am Chem Soc 125:856–857. doi:10.1021/ja028589k

Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. Biochemistry 52:1828–1837. doi:10.1021/bi4000244

Chatterjee A, Lajoie MJ, Xiao H, Church GM, Schultz PG (2014) A bacterial strain with a unique quadruplet codon specifying non-native amino acids. Chembiochem 15:1782–1786. doi:10.1002/cbic.201402104

Chen AH, Silver PA (2012) Designing biological compartmentalization. Trends Cell Biol 22:662–670. doi:10.1016/j.tcb.2012.07.002

Chizzolini F, Forlin M, Cecchi D, Mansy SS (2014) Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. ACS Synth Biol 3:363–371. doi:10.1021/sb4000977

Chow BY et al (2010) High-performance genetically targetable optical neural silencing by light-driven proton pumps. Nature 463:98–102. doi:10.1038/nature08652

Christian M et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761. doi:10.1534/genetics.110.120717

Copeland MF, Politz MC, Pfleger BF (2014) Application of TALEs. CRISPR/Cas and sRNAs as trans-acting regulators in prokaryotes. Curr Opin Biotechnol 29:46–54. doi:10.1016/j.copbio.2014.02.010

Deamer DW (1997) The first living systems: a bioenergetic perspective. Microbiol Mol Biol Rev 61:239–261

Del Bianco C, Mansy SS (2012) Nonreplicating protocells. Acc Chem Res 45:2125–2130. doi:10.1021/ar300097w

Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333:470–474. doi:10.1126/science.1206938

Dhami K et al (2014) Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. Nucleic Acids Res 42:10235–10244. doi:10.1093/nar/gku715
Du L, Gao R, Forster AC (2009) Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase. Biotechnol Bioeng 104:1189–1196. doi:10.1002/bit.22491
Dueber JE et al (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27:753–759. doi:10.1038/nbt.1557
Edge MD et al (1981) Total synthesis of a human leukocyte interferon gene. Nature 292:756–762
Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822. doi:10.1038/346818a0
Famulok M, Mayer G (2014) Aptamers and SELEX in chemistry and biology. Chem Biol 21:1055–1058. doi:10.1016/j.chembiol.2014.08.003
Folcher M et al (2014) Mind-controlled transgene expression by a wireless-powered optogenetic designer cell implant. Nat Commun 5:5392. doi:10.1038/ncomms6392
Forlin M, Lentini R, Mansy SS (2012) Cellular imitations. Curr Opin Chem Biol 16:586–592. doi:10.1016/j.cbpa.2012.10.020
Frank S, Lawrence AD, Prentice MB, Warren MJ (2013) Bacterial microcompartments moving into a synthetic biological world. J Biotechnol 163:273–279. doi:10.1016/j.jbiotec.2012.09.002
Gibson DG et al (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319:1215–1220. doi:10.1126/science.1151721
Gilham PT, Khorana HG (1958) Studies on polynucleotides. I. A new and general method for the chemical synthesis of the C5′-C3′ internucleotidic linkage. Syntheses of deoxyribodinucleotides. J Am Chem Soc 80(23):6212
Glass JI et al (2006) Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425–430. doi:10.1073/pnas.0510013103
Good MC, Zalatan JG, Lim WA (2011) Scaffold proteins: hubs for controlling the flow of cellular information. Science 332:680–686. doi:10.1126/science.1198701
Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold switches: de-novo-designed regulators of gene expression. Cell 159(4):925–939. doi:10.1016/j.cell.2014.10.002
Gupta S, Bram EE, Weiss R (2013a) Genetically programmable pathogen sense and destroy. ACS Synth Biol 2:715–723. doi:10.1021/sb4000417
Gupta A et al (2013b) Encapsulated fusion protein confers “sense and respond” activity to chitosan-alginate capsules to manipulate bacterial quorum sensing. Biotechnol Bioeng 110:552–562. doi:10.1002/bit.24711
Huang X et al (2013) Interfacial assembly of protein-polymer nano-conjugates into stimulus-responsive biomimetic protocells. Nat Commun 4:2239. doi:10.1038/ncomms3239
Hwang IY et al (2014) Reprogramming microbes to be pathogen-seeking killers. ACS Synth Biol 3:228–237. doi:10.1021/sb400077j
Ichida JK, Horhota A, Zou K, McLaughlin LW, Szostak JW (2005) High fidelity TNA synthesis by Therminator polymerase. Nucleic Acids Res 33:5219–5225. doi:10.1093/nar/gki840
Isaacs FJ et al (2004) Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22:841–847. doi:10.1038/nbt986
Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. Nat Biotechnol 24:545–554. doi:10.1038/nbt1208
Itakura K et al (1977) Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science 198:1056–1063
Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. Proc Natl Acad Sci USA 102:15971–15976. doi:10.1073/pnas.0503868102

Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4:220. doi:10.1038/msb.2008.57

Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14:49–55. doi:10.1038/nrm3486

Karig DK, Iyer S, Simpson ML, Doktycz MJ (2012) Expression optimization and synthetic gene networks in cell-free systems. Nucleic Acids Res 40:3763–3774. doi:10.1093/nar/gkr1191

Keating CD (2012) Aqueous phase separation as a possible route to compartmentalization of biological molecules. Acc Chem Res 45:2114–2124. doi:10.1021/ar200294y

Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. Annu Rev Microbiol 64:391–408. doi:10.1146/annurev.micro.112408.134211

Khorana HG (1965) Polynucleotide synthesis and the genetic code. Fed Proc 24:1473–1487

Khorana HG et al (1972) Studies on polynucleotides. 103. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. J Mol Biol 72:209–217

Kim DF, Green R (1999) Base-pairing between 23S rRNA and tRNA in the ribosomal A site. Mol Cell 4:859–864

Kimoto M, Yamashige R, Matsunaga K, Yokoyama S, Hirao I (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet. Nat Biotechnol 31:453–457. doi:10.1038/nbt.2556

Kobori S, Ichihashi N, Kazuta Y, Yomo T (2013) A controllable gene expression system in liposomes that includes a positive feedback loop. Mol Biosyst 9:1282–1285. doi:10.1039/c3mb70032a

Konermann S et al (2013) Optical control of mammalian endogenous transcription and epigenetic states. Nature 500:472–476. doi:10.1038/nature12466

Kotula JW et al (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. Proc Natl Acad Sci USA 111:4838–4843. doi:10.1073/pnas.1321321111

Krueger AT, Kool ET (2009) Redesigning the architecture of the base pair: toward biochemical and biological function of new genetic sets. Chem Biol 16:242–248. doi:10.1016/j.chembiol.2008.12.004

Kwon YC, Hahn GH, Huh KM, Kim DM (2008) Synthesis of functional proteins using Escherichia coli extract entrapped in calcium alginate microbeads. Anal Biochem 373:192–196. doi:10.1016/j.ab.2007.10.045

Lai YT et al (2014) Structure of a designed protein cage that self-assembles into a highly porous cube. Nat Chem 6:1065–1071. doi:10.1038/nchem.2107

Lajoie MJ et al (2013) Genomically recoded organisms expand biological functions. Science 342:357–360. doi:10.1126/science.1241459

Langecker M et al (2012) Synthetic lipid membrane channels formed by designed DNA nanostructures. Science 338:932–936. doi:10.1126/science.1225624

Leal NA et al (2014) Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. ACS Synth Biol. doi:10.1021/sb500268n

Lentini R et al (2013) Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. ACS Synth Biol 2:482–489. doi:10.1021/sb400003y

Lentini R et al (2014) Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour. Nat Commun 5:4012. doi:10.1038/ncomms5012

Levskaya A et al (2005) Synthetic biology: engineering Escherichia coli to see light. Nature 438:441–442. doi:10.1038/nature04405

Li L et al (2014) Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. J Am Chem Soc 136:826–829. doi:10.1021/ja408814g

Liang JC, Bloom RJ, Smolke CD (2011) Engineering biological systems with synthetic RNA molecules. Mol Cell 43:915–926. doi:10.1016/j.molcel.2011.08.023

Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444. doi:10.1146/annurev.biochem.052308.105824

Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci USA 108:8617–8622. doi:10.1073/pnas.1015741108
Luisi PL (2003) Autopoiesis: a review and a reappraisal. Naturwissenschaften 90:49–59. doi:10.1007/s00114-002-0389-9
Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nat Methods 10:957–963. doi:10.1038/nmeth.2649
Malyshev DA et al (2014) A semi-synthetic organism with an expanded genetic alphabet. Nature 509:385–388. doi:10.1038/nature13314
Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11:181–190. doi:10.1038/nrg2749
Martini L, Mansy SS (2011) Cell-like systems with riboswitch controlled gene expression. Chem Commun (Camb) 47:10734–10736. doi:10.1039/c1cc13930d
Meinhardt S et al (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. Nucleic Acids Res 40:11139–11154. doi:10.1093/nar/gks806
Michelson AM, Todd AR (1955) Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′: 5′-internucleotidic linkage. J Chem Soc 2632–2638. doi:10.1039/JR9550002632
Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501. doi:10.1126/science.1178817
Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464:441–444. doi:10.1038/nature08817
Ni R, Chau Y (2014) Structural mimics of viruses through peptide/DNA co-assembly. J Am Chem Soc. doi:10.1021/ja507833x
Nielsen AA, Voigt CA (2014) Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. Mol Syst Biol 10:763
Nihongaki Y, Yamamoto S, Kawano F, Suzuki H, Sato M (2015) CRISPR-Cas9-based photoactivatable transcription system. Chem Biol. doi:10.1016/j.chembiol.2014.12.011
Noireaux V, Libchaber A (2004) A vesicle bioreactor as a step toward an artificial cell assembly. Proc Natl Acad Sci USA 101:17669–17674. doi:10.1073/pnas.0408236101
Noireaux V, Bar-Ziv R, Libchaber A (2003) Principles of cell-free genetic circuit assembly. Proc Natl Acad Sci USA 100:12672–12677. doi:10.1073/pnas.2135496100
Nomura Y, Yokobayashi Y (2007) Reengineering a natural riboswitch by dual genetic selection. J Am Chem Soc 129:13814–13815. doi:10.1021/ja076298b
Ogawa A (2011) Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. RNA 17:478–488. doi:10.1261/rna.2433111
Passioura T, Suga H (2013) Flexizyme-mediated genetic reprogramming as a tool for noncanonical peptide synthesis and drug discovery. Chemistry 19:6530–6536. doi:10.1002/chem.201300247
Pinheiro VB et al (2012) Synthetic genetic polymers capable of heredity and evolution. Science 336:341–344. doi:10.1126/science.1217622
Politz MC, Copeland MF, Pfleger BF (2013) Artificial repressors for controlling gene expression in bacteria. Chem Commun (Camb) 49:4325–4327. doi:10.1039/c2cc37107c
Polstein LR, Gersbach CA (2012) Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. J Am Chem Soc 134:16480–16483. doi:10.1021/ja3065667
Polstein LR, Gersbach CA (2014) Light-inducible gene regulation with engineered zinc finger proteins. Methods Mol Biol 1148:89–107. doi:10.1007/978-1-4939-0470-9_7
Polstein LR, Gersbach CA (2015) A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat Chem Biol. doi:10.1038/nchembio.1753

Qi LS et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi:10.1016/j.cell.2013.02.022
Rackham O, Chin JW (2005) A network of orthogonal ribosome x mRNA pairs. Nat Chem Biol 1:159–166. doi:10.1038/nchembio719
Robertson DL, Joyce GF (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. Nature 344:467–468. doi:10.1038/344467a0
Rodrigo G, Landrain TE, Jaramillo A (2012) De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. Proc Natl Acad Sci USA 109:15271–15276. doi:10.1073/pnas.1203831109
Rothemund PW (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440:297–302. doi:10.1038/nature04586
Sachdeva G, Garg A, Godding D, Way JC, Silver PA (2014) In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Res 42:9493–9503. doi:10.1093/nar/gku617
Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27:946–950. doi:10.1038/nbt.1568
Schlippe YV, Hartman MC, Josephson K, Szostak JW (2012) In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. J Am Chem Soc 134:10469–10477. doi:10.1021/ja301017y
Schöning K et al (2000) Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3′→2′) oligonucleotide system. Science 290:1347–1351
Seeman NC (1982) Nucleic acid junctions and lattices. J Theor Biol 99:237–247
Sefah K et al (2014) In vitro selection with artificial expanded genetic information systems. Proc Natl Acad Sci USA 111:1449–1454. doi:10.1073/pnas.1311778111
Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE (2009) Optimization of an unnatural base pair toward natural-like replication. J Am Chem Soc 131:3246–3252. doi:10.1021/ja807853m
Shimizu Y et al (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755. doi:10.1038/90802
Shin J, Noireaux V (2012) An E. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells. ACS Synth Biol 1:29–41. doi:10.1021/sb200016s
Smith HO, Hutchison CA, Pfannkoch C, Venter JC (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci USA 100:15440–15445. doi:10.1073/pnas.2237126100
Sparkman-Yager D, Correa-Rojas RA, Carothers JM (2015) Kinetic folding design of aptazyme-regulated expression devices as riboswitches for metabolic engineering. Methods Enzymol 550:321–340. doi:10.1016/bs.mie.2014.10.038
Spirin AS, Swartz JR (2007) Cell-free protein synthesis systems: historical landmarks, classification, and general methods. In: Cell-free protein synthesis: methods and protocols. Wiley, New York. doi:10.1002/9783527622702.ch1
Stanton BC et al (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. Nat Chem Biol 10:99–105. doi:10.1038/nchembio.1411
Stögbauer T, Windhager L, Zimmer R, Rädler JO (2012) Experiment and mathematical modeling of gene expression dynamics in a cell-free system. Integr Biol (Camb) 4:494–501. doi:10.1039/c2ib00102k
Suess B, Fink B, Berens C, Stentz R, Hillen W (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res 32:1610–1614. doi:10.1093/nar/gkh321
Taylor AI et al (2014) Catalysts from synthetic genetic polymers. Nature. doi:10.1038/nature13982
Terasaka N, Hayashi G, Katoh T, Suga H (2014) An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. Nat Chem Biol 10:555–557. doi:10.1038/nchembio.1549

Topp S, Gallivan JP (2007) Guiding bacteria with small molecules and RNA. J Am Chem Soc 129:6807–6811. doi:10.1021/ja0692480
Torino D, Martini L, Mansy SS (2013) Piecing together cell-like systems. Curr Org Chem 17:1751–1757. doi:10.2174/13852728113179990082
Torre P, Keating CD, Mansy SS (2014) Multiphase water-in-oil emulsion droplets for cell-free transcription-translation. Langmuir 30:5695–5699. doi:10.1021/la404146g
Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
Wan W et al (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. Angew Chem Int Ed Engl 49:3211–3214. doi:10.1002/anie.201000465
Wang K, Neumann H, Peak-Chew SY, Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 25:770–777. doi:10.1038/nbt1314
Wang HH et al (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898. doi:10.1038/nature08187
Whitaker WR, Davis SA, Arkin AP, Dueber JE (2012) Engineering robust control of two-component system phosphotransfer using modular scaffolds. Proc Natl Acad Sci USA 109:18090–18095. doi:10.1073/pnas.1209230109
Whitaker WR, Lee H, Arkin AP, Dueber JE (2014) Avoidance of truncated proteins from unintended ribosome binding sites within heterologous protein coding sequences. ACS Synth Biol. doi:10.1021/sb500003x
Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428:281–286. doi:10.1038/nature02362
Wörsdörfer B, Woycechowsky KJ, Hilvert D (2011) Directed evolution of a protein container. Science 331:589–592. doi:10.1126/science.1199081
Yang Z, Chen F, Chamberlin SG, Benner SA (2010) Expanded genetic alphabets in the polymerase chain reaction. Angew Chem Int Ed Engl 49:177–180. doi:10.1002/anie.200905173
Ye H, Daoud-El Baba M, Peng RW, Fussenegger M (2011) A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. Science 332:1565–1568. doi:10.1126/science.1203535
Yu H, Zhang S, Chaput JC (2012) Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. Nat Chem 4:183–187. doi:10.1038/nchem.1241

Index

A
Ab initio prediction methods, 284
Ac/Ds system, 109
O-Acetylserine synthesis, 147
Acyl homoserine lactone (AHL), 35
Adaptamer-directed gene disruptions, 88
Adaptamers, 86–87
Adaptive laboratory evolution (ALE), 212, 214
Aliphatic polyesters, 313–317
  drawback, 313
  polycaprolactone, 317–319
  polydioxanone, 319
  poly-α-hydroxy esters
    polyglycolides, 314–315
    polylactides, 315–317
    ring-opening polymerization of cyclic lactones, 313, 314
    synthesis, 313
  properties, 312
  with reactive groups, 313
Allele segregation, 6
Amber codon, 184, 185
AMBER force fields, 288
Aminoacylation, quality control of, 155
Aminoacyl-tRNA synthetases (aaRSs), 155
AND gate, 29
Aptamers, 80, 86–87
Aptaswitches, 80
ara promoters, 4
Archaea, selenocysteine, 158
Argonaute proteins, 83
Aromatic ncAAs, 147
Artificial cells, 348–351
  advantages, 348
  E. coli, 349
  gene expression, 349
  impermeable signaling molecule, 349
  internal organization, 350
Artificial piggyBAC transposon, 109
Artificial riboswitches, 340
Artificial RNAs, controlling cellular activity with, 339–341
Artificial transcription factors, 22
  lysine-specific demethylase, 1, 19
  prokaryotic biosensors, 18
  RNA-guided DNA binding with CRISPR technology, 23–24
  TALE (*see* Transcription activator-like effector (TALE))
  zinc finger proteins, 20–22
Aspergillus niger, 237
Attenuator toggle mechanism, 13
AutoDock (Scripps), 286

B
Bacterial microcompartments (BMCs), 272–276
Bacterial promoters, engineering of, 68
Bacterial quorum sensing (QS) system, 223
Bc-sRNAs, 85
BglBrick (BBb) assembly, 122
BioBrick (BB) standard assembly, 120–122
Biochemical computer, 32
Biodegradable biopolymers, 308
Biomaterials, 308
Bioorthogonal conjugation reaction, 180–182
Bioorthogonal noncanonical amino acid tagging (BONCAT), 182, 183
Biopolymers
  biosynthesis, 308
  chemical synthesis, 308
Biosensors, 26–27
Biosyn, 320
Bipedal DNA walkers, 91
Bistable genetic toggle switch, 32

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

Bistable toggle switch, 233
Botrytis cinerea, 85
Bottom-up approach, 49
Buffer gate, 28
Burn-the-bridge methods, 93

C
Caldicellulosiruptor, 241
Canavanine, 151
Canonical amino acids (cAAs)
  L-enantiomers, 145
  and genetic code, 145
  individual side chains, 146
  physicochemical properties and chemical diversity, 145–146
  proline, 146
Canonical ensembles. *See* NVT ensembles
Carbon storage regulator (Csr) system, 221
Cell factory, 212
  cellular stress, 222
  cellular system, integrity of, 222
  construction, at genomic level, 213
    adaptive laboratory evolution, 214
    conjugative assembly genome engineering, 215
    environmental conditions, 214
    global transcription machinery engineering, 215
    low-mutation-rate strains, 214
    MDS42 strain, 214
    multiplex automated genome engineering, 215
    trackable multiplex recombineering, 215
  development, 219
  novel signalling functions, for old molecules, 223–224
  oxidative stress, 222
  quantitative settings, 221
Cell-free protein synthesis (CFPS), 173, 174
Cellular activity
  with artificial RNAs, 339–341
  with engineered proteins, 341–343
Cellular metabolism, 218
Cellulosic biomass, 299
Cellulosome, 270
CHARMM force fields, 288
Chemically inducible chromosomal evolution (CIChE), 6, 7
Chimeric promoters, construction of, 68, 69
Chromosomal integration site, effect of, 10
Chromosomally integrated construct, transcription of, 8
Circular polymerase extension cloning (CPEC), 126
Cis-acting RNA-based regulatory systems, 340
Classic Turbo Yeast, 260
Click chemistry, 181
CLOCKS, 34
Closed-loop glucose controller circuits, 45
Clostridium thermocellum, 241
Combinatorial transcriptional engineering (COMPACTER) strategy, 260
Combined residue- and site-specific ncAA incorporation technique, 198
Comparative approach. *See* Homology modeling
Condition-specific codon optimisation approach, 218
Conjugative assembly genome engineering (CAGE), 215
Copoly(ester amide) synthesis, 321
Core promoter, 11
Corynebacterium glutamicum, 236
CRISPR-associated (*cas*) genes, 23
CRISPR–Cas9 technology
  adaptive immunity in bacteria, 118
  description, 117
  drawback, 119
  targeted genome editing, 118–119
CRISPR interference (CRISPRi), 23
Cyanobacterial photoreceptor, 35
Cys_2His_2 zinc finger, 20
Cytomegalovirus promoters, 5

D
dCas9 protein, 24
De novo approaches. *See* Ab initio prediction methods
Decoy operators, 7–8
Delta-5-acetylornithine synthesis, 147
Dermabond, 324
Desaminotyrosyl-tyrosine alkyl esters, chemical structure, 327
Dicer. *See* Ribonuclease III enzyme
Disposable medical devices, 310
DNA assembly methods, 120, 260
  restriction endonuclease-based cloning methods, 120
  restriction-/ligation-independent assembly strategies, 125–130
  uses, 119
DNA barcodes, 88–89
DNA elements, 66
DNA machines, 89–91
DNA motor, 93
DNA origami technology, 91
DNA pliers, 90
DNA sequencing, 232

DNA walker, 91–94
Dock (UCSF) algorithm, 286
Drug-eluting stents (DES), 309
Dynamic metabolic valve, 266

E
E. coli
cell factory, 212
endogenous trytophan synthase, 193
global transcription machinery engineering, 216
MAGE technology, 215
mevalonate production, 217
o-pair library, positive and negative screening, 187
para-amino-L-phenylalanine biosynthesis, 193
preventing cholera infection, 43
selenocysteine, 158
transcriptomics, 10
Edge detector circuit, 36
Electrospray ionization mass analysis coupled to liquid chromatography (ESI LC-MS), 176
Endogenous regulatory systems, for engineered transcription, 5
Error-prone PCR (ep-PCR), 72
Eukarya, selenocysteine, 158
Eukaryotic promoters, engineering of
galactose-inducible promoter library, 74
random mutagenesis, 74
structure of, 73
TFBS effects, 75, 76
upstream activating/repressing sequences, 74
Yarrowia lipolytica, 74
Expression platform domain, 80
Extracytoplasmic function (ECF) σ-factors, 16
Extremophiles
biofuel production, 242–243
diversity of, 239–240
lignocellulosic crops, 241
M. sedula, 241
proof-of-concept, 241

F
FapR sensor, 48
Feed-forward loops (FFLs), 41
FlexX (BioSolveIT) algorithm, 286
Fluorescence-activated cell sorting (FACS), 13
Fluorescent noncanonical amino acid tagging (FUNCAT), 182
*Fok*I nuclease, 115
FoldX, 296, 299, 300
Free energy, 294
description, 293
FoldX, 296
free energy perturbation, 294
path-dependent methods, 294
path-independent methods, 294
steered molecular dynamics, 294–296
Free energy perturbation (FEP), 294
Frozen accident theory of Crick, 154

G
Gene expression control, with proteins, 341, 342
Gene expression control, with RNAs, 339, 340
Gene synthesis, 232–ENF
Gene therapy, 312
Genetic circuits, 25
biosensors, 26–27
bistable genetic toggle switch, 31
components, 39
clock generators, 33
industrial bioreactor, 42
lambda phage state-switching circuit, synthetic version of, 36
logic gates, 27–30
multicellular genetic circuits, 35
optical edge detection circuit, 35
optimization, 37
parallel network loops, 40–42
quorum sensing, 34
repressilator, 31
serial signal processing parts, 39–40
transfer functions of discrete components, 37–39
volatile and nonvolatile memory, 31–33
Genetic code, 232
Genetic relaxation oscillators, 33
Genetically encoded RNA scaffolds, 341
Genomatica's process, 275
Genome editing strategy, 102
CRISPR–Cas9 technology, 117–119
homing endonucleases, 114
TALE nucleases, 116–117
zinc-finger nucleases, 115–116
Genome engineering, 102, 103, 107, 112, 114
genome editing strategy (*see* Genome editing strategy)
non-programmable (*see* Non-programmable genome engineering)
principles, 106
prokaryotic (*see* Prokaryotic genome engineering)
RNA interference, 110–111

Genome mining, 6
Gibson isothermal assembly, 129
glmS riboswitch, 80
Global transcription machinery engineering (gTME), 16–18, 215
Gluconate, 220
Golden Gate cloning (GG) method, 122
GoldenBraid (GB) method, 123, 124
GROMOS force fields, 289
Group II introns, 112–113
Growth mediated buffering effect, 267
Guide RNA (gRNA), 118

H
HADDOCK methods, 287
Hairpin aptamers, 81
Hammerhead ribozyme, 78, 79
High-throughput screening technology, 37
High-throughput sequencing technologies, 5
Hill equation, 37
Homing endonucleases, 114
Homologous recombination (HR), 102
Homology modeling methods, 284–285
Host-mediated DNA repair, 102
Human Krüppel-associated box (KRAB), 18
Hybrid promoters, 69
Hybrid transcription factor, design of, 18
Hydrolysis of cellulose, 270
Hyperthermophilic archaeon, 243
Hyperuricemia, 45

I
Industrial microbiology, 212
Intergenic adaptamers, 87
Intron-encoded homing nucleases, 114
In vivo DNA assembly techniques, 129–130
Isobaric–isothermal ensemble. *See* NPT ensemble

J
Jarzynski's equality, 295, 296

K
Keio collection, 235

L
lac operon, 4
LacI repressor, concentration of, 10
lac promoters, 4
L-Lactic acid, 315
Lactococcus lactis Ll.LtrB group II intron, 112
Lambda phage state-switching circuit, synthetic version of, 36
Lambda Red recombineering system, 113
Large intergenic noncoding RNAs (lincRNAs), 85
LexA, 43
Ligase cycling reaction (LCR), 128–129
Ligation-independent cloning (LIC), 127–128
Ligation-independent method, for high-throughput TALE assembly, 117
Light-inducible transcriptional effectors (LITEs), 18
Logic gates, 27
  AND gate, 28, 30
  buffer gates, 27
  description, 27
  dynamic range, 27
  NAND, 29
  NOR gate, 29, 30
  NOT gates, 27
Long noncoding RNAs (lncRNAs), 85
Lysine-specific demethylase 1 (LSD1), 19

M
Magnetite cationic liposomes, 44
Malonyl-CoA sensor-actuator system, 264–265
Mariner-type transposon Sleeping Beauty, 109
Membrane-bound hydrogenase (MBH), 242
Memory, 32
Metabolator, 33
Metabolic controls, 220
  codon optimisation strategy, 218
  genetic circuits
    dynamic controls, 220
    on-off methods, 220
  perturbing central regulatory systems, 221
  recombinant productions, 220–221
Metabolic engineering, 218
Metabolic pathway engineering, 260–267, 269, 271, 272
  aim of, 256
  bacterial microcompartments, 272–275
  cells metabolism, 256
  classical methods, 257
  combined approaches, 275
  commercial application, 275–276
  enzymes activity, 257
  genome level, 258
  nucleic acids

bacterial outer membrane vesicles scaffold, 272
plasmid based scaffold, 271, 272
RNA molecule based scaffold, 272
protein level, 269
protein scaffolds, 270
spatial organisation, 269
synthetic biological parts, 258
transcription level
fungal xylose pathway, 260
malonyl-CoA sensor-actuator system, 264–265
multivariate modular metabolic engineering, 262
promoter strength, 260–261
toggle switch, 262–263
xylose utilisation, 261
translational level
dynamic metabolic valve, 266
riboswitch, 267
synthetic ribosome binding sites, 266
Metallosphaera sedula, 241
Methanococcus jannaschii, 240
Michaelis–Menten kinetics, 298
Microbial synthetic platform cells, 231, 239–243
alcohols, 237
biodiesels, 237, 238
chassis cells, 236–237
cutting-edge synthetic biology technologies, 238
extremophiles
diversity of, 239–240
lignocellulosic crops, 241
M. sedula, 241
novel platform cells, biofuel production, 242–243
proof-of-concept, 241
fuels, 230
genome editing, 234–235
minimal genome, 235
synthetic cells (*see* Synthetic cells)
Microcanonical ensemble. *See* NVE ensemble
MicroRNAs (miRNAs), 82, 110, 111
Drosophila, 84
gene silencing, 84
lsy-6 gene functions, 85
pathway, 83
post-transcriptional repressors, 84
transcription factors, 84
Modular Cloning Strategy (MoClo), 123
Molecular dynamics (MD), 291
acetylcholinesterase, 296
basics, 289–290
description, 289
simulation
energy minimization step, 291
generic input files, 291
initialization, 291
NPT ensemble, 291
NVE ensemble, 291
NVT ensemble, 291
time scale limitations, 292
statistical ensemble, 290
Molecular engineering tools, 102
Molecular mechanics, 287–289
Molecular modeling, 283–289
description, 283
docking
AutoDock (Scripps), 286
AutoDock Vina, 286
Dock (UCSF), 286
flexibility problem of receptors, 287
FlexX (BioSolveIT) algorithm, 286
intermolecular biological complex, 3D models, 285
molecular mechanics, 287–289
pose generation, 285
pose selection, 285
scoring step, 286
protein structure prediction
ab initio prediction methods, 284
homology modeling methods, 284–285
limitations, 283
Molecular spiders, 92
Multiplex automated genome engineering (MAGE), 113, 215
Multivariate modular metabolic engineering (MMME), 262
Mycoplasma mycoides, 237

N

N-acyl homoserine lactone, for intercellular communication, 26
Nanomachines
applications, 90
limitations, 90
Natural transcription factors
ligand-binding specificities, 342
programmability, 342
Natural *vs.* artificial riboswitches, 340
Negative autoregulation (NAR), 40
Nicking endonucleases, 128
Nitrile hydratase, 299
Non-biodegradable biopolymers, 308

Noncanonical amino acids (ncAAs), 148, 183. *See also* Residue-specific noncanonical amino acid incorporation
aminoacylation efficiency, 150
antimetabolites, 150
aromatic, 147
biological functions, 150–151
biosynthesis, 147
chemical diversity, 147
heterocycles, 147
incorporation in in vitro expression systems, 173–174
incorporation into proteins, 150
occurrence, 146
site-specific incorporation (*see* Site-specific noncanonical amino acid incorporation)
structures, 147
Noncoding RNAs, 82
Nonhomologous end joining (NHEJ), 102
Non-programmable genome engineering
site-specific recombinases, 107–109
transposons, 109–110
Non-ribosomal peptides (NRPs), 151–152
Nonsense codons, 184
NOT gate, 28, 38
NPT ensembles, 292
Nucleic acids, 339. *See also* Xeno-nucleic acids
NVE ensemble, 291
NVT ensembles, 292
Nylons, 236

O
(2-Octyl cyanoacrylate), 324
Oligomerized Pool ENgineering (OPEN), 21
Operator buffers, 8
Optical edge detection circuit, 35
Optical tracing, 312
OptKnock algorithm, 276
Orthogonal factors, 6
Orthopedic devices, 309
Overlap extension PCR (OE PCR), 125
Overlap-based cloning methods, 120
Oxidative stress, 222

P
Path-dependent methods, 294
Path-independent methods, 294
Pharma-protein production, 216
Phi-X174 bacteriophage genome, 338
Phosphorylation, of *Bacillus subtilis*, 218
Phosphoserine (Sep), 160
Photoreceptor, 35
Pichia pastoris, methanol-inducible P_{AOX1} promoter of, 75
PIWI-interacting RNAs (piRNAs), 82, 83
Plant pathogen-mediated promoter activation, 116
Poly-α-hydroxy esters
ring-opening polymerization of cyclic lactones, 313
synthesis, 313
Poly(alkyl cyanoacrylates) (PCA), 324–325
Poly(amino acid), 327
Polyanhydrides, 323
melt condensation, 323
orthopedic applications, 323
Poly(aspartic acid) (PAA), 329
Polycaprolactone, 317–319
ROP synthesis, 313
Polycaprolactone (PCL) modifications, 318
Poly[(carboxyphenoxy propane)-sebacic acid] structure, 323
Polydioxanone (PDS), 319
Poly(ester amide), 320
Polyglycolides
mechanical properties, 314
properties, 314
resorbable suture development, 314
Poly(lactide-*co*-glycolide) (PLGA), 316
Polylactides, 315–317
Poly(L-glutamic acid), 328
Poly(L-lactide) (PLLA), 315
Poly orthoesters, 321–322
Poly(orthoester) I (POE I), 321
Poly(orthoester) II (POE II), 321
Poly(orthoester) III (POE III), 321
(Poly)peptides, 151, 156–160
genetic code, 152–155
non-ribosomal peptides, 151–153
ribosomal translation, 153, 155–156
standard genetic code, 174
changing of, 160
engineering (*see* Residue-specific noncanonical amino acid incorporation)
phosphoserine, 160
pyrrolysine, 159–160
selenocysteine, 157–159
variations, 156, 157
Polyphosphazenes, 325–326
Polyphosphoesters, 326–327
in vitro cytocompatibility, 326
in vivo tissue compatibility, 326
structure, 326
Polysaccharides, 308
Poly(trimethylene carbonate), 319–320

Pose generation, 285
Pose selection, 285
Positive autoregulation (PAR), 41
Product expulsion mechanism, of cellulose-degrading enzyme, 298
Prokaryotic biosensors, 18
Prokaryotic genome engineering
  group II introns, 112–113
  recombineering, 113–114
Prokaryotic promoters, engineering of, 70
  chemical DNA synthesis, 72
  construction of libraries, 70, 71
  lycopene, 72
  mutagenic PCR, 72
  random mutagenesis, 70, 72, 73
  rational engineering strategies, 70
  statistical method, 72
  strong promoters, creation of, 71
  TFBSs, arrangement and type of, 72
  UP elements, 73
Proline, 178
Prolonged overlap extension PCR (POE PCR), 126
Prosthetic networks. *See* Closed-loop glucose controller circuits
Protein engineering, 282. *See also* Molecular modelling
  noncanonical amino acids for, 161
  rational design
    computer assistance, 282
    structure and function complementarity, 282
Protein–protein docking methods, 287
Proteins, 308
  building blocks
    canonical amino acids, 144–146
    noncanonical aminoacids, 146–151
  expression
    optimization, 67
    synthetic control, 66
  folding and stability, manipulation by ncAAs, 178–180
  in-frame stop codons, 186
  motions, 293
  pharma-protein production, 216
  post-translational modifications, 217–218
  post-translational regulation, of level and local concentrations, 217
  scaffolding, 66
Proteome residue-specific noncanonical amino acid incorporation, 182–183
Pseudo-polyamino acids, 327–328
Pyrrolysine (Pyl), 159–160
Pyrrolysyl-tRNA synthetase (PylRS), 189–190

Q
Quorum sensing, 34

R
Random mutagenesis, 70, 72
Rational engineering strategies, 70
Recombinant cell factory, 213
Recombinase addressable data (RAD) module, 32
Recombinase-mediated cassette exchange (RMCE), 108
Recombineering, 113–114
Registry of Standard Biological Parts, 49
Relaxation oscillator, 34
Repeat variable diresidues (RVDs), 116
Reporter-based selection techniques, 13
Repressilator, 233
Residue-specific noncanonical amino acid incorporation, 174
  AaRSs, 174
  electrospray ionization mass analysis coupled to liquid chromatography, 176
  enzyme activity, manipulation of, 177–178
  fluorescence property, manipulation of, 180
  prerequisite, 175
  protein folding and stability, manipulation of, 178–180
  proteome, 182–183
  reactive ncAAS and bioorthogonal conjugation reaction, 180–182
  Seleno-ncAAs for protein structure determination, 176
  supplementation-based incorporation, 175
  into target proteins, 175
  translation quality control, 175
Restriction endonuclease (RE)-based assembly methods, 120
Restriction endonuclease-based cloning methods, 120
Restriction-/ligation-independent assembly strategies
  circular polymerase extension cloning, 126
  Gibson isothermal assembly, 129
  in vivo DNA assembly techniques, 129–130
  ligase cycling reaction, 128–129
  ligation-independent cloning, 127–128
  overlap extension PCR, 125
  sequence- and ligation-independent cloning, 127
Retroactivity, 8

Rewritable recombinase addressable data (RAD) SET/RESET latch, 32
Ribonuclease III enzyme, 83
Ribosomal translation, 155–156
Ribosome
  ribosomal translation, 156
  stalling, 13
Riboswitch, 267
  description, 80
  in DNA sequence, 81
  glmS riboswitch, 80
  principle, 80
  structure, 80
Ribozymes
  catalyzation, 79
  description, 78
  features, 78
  hammerhead ribozyme, 78, 79
  self-cleaving ribozymes, 78
  self-spicing introns, 78
  site-specific RNA self-cleavage, 79
RNA elements, 66
RNAP II, 12
Rosetta methods, 287

S
Saccharomyces cerevisiae
  cell factory, 212
  global transcription machinery engineering, 216
S-adenosyl-L-ethionine (SAE), 150
S-adenosyl-L-methionine (SAM), 150
Seamless ligation cloning extract (SLiCE), 127
Seed sequence, 84
Selection marker recycling, 109
Selenocysteine (Sec), 157–159
Selenocysteine insertion sequence (SECIS) element, 158, 159
Selenoproteins, 157
SELEX procedure, 86
SELEX technique, 13
Self-cleaving ribozymes, 78
Self-spicing introns, 78
Sense codon recoding, 192–193
Sequence- and ligation-independent cloning (SLIC), 127
Sequence-based methods, 296
Serine/threonine protein kinase, 297
Simian virus 40 promoters, 5
Single-cell genome sequencing, 232
Site-specific noncanonical amino acid incorporation, 187–193
  L-N^6-acetyllysine, 197
  3,4-dihydroxy-L-phenylalanine, 197
  enzyme activity, manipulation of, 197
  O-phospho-L-serine, 197
  orthogonal pairs
    from other aaRSs, 190
    barnase gene, 188
    biosynthesis, 193
    fluorescence-activated cell sorting, 188
    generation of, 187
    negative selection step, 187
    orthogonal ribosomes, 190, 191
    positive selection step, 187
    from pyrrolysyl-tRNA synthetase, 189–190
    sense codon recoding, 192–193
    strains with synthetic genetic codes, 191
    from tyrosyl-tRNA synthetase, 188–189
  phage display system, 197
  principle, 183–187
  protein structure, dynamics and functions, 196–197
  reactive ncAAs, 194–196
Site-specific recombinases, 107–109
Small/short interfering RNAs (siRNA), 82, 84, 110, 111
Small RNAs (sRNAs)
  description, 82
  gene expression, regulation of, 85
  genome stability, 85
  in prokaryotes, 81
  processing parameters, 82–84
  recognition sites of, 84
Small-molecule-sensing allosteric ribozymes, 79
Software tools, DNA-binding domains, 19
Splicing by overlap extension PCR (SOE). *See* Overlap extension PCR (OE PCR)
Steered molecular dynamics (SMD), 294–297
Stimuli-responsive polymers, 311
Stop codon suppression (SCS), 184
Streptavidin-biotin-based detection systems, 86
Streptococcus pyogenes, CRISPR/Cas-system, 23
Supplementation-based incorporation (SPI), 175
Suppressor tRNA, 158
Synthetic Biology Open Language, 49
Synthetic biopolymers, 327–328
  aliphatic polyesters, 308, 312–319
  bioimaging, 312
  biomedical applications, 309
    disposable medical devices, 310
    drug-eluting stents, 309
    orthopedic devices, 309
  drug delivery and release control, 311
  gene delivery, 312
  optical tracing, 312
  poly(alkyl cyanoacrylates), 324–325

poly(amino acid), 327
poly(aspartic acid), 328
poly(L-glutamic acid), 328
pseudo-polyamino acids, 327–328
polyanhydrides, 322–324
poly(ester amide), 320
poly(orthoesters), 321–322
polyphosphazenes, 325–326
polyphosphoesters, 308, 326–327
poly(trimethylene carbonate), 319–320
tissue engineering, 310
Synthetic cells
biological parts, 231–232
DNA sequencing, 232
gene synthesis, 232–ENF
genetic circuit, 233–234
Synthetic eukaryotic regulatory domains, 19
Synthetic promoters
classification strategy, 68
engineering, 67
eukaryotic promoters, 73–77
hybrid promoters, 68–70
libraries, 68
metabolic pathways, optimization of, 68
prokaryotic promoters, 70–73
Synthetic small RNAs, 231
Synthetic toggle switch, 31
Synthetic Yeast 2.0 programme, 213
Synthetically reconstructed Frog Prince, 109
Systematic evolution of ligands by exponential enrichment procedure, 86

T

T7 promoter–polymerase pair, 4
tac/*trc* promoters, 69
Tandem gene duplication, 6–7
Targetrons. *See* Group II introns
Taxadiene, 47
Terminal inverted repeats (TIRs), of DNA transposons, 109
Terminators, 77–78
tet promoters, 4
Tet-OFF module, 18
Theophylline, 350
Theophylline aptamer, 81
Theophylline-sensitive attenuator, 14
Thermus aquaticus, 240
Three-hybrid sensor systems, 26
Time-delay oscillator, 34
Tissue engineering, in synthetic biopolymers, 310
Toggle switch, 262–263
Top-down approach, 49
Trackable multiplex recombineering (TRMR), 215
Transcribed ultraconserved regions (T-UCRs), 85
Transcription, 15, 16
bacterial *vs*. eukaryotic promoter architecture, 11
of chromosomally integrated construct, 8
control, 12
core promoters, 11
exogenous promoter template, 12
of heterologous genes, 5
initiatio rates, modulation of, 12
machinery engineeering, 15–18
bacterial σ factors, 15
extracytoplasmic function σ factors, 16
functionally relevant regulatory networks, 15
T7 phage RNAP, 15
gTME (*see* Global transcription machinery engineering (gTME))
promoter specific factors, 11
regulation through termination, 13–14
RNA riboregulators, 14
ten-eleven translocation 1 (TET1) protein, 19
Transcription activator-like effector (TALE), 22–23, 116–117
Transcription attenuation, 14
Transcription factor, 18
artificial (*see* Artificial transcription factors)
decoy operators, 7–8
mining workflow, 6
titration effect, 8, 9
Transcription process, components of, 2, 3
Transcriptional engineering, in industrial applications, 47–49
Transcriptional engineering, in medicine, 43, 45, 46
antibody-based assays, 44
cancer treatment, 45
bacterial/viral vectors, 45
chemical inducers, 46
Clostridium novyi, 46
E. coli fdhF promoter, 45
ideal pharmaceutical, 45
radiation therapy, 46
tissue-specific promoters, 46
combatting infectious disease, 43–44
genetic memory elements, 44
type II diabetes treatment, 44
Transfer RNAs (tRNAs), 155
Translation termination, 184

Transposition, 109, 110
Transposon-mediated mutant library, 216
Transposons, 109–110
Trimethylene carbonate, 319, 320
trp attenuator, 13
Truth table, 36
Twister RNA, 79
Two-hybrid sensor system, 26
Type II restriction endonucleases, 120
  BglBrick (BBb) assembly, 122
  BioBrick standard assembly, 121–122
Type IIS RE-based cloning methods
  general Modular Cloning strategy, 123
  Golden Gate cloning method, 122
  GoldenBraid method, 123
Tyrosine-derived polyarylates, 328
Tyrosyl-tRNA synthetase (TyrRS), 188–189

U
Upstream promoter (UP) elements, 69
Uridine-specific excision reagent (USER) cloning/fusion approach, 128

V
Vibrio cholerae, 43
Viral polymerases, 5

W
Water-soluble cationic polyphosphoesters, 326
Watson–Crick base pairing, 23, 192

X
Xeno-amino acids, 346–348
Xenobiotic life
  artificial cells, 348–351
  artificial RNAs, controlling cellular activity with, 339–341
  engineered proteins, controlling cellular activity with, 341–343
  xeno-nucleic acids, 343–346
Xeno-nucleic acids, 343–346
Xenopus laevis transcription factor IIIA (TFIIIA), 115

Z
Zinc finger proteins (ZFPs)
  amino acid residue specificity, 20
  Cys_2His_2 zinc finger, 20
  description, 20
  in eukaryotic organisms, 20
  OPEN, 21
Zinc finger Targeter (ZiFiT), 21
Zinc-finger nucleases (ZFN), 115–116