Vijai Singh *Editor*

Advances in Synthetic Biology

Springer

Advances in Synthetic Biology

Vijai Singh
Editor

Advances in Synthetic Biology

Springer

Editor
Vijai Singh
Department of Biosciences
School of Science
Indrashil University,
Mehsana, Gujarat, India

ISBN 978-981-15-0080-0 ISBN 978-981-15-0081-7 (eBook)
https://doi.org/10.1007/978-981-15-0081-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

I am delighted to write a brief introductory message for the *Advances of Synthetic Biology*, a scholarly presentation in the rapidly evolving field of synthetic biology.

Synthetic biology is a science of engineering biological systems from elementary parts like genes, RNA and proteins towards building higher order structures ranging from biomolecular interactions to pathways to multicellular networks. The key is to use data from reductionist approaches like molecular biology and biochemistry integrated into computational databases and models to generate standards and composition protocols for engineering biological systems. Once the rational design principles of constructing biological systems are standardized, applications in the area of health, energy and environment would be a natural consequence. Currently efforts are also under way to design cellular factories towards an uninterrupted production of chemicals and biochemicals at an affordable cost.

This book fulfils an important unmet need to bring out key ideas and approaches towards building functional biomolecular interactions and networks. This book deals with the design of synthetic promoter, ribosome binding site, transcription factor, small non-coding RNA, transcription terminator and so on. Synthetic parts, devices, systems, oscillators, biological gates, synthetic small regulatory RNAs (riboregulators and riboswitches), DNA sequencing, design and assembly of synthetic genome, biosynthetic pathways, synthetic biology toolbox, computation, extended genetic code, CRISPR-Cas systems, cell-free protein synthesis systems, microfluidics and ethical issues have been presented using an easy-to-understand narrative.

Dr. Vijai Singh has put tremendous effort into compiling a valuable literature covering key concepts and findings in synthetic biology. Through strong support from Springer Nature, I believe this book will be really useful for the students, researchers, clinicians, societal stakeholders and policymakers. My best wishes to the author and readers for enjoying this intellectual journey.

School of Biotechnology, Jawaharlal Nehru University,
New Delhi, India

Pawan K. Dhar

Preface

Synthetic biology is a newly emerging field that employs the application of engineering principles to biology. Recent advances in synthetic biology have led to the development of several synthetic gene circuits for a wide range of applications including disease diagnostics, disease treatment, production of biomolecules, therapeutics, vaccines, biomaterials, biofuels and fine chemicals. It has allowed us to easily manipulate or build a novel function or upgrade the existing gene function by using a synthetic promoter, ribosome binding sites, genes, transcription terminators, small RNAs and many more in a wide number of organisms and cell types. This book covers several important areas of synthetic biology which allow us to easily read and understand. It covers the introduction of synthetic biology and how we can build a synthetic gene network and create novel functions within a cell. It offers to design some key cell components including synthetic promoters, ribosome binding sites, transcription factors, small non-coding RNAs, transcription terminators and many more.

The genetic parts are assembled to create a synthetic oscillator that endows the cell with oscillatory behaviour that could be fine-tuned by using an inducer molecule or altering the temperature or pH. The current state of the art describes how to build an oscillator and characterize it in prokaryotic and eukaryotic cells. To grasp a better control over the cell, this book contains a section for the design and building of biologic gates and computation. Small non-coding RNA plays a key role in gene repression and activation. This book covers both small RNAs including riboregulator and riboswitch in detail along with their applications in biomedical, therapeutics and biotechnology. The expansion of tools and technologies in synthetic biology is on high priority. This book covers the tools and technologies involved in different DNA sequencing, cloning methods, small regulatory RNA, expansion of genetic code, CRISPR-Cas systems, etc. that have been developed for accelerating synthetic biology research. This book also covers cell-free protein synthesis systems, microfluidics and ethical issues associated with it.

This book highlights and explores several aspects of synthetic biology in a way that can help future investigators, researchers, students and stakeholders to perform their research effortlessly. The book is a compilation of 20 chapters that have been written by eminent scientists from seven countries including Bulgaria, France, India, Japan, South Africa, the UK and the USA. It makes use of a rich literary text of

excellent depth, clarity and coverage. This book can be an excellent basis which may allow scientific knowledge to grow, widen and accelerate synthetic biology. This book offers not only a better understanding of synthetic biology but also triggers unanswered questions. Though monumental efforts have been invested to make this book user-friendly, we are aware that the first version always comes with bugs. We would be pleased to receive comments and suggestions to improve the book further.

Mehsana, Gujarat, India Vijai Singh

Acknowledgements

I would like to express my sincere gratitude and deep appreciation to Dr. J.S. Yadav, Director (Research), Indrashil University, India, for extending his outstanding support and motivation to complete this book. I am also grateful to Prof. Pawan K. Dhar who gave me outstanding personal and professional support as well as inspiration to accomplish this book.

I am delighted to thank all the authors for their excellent contributions to this book. I would like to thank Dr. Bhavik Sawhney (Associate Editor—Biomedicine) and Ms. RaagaiPriya ChandraSekaran (Production Coordinator) from Springer for their excellent management of this project.

I greatly appreciate the support of my students Nisarg Gohil and Gargi Bhattacharjee, whose discussion and comments helped to shape this book. I thank Vimal C. Pandey, Satya Prakash and those whose names do not feature here but have directly or indirectly contributed in shaping this project.

I wish to express my gratitude to my beloved wife Pritee Singh for her endless support, patience and inspiration. Lots of affection for my kids Aaradhya and Ayush who missed me during this project.

I would like to warmly thank the faculty and staff of Indrashil University for providing a great working environment. Last but not least, my sincere thanks to GOD for his supreme POWER and endowing me to live with joy and victory in the shape of this book.

Contents

About the Editor

Vijai Singh is an Associate Professor at the Department of Biosciences, School of Science, Indrashil University, Mehsana, Gujarat, India. He previously served as a Postdoctoral Fellow at the Institute of Systems and Synthetic Biology, France and at the School of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology, South Korea. Dr. Singh has also served as an Assistant Professor at Department of Biotechnology, Invertis University, India and the Department of Biological Sciences and Biotechnology, Institute of Advanced Research, India. He received his Ph.D. in Biotechnology from Dr. A.P.J. Abdul Kalam Technical University/National Bureau of Fish Genetic Resources, Lucknow, India.

Dr. Singh has designed and characterized a number of synthetic oscillators, gene regulatory networks, and biosynthetic pathways in *Escherichia coli*. Currently, his laboratory is focusing on the construction of a novel biosynthetic pathway for the production of pigments and chemicals. Additionally, his laboratory works on development of CRISPR-based platform for the diagnosis and eradication of multidrug resistant pathogens. Dr. Singh has published 72 articles, 23 book chapters, and 2 books. He currently serves as a member of the editorial board and reviewer for a number of peer-reviewed journals.

Introduction to Synthetic Biology

1

Vijai Singh

Abstract

Synthetic biology is a newly growing field which allows us to design non-natural parts, devices and circuits for biotechnological applications. These novel systems can help to find a solution for current challenges that we are facing in context to fulfilling the demand of drugs, vaccines, precise diagnosis, fine chemicals, biofuels and so on. In the past decade, a number of parts, devices and systems have been engineered and characterized in many organisms. Currently, a number of research groups are focusing on the development of new technologies/assays, including CRISPR-Cas9, riboregulators, riboswitches, cell-free protein synthesis and microfluidics that can accelerate synthetic biology research and its applications. This chapter highlights the progress, challenges and applications of parts, devices, circuits and tools towards biological, biomedical, therapeutic and industrial purpose.

Keywords

Promoter · Transcription factor · CRISPR-Cas9 · Riboregulators · Riboswitches · Circuits · Gene regulation · Gene network

1.1 Introduction

Synthetic biology is a new field that incorporates engineering principles in biology. In the past decade, biological parts, devices and systems have been engineered and tested in many organisms (Endy 2005; Purnick and Weiss 2009; Khalil and Collins 2010). Novel small genetic parts such as promoters (Lutz and Bujard 1997; Alper et

V. Singh (✉)
Department of Biosciences, School of Science, Indrashil University, Mehsana, Gujarat, India
e-mail: vijai.singh@indrashiluniversity.edu.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_1

al. 2005), proteins, RNAs (Basu et al. 2005; Pfleger et al. 2006) and scaffolds (Park et al. 2003; Dueber et al. 2009) have been engineered and well characterized. These parts are reusable and can be assembled together to build novel devices and complex circuits such as oscillators (Elowitz and Leibier 2000; Stricker et al. 2008), biologic gates (Tamsir et al. 2011; Moon et al. 2012), riboswitches (Bayer and Smolke 2005; Wang et al. 2008; Patel et al. 2018) and riboregulators (Isaacs et al. 2004; Callura et al. 2010; Patel et al. 2018) for controlling the cellular behaviour and use them for biotechnological applications.

In recent times, developing synthetic biology technologies is on high priority. It has gained much scientific and public attention towards building or modifying organisms with highly predictive phenotype. A pressing need arises to build advance systems that can solve major issues of health, environment and energy. A wide range of gene cloning methods (Li and Elledge 2007; Sleight et al. 2010), genome assembly (Gibson et al. 2009), genome designing (Gibson et al. 2010), cell-free protein synthesis (Carlson et al. 2012; Moore et al. 2017), microfluidics platform (Balagaddé et al. 2005; Marcus et al. 2006; Maerkl and Quake 2007; Stricker et al. 2008) and CRISPR-Cas9 system (Jinek et al. 2012; Mali et al. 2013; Cong et al. 2013; Jiang et al. 2013; DiCarlo et al. 2013; Bikard et al. 2014; Jakočiūnas et al. 2015; Singh et al. 2017, 2018) have been introduced. This chapter highlights the recent advances, challenges and future potential of synthetic biology for industrial, therapeutic, biomedical and biotechnological applications.

1.2 Engineering and Characterization of Synthetic Parts

Currently, a wide range of synthetic parts including promoter, RBS, transcription factor, small non-coding RNA and transcription terminator have been engineered and tested in many model organisms and cell types.

1.2.1 Synthetic Promoter

A promoter is a portion of specific DNA sequences where RNA polymerase binds and starts transcription process. The main component of *Escherichia coli* promoter is −35 and −10 (Pribnow box) region and operator region (repressor or activator binds). These play a key role in activating or repressing gene expression. Inducible, constitutive and hybrid promoters are mainly used in synthetic biology for designing and construction of synthetic gene circuits, biosynthetic pathways and complex devices for biotechnological applications. An inducible promoter is an important promoter for gene expression. A number of wild types of inducible promoters (pLlac01, pLtet01, pBAD, etc.) have been modulated and re-designed for holding tight control over the gene expression. The pLlac01 promoter contains the operator for LacO where LacI repressor binds and stops transcription. It can be activated by using the IPTG inducer molecule. Similarly, pLtet01 is regulated by TetR and

is induced by anhydrotetracycline (aTc), while pBAD is regulated by araC and activated by arabinose (Lutz and Bujard 1997; Gardner et al. 2000; Khlebnikov et al. 2001; Guet et al. 2002; Stricker et al. 2008).

Constitutive promoters do not possess any operator site for binding of repressor or activator. This promoter is unregulated by any transcription factor; therefore, it continuously expresses the gene. Seven constitutive promoters have been identified in *E. coli* (Liang et al. 1999). Registry of Standard Biological Parts has a series of constitutive synthetic promoters (BBa_J23100 to BBa_J23119) that are regularly used in synthetic biology and metabolic engineering for constructing tunable circuits (Carrera et al. 2011). In the past decade, number of hybrid synthetic promoters have been designed and characterized in many organisms. These promoters contain operators for two transcription factors. In the absence of one inducer molecule, the gene will not be fully expressed. In order to get full activation, both the inducers molecules are required (Kuhlman et al. 2007). pLlac_lux is one such example that contains two operators sites, LacI and luxR, which are induced by IPTG and CO6HSL molecules. pLac_Tet that is repressed by LacI and TetR is in turn induced by IPTG and aTc molecules in order to get full gene expression. Therefore, promoters are essential components in synthetic biology and metabolic engineering. In order to make more standard biological parts, it should be more standardized and modulated so that it can be easily used in biotechnological, therapeutic and industrial applications.

1.2.2 Ribosome-Binding Site

A ribosome-binding site (RBS) is present upstream of messenger RNA and allows ribosome to bind there and begin translation. RBS is also known as Shine–Dalgarno sequence (SD) or Shine–Dalgarno box (Shine and Dalgarno 1975). A single mutation can affect the translation efficiency by weaker binding of mRNA-ribosome pairing efficiency. SD sequences are present both in bacteria and in archaea. The SD sequences regulate the rate of translation. In a study, sequence AAAGAGGAGAAA is considered as a stronger RBS of *E. coli* and was used for building a synthetic oscillator (Elowitz and Leibier 2000). A library of synthetic RBS has been engineered and tested in a number of organisms with different efficiencies. The different strength of RBS can help to regulate the single gene expression or cluster of genes with a wide range of biological functions.

1.2.3 Transcription Factor

A transcription factor is a protein that directly binds with the promoter sequences of the operator region and starts or stops the gene expression. It can perform its function either alone or with the help of complex protein molecules by activating or repressing target gene (Roeder 1996). In synthetic biology, LacI, TetR, Lambda CI and AraC are used for the construction of circuits (Gardner et al. 2000; Guet et al.

2002). For instance, lac operon in *E. coli* contains a number of genes including lacZ, lacY and lacA. These genes express enzymes β-galactosidase, lactose permease, and thiogalactoside transacetylase or galactoside O-acetyltransferase, which is mainly repressed by LacI repressor and activated by lactose or IPTG molecules (Jacob and Monod 1961). Similarly, TetR is another repressor that can bind with the tetO operator region and is induced by aTc molecule (Helbl et al. 1995). TetR is extensively used in synthetic biology for construction of gene networks because of its tight control for fine-tuning of expression of gene (Elowitz and Leibier 2000; Gardner et al. 2000; Swinburne et al. 2008). In addition, luxR is modulated by *N*-acyl-L-homoserine lactone (AHL) molecules. It is considered as either activator or repressor (Koch et al. 2005; Gohil et al. 2018). Another important transcription factor AraC can bind to the araBAD promoter and get activated in the presence of arabinose (Lee et al. 1981). All these above transcription factors are useful in building of regulatory synthetic devices, circuits and pathway or to re-design an existing pathway.

1.2.4 Protein Degradation Tag

Protein degradation plays a crucial role in the reduction of protein overload in the cellular systems. It is regulated by degradation machinery which can clear the protein and maintain the cellular physiology and regulation. The turnover rate is important for balancing the protein functioning and visualizing signal. In the case of synthetic circuits for measuring network dynamics, the transcription factor and fluorescence encoding genes should be fused with degradation tag ssrA. The ssrA is a common degradation tag that is degraded by ClpXP degradation machinery. It decreases the protein half-life and is used for measuring network dynamics (Elowitz and Leibier 2000; Stricker et al. 2008; Danino et al. 2010). Registry of standard biological parts (http://parts.igem.org/Main_Page) has a wide range of weak and strong degradation tags. It is physically available and also can be designed in the form of genes or primers. In order to make a fast dynamics cellular system, degradation tag should be incorporated and used for measuring cellular network function.

1.2.5 Transcription Terminator

Transcription terminator is an important component of the cells. It can stop the running RNAP for transcription process. In prokaryotes, it is mainly rho-dependent and rho-independent types. Rho-independent transcription comprises a palindromic region that creates a G-C rich base pair stem loop followed by T bases. This loop causes RNAP to pause transcription of poly-A tail (Watson 2004). Rho-dependent transcription terminator requires Rho factor which can form a hexameric ring wherein two RNA-binding sites are involved (Skordalakes and Berger 2003). In *E. coli* chromosome, two sets of unidirectional DNA replication pause (Ter)

sites are present that contain the replication fork which controls the termination of chromosome replication (Duggin and Bell 2009). A library of synthetic terminators is available that may be used for designing and construction of synthetic circuits.

1.3 Engineering and Characterization of Synthetic Devices and Systems

1.3.1 Small Non-coding RNA

Small non-coding RNA is a key element present in prokaryotic and eukaryotic organisms. It can either activate or repress gene expression. These small regulatory elements are commonly known as riboswitches or riboregulators, which have been described below.

1.3.1.1 Riboswitches

A riboswitch is a small non-coding regulatory RNA which is mainly present upstream of mRNA and forms a loop that can be dynamically changed in the presence of ligand molecules and activate or repress gene function (Nudler and Mironov 2004; Vitreschak et al. 2004; Tucker and Breaker 2005; Batey 2006; Patel et al. 2018). Riboswitches contain two parts: (1) an aptamer which is responsible for the binding of small ligand molecules and (2) an expression platform which can undergo dynamic changes in response to modulations in the aptamer that allow either activation or repression of a gene function. There are different types of ligand molecules, including protein, amino acids, chemicals, metals, antibiotics, etc. that can bind with small regulatory RNA and dynamically change the secondary structure of RNA allowing them to activate or repress gene function, detect metals, chemicals, etc. There are a number of well identified and characterized riboswitches in wide range of organisms. One such example is that of cobalamin riboswitch that binds with adenosylcobalamin and regulates the biosynthesis and transport of cobalamin (Nou and Kadner 2000).

Glycine riboswitch is known to regulate glycine metabolism upon the binding of glycine molecules (Sherman et al. 2012). Similarly, lysine riboswitch is regulated by binding of lysine molecule, and it regulates lysine biosynthesis. NiCo riboswitch binds with two metals ions nickel (Ni) and cobalt (Co) and is used for metal detection and biosensing. A wide range of naturally occurring riboswitches has been identified and tested. Many riboswitches have been artificially designed, characterized and used in metabolic engineering for monitoring lysine concentration (Yang et al. 2013), therapeutics, biosensing and synthetic biology applications.

1.3.1.2 Riboregulators

A riboregulator is a small non-coding RNA molecule that responds to signal by Watson–Crick base pairing. Under normal condition, mRNAs are cis-repressed and the small non-coding RNA can open the secondary structure allowing them to bind to the ribosome and start the translation process. It can act at different

stages of transcription, translation and post-transcription. It plays a key role in biotechnology, therapeutic and industrial applications (Patel et al. 2018). The first synthetic riboregulator was designed, constructed and characterized in *E. coli*. Small cis-regulatory complementary sequences were inserted upstream of the target gene. In transcription, cis-repressed sequences form a stem loop at the 5′ untranslated region (UTR) of mRNA which tends to interfere with ribosome binding. A small RNA is expressed in trans target which binds with UTR and alters the stem-loop structure, allowing the activation of target gene function (Isaacs et al. 2004).

Similarly, riboregulator has been designed that can confer biologic gates function (Rinaudo et al. 2007) in mammalian cells. Nechooshtan et al. constructed a pH-responsive riboregulator that can respond to change in the pH, thereby causing activation or repression of gene function (Nechooshtan et al. 2009). In metabolic engineering, for increasing the carbon flux or redirecting carbon flux towards desired products, it is important to knock-out non-essential gene or down-regulate essential genes (Gohil et al. 2017, 2019; Panchasara et al. 2018). Na et al. designed and constructed a library of small synthetic RNAs in *E. coli* for increasing tyrosine and cadaverine production (Na et al. 2013). Therefore, riboregulator can be designed to accomplish wider applications in the biomedical, diagnostic, therapeutic and industrial setting.

1.3.2 Synthetic Oscillator

The first synthetic circuits such as repressilator (Elowitz and Leibier 2000) and toggle switch (Gardner et al. 2000) were designed and characterized in the year 2000 in *E. coli*. This groundbreaking work laid the foundation of synthetic biology. Over the past decade, number of synthetic devices and systems have been designed and characterized in different organisms and cell types. It is widely used in a number of applications in synthetic biology and metabolic engineering for fine-tuning or up-regulation of production of the metabolites, biofuels, chemicals, therapeutics and much more.

Ever since its discovery, synthetic oscillators have opened up a new avenue for a digitalized biological living system. It has shown predictive behaviour. The oscillator has the potential to deliver 'one dose per day' in a time-dependent manner which rather depends on the programmed circuits. Building oscillators mainly requires negative or positive/negative feedback in circuits. Elowitz and Leibler engineered three genes-based synthetic oscillator which they called as repressilator. It contains LacI, TetR and lambda CI transcription factors and its corresponding promoters that were induced by IPTG. They used GFP as a reporter to monitor oscillatory behaviour, and it was found that 40% cells were oscillating with a period of 160 ± 40 (mean ± s.d.) min and cell division took 50–70 min (Elowitz and Leibier 2000). Repressilator was constructed based on negative feedback, and taking advantage of this, Stricker et al. built a fast, robust and tunable synthetic oscillator in *E. coli*. They used LacI as a repressor and AraC as an activator and pLac-ara as a synthetic promoter. For monitoring the oscillation, GFP with degradation tag

was used to obtain fast dynamics of systems. They observed robust oscillation at 2 mM IPTG at 37 °C within 13–58 min of period by changing 0.1–3.0% of arabinose concentrations. Oscillation was independent of cell cycle (Stricker et al. 2008).

Similarly, a synchronized synthetic oscillator has been built based on quorum-sensing molecules in *E. coli*. A mean oscillatory period of 90 ± 6 min and mean amplitude 54 ± 6 GFP arbitrary units were obtained during high flow rate. At a low flow rate, 55 ± 6 min oscillating period with the amplitude of 30 ± 9 GFP was obtained. This oscillation was based on degrade-and-fire dynamics (Danino et al. 2010). A wide range of synthetic oscillators has been designed and constructed in prokaryotic and eukaryotic organisms. It can be used and expanded towards therapeutic, industrial and biotechnological applications.

1.3.3 Biologic Gates

In electrical engineering, logic gates based large-scale digital circuits using Boolean gates are a standard method and have been used for a long time in computer technology. In biology, an orthogonal AND gate was constructed in *E. coli*. Two co-activating genes (hrpR and hrpS) which were controlled by input promoter and a σ(54)-dependent hrpL promoter as output were used. The hrpL promoter was activated when both genes were expressed that allowed AND gate to function. It was modulated by applying another promoter output that connected the output to NOT gate to achieve a NAND gate function (Wang et al. 2011). Similarly, LuxI-LuxR quorum-sensing system (Miller and Bassler 2001; Gohil et al. 2018) was used to generate AND gate function. It could further enhance the specificities by 1.5-fold (Sayut et al. 2011).

In addition, Tamsir et al. designed and constructed XOR logic gate in *E. coli* using orthogonal quorum-sensing sender and receiver devices. Four strains were used that carried different logic gates that had the ability to perform XOR gate function. In the experiment, cell 1 carried NOR and used Ara and aTc input that expressed LasI output. This cell was wired with NOR in cells 2 and 3 through the 3OC12-HSL (N-3-oxododecanoyl homoserine lactone). When, cells 2 and 3 used Ara and aTc as a second input, the output of NOR in cells 2 and 3 was RhlI that produced C4-HSL (*N*-butyryl-homoserine lactone). Cell 4 acted as a buffer gate. The output could be monitored using YFP reporter (Tamsir et al. 2011). In a study, Moon et al. constructed a two input-based AND gate in *E. coli* for creating a more complex programme (Moon et al. 2012). A number of biologic gates have also been designed and constructed in bacteria to mammals. These engineered biologic gates have a wide range of applications in biomedical, therapeutic, reprogramming and also biological computation.

1.4 Advances in Gene Cloning and Genome Assembly Methods

In synthetic biology, gene cloning and expression are key factors that always constrain and limit the proceedings of the experiment. Synthetic gene circuits need to be fine-tuned and optimized in order to attain predictive function. Therefore, apart from the routine gene cloning techniques including sticky end, blunt end and T-A cloning, the recently established gene cloning techniques to accelerate synthetic biology research have been described briefly.

1.4.1 Ligase-Independent Cloning Technique

Ligase-independent cloning (LIC) is restriction enzyme-free cloning technique, which is simple, fast and easy to use it. It uses T4 DNA polymerase for creating a 10–15 bases single overhang in a vector which allows the insert to join easily. LIC was initially used for inter-Alu fragment gene cloning from hybrid and human cells (Aslanidis and de Jong 1990; Haun et al. 1992). Similarly, sequence and ligation-independent cloning (SLIC) has been developed based on homologous recombination for multiple fragments cloning in a single tube reaction (Li and Elledge 2007). In-fusion PCR cloning developed by Clontech Laboratories, USA, is yet another restriction enzyme-free cloning technique, which is simple, rapid and efficient for assembling multiple genes together in a single reaction tube. Primers are designed that contain 15–20 bp homology with the corresponding genes. Insert and plasmid is incubated together for 15 min at 50 °C in the presence of an in-fusion enzyme that can generate overhang and get annealed. This reaction mixture can be directly transformed into the competent cell and clones can be screened based on the marker (Sleight et al. 2010).

1.4.2 Gibson Assembly

Gibson assembly developed by Gibson et al. (2009) is a simple, rapid and efficient technique in synthetic biology. It can be performed in a single tube using three enzymes including T5 exonuclease (chews back DNA), DNA polymerase (adds bases to fill gaps) and Taq DNA ligase (seals nick). Insert and plasmid need to contain 20–40 bases homology, which can be mixed with Gibson master mix and incubated at 50 °C for 30–60 min. Then, this reaction mix can directly be transformed into competent cells and screened for recombinant clones. Currently, Gibson assembly is commercially available in the market and is used for constructing complex circuits (Gibson et al. 2010).

1.5 CRISPR-Cas Systems

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein (Cas) are found within the genome of bacteria and archaea. It is an RNA-mediated immune system of prokaryotes that protects them from infection of bacteriophage and plasmids (Barrangou et al. 2007; Horvath and Barrangou 2010; Barrangou and Marraffini 2014; Bhattacharjee et al. 2019, 2020). CRISPR-Cas systems are divided into classes I and II, six types and 18 subtypes. All the CRISPR systems contain different types of targeted DNA, RNA or both along with signature proteins (Makarova et al. 2015; Hille et al. 2018).

Currently, the type II CRISPR-Cas9 is widely being used in synthetic biology, metabolic engineering and biotechnology. It has been used for genome editing and regulation in a wide range of organisms (Jinek et al. 2012; Mali et al. 2013; Cong et al. 2013; Jiang et al. 2013; DiCarlo et al. 2013; Bikard et al. 2014; Jakočiūnas et al. 2015). It is a simple, rapid and sensitive tool that requires expression of Cas9 protein, single-guide RNA and PAM sequences at the target region. The Cas9-sgRNA complex binds to the target region and generates a double-stranded break, which later repairs by either the NHEJ or HDR pathway. In this way, genome modification can be executed (Singh et al. 2017). CRISPR-Cas9 system has been used for genome editing of *Drosophila* (Ren et al. 2014; Port et al. 2014), zebrafish (Hwang et al. 2013; Hisano et al. 2015; Liu et al. 2017; Cornet et al. 2018), removal of viruses including human papillomavirus (Kennedy et al. 2014), hepatitis B virus (Lin et al. 2014; Zhen et al. 2015), latent Epstein–Barr virus (Wang and Quake 2014), HIV-1 (Ebina et al. 2013; Zhu et al. 2015), repairing of defective genes (Long et al. 2014; Nelson et al. 2016; Guan et al. 2016) and many more.

By mutating two active regions of the Cas9 moiety, scientists have created yet another variant of Cas9 called the dead Cas9, which has lost its ability to generate a DSB on DNA but it retains binding ability of target DNA. This is known as CRISPR interference (CRISPRi), which has been extensively used for gene regulation, epigenetic modifications, high-throughput screening and imaging genomic loci in wide range of organisms (Qi et al. 2013; Bikard et al. 2013; Chen et al. 2013; Gilbert et al. 2013; Ma et al. 2015).

The type VI class 2, Cas13a has been identified and used for development of simple, rapid and ultrasensitive tool for early pathogens detection. This is attributed as Specific High-Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK). SHERLOCK has been used for ultrasensitive detection up to attomolar concentration of Zika virus, Dengue virus, *E. coli* and *Pseudomonas aeruginosa* and to discriminate antibiotic-resistant *Klebsiella pneumoniae* as well as mutations in cancer (Gootenberg et al. 2017, 2018; Myhrvold et al. 2018; Khambhati et al. 2019a). CRISPR-based platform can be extended in many more serious pathogens for timely diagnosis in order to properly treat diseases.

1.6 Cell-Free Protein Synthesis System

Cell-free protein synthesis (CFPS) system is used in synthetic biology and metabolic engineering for biomedical, therapeutic and industrial applications. It is membraneless simple, fast and high-throughput tool for production of toxic products, chemicals, assembly of bacteriophages, biosynthetic pathways, incorporation of toxic non-natural amino acids in protein for tagging or improving the potency of protein and many more. CFPS systems, also known as cell-free transcription-translation (TX-TL) systems or cell-free expression systems, have emerged as a powerful tool for performing research without the use of living organisms. It can be used for direct control of transcription, translation, post-translational modification and metabolism in an open source (Carlson et al. 2012; Moore et al. 2017; Khambhati et al. 2019b).

It requires RNA polymerase, energy sources, cofactors, substrates, DNA or plasmid and translational machinery including ribosome, translation factors and tRNA synthases. CFPS system is commercially available and used for a number of synthetic biology applications. CFPS has been used for incorporation of toxic amino acids for production of proteins such as canavanine (Worst et al. 2015), production of therapeutics (Zawada et al. 2011), assembly of bacteriophage (Shin et al. 2012), building of orthogonal genetic codes (Des Soye et al. 2015; Chemla et al. 2015), diagnostic of Dengue, Zika and Ebola viruses (Pardee et al. 2014, 2016; Gootenberg et al. 2017), testing of synthetic gene networks (Pardee et al. 2014; Takahashi et al. 2015) and many more. CFPS system has the potential to up-regulate and accelerate synthetic biology research in the near future towards biomedical, therapeutic, industrial and biotechnological applications.

1.7 Microfluidics

Microfluidics is a rapidly growing and powerful tool with a number of applications including PCR, cloning, diagnostic and monitoring of single-cell dynamics. It has the potential to reduce cost and reagent consumption. Microfluidics chips are prepared using PDMS (polydimethylsiloxane) and curing reagent, which is a transparent polymer. Microfluidics is getting popular in synthetic biology and has gained a lot of scientific attention. It has been used for continuous measurement of synthetic network by continuous supplying of growth media to the cells (Balagaddé et al. 2005; Stricker et al. 2008).

Microfluidics was used for mRNA extraction from a single cell for synthesis of cDNA followed by PCR. It can allow us to accelerate error-free high-throughput experiment (Marcus et al. 2006; Maerkl and Quake 2007). Currently, microfluidics platform is being used for testing and screening of high-value chemicals and therapeutics and in synthetic biology experiments at single-cell measurement of cellular behaviour (Stricker et al. 2008; Danino et al. 2010; Shen et al. 2015; Rodrigo et al. 2017). A number of companies are developing microfluidics chips for high-throughput detection of pathogens for fitting treatment. Currently, the price of chips

is relatively high because of requisite raw materials and expertise, but in the near future, it is expected to go down due to high market demand and customized chip development.

1.8 Conclusion and Future Remarks

Synthetic biology has the potential to reduce the cost of high-value biomolecules for human and animal uses. Synthetic promoter libraries have been used for a wide spectrum of product production in a tunable way. Small genetic parts can be used for accelerating the construction of biological networks and biosynthetic pathways. Synthetic devices and circuits can be further used for digital control or reprogramming of cellular machinery for tight and tunable gene expression towards healthy cell growth and high production. Currently, burgeoning synthetic biology technologies are on high priority. CRISPR-Cas9 allows us for genome editing towards biotechnological, therapeutic and biomedical applications. Cell-free protein synthesis and microfluidics can be a useful and powerful tool for accelerating synthetic biology research towards finding a solution for food, health and energy.

References

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

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

Balagaddé FK, You L, Hansen CL et al (2005) Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. Science 309:137–140

Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 54:234–244

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

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

Batey RT (2006) Structures of regulatory elements in mRNAs. Curr Opin Struct Biol 16:299–306

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol 23:337–343

Bhattacharjee G, Khambhati K, Singh V (2019) RNA-guided CRISPR-Cas9 system for removal of microbial pathogens. In: Tripathi V, Kumar P, Tripathi P, Kishore A (eds) Microbial genomics in sustainable agroecosystems. Springer, Singapore, pp 227–248

Bhattacharjee G, Mani I, Gohil N et al (2020) CRISPR technology for genome editing. In: Faintuch J, Faintuch S (eds) Precision medicine for investigators, practitioners and providers. Academic Press, London, pp 59–69

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

Bikard D, Euler CW, Jiang W et al (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol 32:1146–1150

Callura JM, Dwyer DJ, Isaacs FJ et al (2010) Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. Proc Nat Acad Sci U S A 107:15898–15903

Carlson ED, Gan R, Hodgman CE et al (2012) Cell-free protein synthesis: applications come of age. Biotechnol Adv 30:1185–1194
Carrera J, Rodrigo G, Singh V et al (2011) Empirical model and *in vivo* characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. Biotechnol J 6:773–783
Chemla Y, Ozer E, Schlesinger O et al (2015) Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. Biotechnol Bioeng 112:1663–1672
Chen B, Gilbert LA, Cimini BA et al (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR-Cas system. Cell 155:1479–1491
Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR-Cas systems. Science 339:819–823
Cornet C, Di Donato V, Terriente J (2018) Combining zebrafish and CRISPR/Cas9: towards a more efficient drug discovery pipeline. Front Pharmacol 9:703
Danino T, Mondragón-Palomino O, Tsimring L et al (2010) A synchronized quorum of genetic clocks. Nature 463:326–330
Des Soye BJ, Patel JR, Isaacs FJ et al (2015) Repurposing the translation apparatus for synthetic biology. Curr Opin Chem Biol 28:83–90
DiCarlo JE, Norville JE, Mali P et al (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343
Dueber JE, Wu GC, Malmirchegini GR et al (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27:753–759
Duggin IG, Bell SD (2009) Termination structures in the *Escherichia coli* chromosome replication fork trap. J Mol Biol 387:532–539
Ebina H, Misawa N, Kanemura Y et al (2013) Harnessing the CRISPRCas9 system to disrupt latent HIV-1 provirus. Sci Rep 3:2510
Elowitz MB, Leibier S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338
Endy D (2005) Foundations for engineering biology. Nature 438:449–453
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342
Gibson DG, Young L, Chuang RY et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
Gibson DG, Smith HO, Hutchison CA et al (2010) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7:901–903
Gilbert LA, Larson MH, Morsut L et al (2013) CRISPR mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154:442–451
Gohil N, Panchasara H, Patel S et al (2017) Book review: recent advances in yeast metabolic engineering. Front Bioeng Biotechnol 5:71
Gohil N, Ramírez-García R, Panchasara H et al (2018) Book review: quorum sensing vs. quorum quenching: a battle with no end in sight. Front Cell Infect Microbiol 8:106
Gohil N, Bhattacharjee G, Khambhati K et al (2019) Engineering strategies in microorganisms for the enhanced production of squalene: advances, challenges and opportunities. Front Bioeng Biotechnol 7:50
Gootenberg JS, Abudayyeh OO, Lee JW et al (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356:438–442
Gootenberg JS, Abudayyeh OO, Kellner MJ et al (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 360:439–444
Guan Y, Ma Y, Li Q et al (2016) CRISPR-Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. EMBO Mol Med 8:477–488
Guet CC, Elowitz MB, Hsing W et al (2002) Combinatorial synthesis of genetic networks. Science 296:1466–1470
Haun RS, Serventi IM, Moss J (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. Biotechniques 13:515–518

Helbl W, Berens C, Hillen W (1995) Proximity probing of Tet repressor to tet operator by dimenthylsulfate reveals protected and accessible functions for each recognized base-pair in the major groove. J Mol Biol 245:538–548
Hille F, Richter H, Wong SP et al (2018) The biology of CRISPR-Cas: backward and forward. Cell 172:1239–1259
Hisano Y, Sakuma T, Nakade S et al (2015) Precise inframe integration of exogenous DNA mediated by CRISPR-Cas9 system in zebrafish. Sci Rep 5:8841
Horvath P, Barrangou R (2010) CRISPR-Cas, the immune system of bacteria and archaea. Science 327:167–170
Hwang WY, Fu Y, Reyon D et al (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229
Isaacs FJ, Dwyer DJ, Ding C et al (2004) Engineered riboregulators enable posttranscriptional control of gene expression. Nat Biotechnol 22:841–847
Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–356
Jakočiūnas T, Bonde I, Herrgård M et al (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metab Eng 28:213–222
Jiang W, Bikard D, Cox D et al (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233–239
Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
Kennedy EM, Kornepati AV, Goldstein M et al (2014) Inactivation of the human papillomavirus E6 or E7 gene in cervical carcinoma cells by using a bacterial CRISPR-Cas RNA-guided endonuclease. J Virol 88:11965–11972
Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11:367–379
Khambhati K, Bhattacharjee G, Singh V (2019a) Current progress in CRISPR-based diagnostic platforms. J Cell Biochem 120:2721–2725
Khambhati K, Bhattacharjee G, Gohil N et al (2019b) Exploring the potential of cell-free protein synthesis for extending the abilities of biological systems. Front Bioeng Biotechnol 7:248
Khlebnikov A, Datsenko KA, Skaug T et al (2001) Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology 147:3241–3247
Koch B, Liljefors T, Persson T et al (2005) The LuxR receptor: the sites of interaction with quorum-sensing signals and inhibitors. Microbiology 151:3589–3602
Kuhlman T, Zhang Z, Saier MH Jr et al (2007) Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. Proc Natl Acad Sci U S A 104:6043–6048
Lee NL, Gielow WO, Wallace RG (1981) Mechanism of araC autoregulation and the domains of two overlapping promoters, Pc and PBAD, in the L-arabinose regulatory region of *Escherichia coli*. Proc Natl Acad Sci U S A 78:752–756
Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
Liang S, Bipatnath M, Xu Y et al (1999) Activities of constitutive promoters in *Escherichia coli*. J Mol Biol 292:19–37
Lin SR, Yang HC, Kuo YT et al (2014) The CRISPR-Cas9 system facilitates clearance of the intrahepatic HBV templates *in vivo*. Mol Ther Nucleic Acids 3:e186
Liu J, Zhou Y, Qi X et al (2017) CRISPR/Cas9 in zebrafish: an efficient combination for human genetic diseases modelling. Hum Genet 136:1–2
Long C, McAnally JR, Shelton JM et al (2014) Prevention of muscular dystrophy in mice by CRISPR-Cas9-mediated editing of germline DNA. Science 345:1184–1188
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

Ma H, Naseri A, Reyes-Gutierrez P et al (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. Proc Natl Acad Sci U S A 112:3002–3007
Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. Science 315:233–237
Makarova KS, Wolf YI, Alkhnbashi OS et al (2015) An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol 13:722–736
Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826
Marcus JS, Anderson WF, Quake SR (2006) Microfluidic single-cell mRNA isolation and analysis. Anal Chem 78:3084–3089
Miller MB, Bassler BL (2001) Quorum sensing in bacteria. Annu Rev Microbiol 55:165–199
Moon TS, Lou C, Tamsir A et al (2012) Genetic programs constructed from layered logic gates in single cells. Nature 491:249–253
Moore SJ, Lai HE, Needham H et al (2017) *Streptomyces venezuelae* TX-TL–a next generation cell-free synthetic biology tool. Biotechnol J 12:1600678
Myhrvold C, Freije CA, Gootenberg JS et al (2018) Field-deployable viral diagnostics using CRISPR-Cas13. Science 360:444–448
Na D, Yoo SM, Chung H et al (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol 31:170–174
Nechooshtan G, Elgrably-Weiss M, Sheaffer A et al (2009) A pH-responsive riboregulator. Genes Dev 23:2650–2662
Nelson CE, Hakim CH, Ousterout DG et al (2016) *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science 351:403–407
Nou X, Kadner RJ (2000) Adenosylcobalamin inhibits ribosome binding to btuB RNA. Proc Nat Acad Sci U S A 97:7190–7195
Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29:11–17
Panchasara H, Patel S, Gohil N et al (2018) Biosynthetic pathway for production of renewable biofuels. In: Mittal V (ed) Biofuels. Central West Publishing, Orange, pp 63–86
Pardee K, Green AA, Ferrante T et al (2014) Paper-based synthetic gene networks. Cell 159:940–954
Pardee K, Green AA, Takahashi MK et al (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165:1255–1266
Park SH, Zarrinpar A, Lim WA (2003) Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. Science 299:1061–1064
Patel S, Panchasara H, Braddick D et al (2018) Synthetic small RNAs: current status, challenges, and opportunities. J Cell Biochem 119:9619–9639
Pfleger BF, Pitera DJ, Smolke CD et al (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24:1027–1032
Port F, Chen HM, Lee T et al (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Nat Acad Sci U S A 111:E2967–E2976
Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422
Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:173–183
Ren X, Yang Z, Xu J et al (2014) Enhanced specificity and efficiency of the CRISPR-Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep 9:1151–1162
Rinaudo K, Bleris L, Maddamsetti R et al (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. Nat Biotechnol 25:795–801
Rodrigo G, Prakash S, Shen S et al (2017) Model-based design of RNA hybridization networks implemented in living cells. Nucleic Acids Res 45:9797–9808
Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21:327–335

Sayut DJ, Niu Y, Sun L (2011) Engineering the logical properties of a genetic AND gate. Methods Mol Biol 743:175–184
Shen S, Rodrigo G, Prakash S et al (2015) Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. Nucleic Acids Res 43:5158–5170
Sherman EM, Esquiaqui J, Elsayed G (2012) An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. RNA 18:496–507
Shin J, Jardine P, Noireaux V (2012) Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synt Biol 1:408–413
Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254:34–38
Singh V, Braddick D, Dhar PK et al (2017) Exploring the potential of genome editing CRISPR-Cas9 technology. Gene 599:1–18
Singh V, Gohil N, Ramírez García R et al (2018) Recent advances in CRISPR-Cas9 genome editing technology for biological and biomedical investigations. J Cell Biochem 119:81–94
Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell 114:135–146
Sleight SC, Bartley BA, Lieviant JA (2010) Designing and engineering evolutionary robust genetic circuits. J Biol Eng 4:12
Stricker J, Cookson S, Bennett MR et al (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456:516–519
Swinburne IA, Miguez DG, Landgraf D et al (2008) Intron length increases oscillatory periods of gene expression in animal cells. Genes Dev 22:2342–2346
Takahashi MK, Chappell J, Hayes CA et al (2015) Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription–translation (TX-TL) systems. ACS Synt Biol 4:503–515
Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 469:212–215
Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. Curr Opin Struct Biol 15:342–348
Vitreschak AG, Rodionov DA, Mironov AA (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet 20:44–50
Wang J, Quake SR (2014) RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection. Proc Natl Acad Sci 111:13157–13162
Wang JX, Lee ER, Morales DR et al (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol Cell 29:691–702
Wang B, Kitney RI, Joly N et al (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat Commun 2:508
Watson JD (2004) Molecular biology of the gene. Pearson India Education, Chennai
Worst EG, Exner MP, De Simone A et al (2015) Cell-free expression with the toxic amino acid canavanine. Bioorg Med Chem Lett 25:3658–3660
Yang J, Seo SW, Jang S et al (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. Nat Commun 4:1413
Zawada JF, Yin G, Steiner AR et al (2011) Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. Biotechnol Bioeng 108:1570–1578
Zhen S, Hua L, Liu YH et al (2015) Harnessing the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated Cas9 system to disrupt the hepatitis B virus. Gene Ther 22:404–412
Zhu W, Lei R, Le Duff Y et al (2015) The CRISPRCas9 system inactivates latent HIV-1 proviral DNA. Retrovirology 12:22

Current Progress in Synthetic Genetic Networks

2

Amir Pandi and Heykel Trabelsi

Abstract

Synthetic genetic networks are the main and most-studied area of synthetic biology. Biological networks or circuits provide modular and scalable tools to design-build-test synthetic biological systems for medical, environmental, and industrial applications. This chapter focuses on introducing and discussing the recent progress in design and application of such devices. This chapter starts with the classification of synthetic genetic networks and the role of each and their pros and cons. Then, recent applications of digital/analog genetic/metabolic circuits are presented in three groups of smart therapeutics, diagnosis, and metabolic engineering. Finally, tools and methods of implementing different classes of synthetic gene circuits are presented with covering the majority of the developed methodologies so far. This chapter brings a complete introduction to synthetic genetic circuits and their recent advances to the audience who aim to get familiar with this fast-growing technology.

Keywords

Synthetic genetic networks · Genetic circuits · Digital and analog computation · Gene and metabolic circuits · Genetic circuits applications · Genetic circuits implementation

2.1 Introduction to Synthetic Genetic Networks

Synthetic genetic networks or gene circuits are advanced tools to implement synthetic biological systems for a variety of medical, industrial, and environmental applications (Brophy and Voigt 2014; Purnick and Weiss 2009). The aim of

A. Pandi · H. Trabelsi (✉)
Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_2

these devices is to engineer biological systems receiving multiple inputs such as nutrition and signals, compute them through its artificial networks, and actuate the integrated outputs responding to the environment (Purcell and Lu 2014). The term "computation" which is used in this context means computing biological signals through a synthetic network composed of biological components. As one of the main tools in the field of synthetic biology, genetic networks have been synthesized for the development of (1) biosensors for detection of biomarkers or pollutants, (2) screening or engineering the dynamic regulation of metabolic pathways, or (3) smart therapeutics (Brophy and Voigt 2014).

Inspired by electrical engineering, several synthetic biological devices have been synthesized since the emergence of the field of synthetic biology (Selberg et al. 2018). These devices mimic the digital or analog computation paradigm by applying different classes of cellular components (Purcell and Lu 2014). To name some of the approaches, the synthetic genetic networks implemented so far consist of single or multilayer logic gates (Nielsen et al. 2016; Guiziou et al. 2019), oscillators (Rosier and de Greef 2015), amplifiers (Bonnet et al. 2013; Zeng et al. 2018; Wang et al. 2015), switches (Gardner et al. 2000) and memory devices (Bonnet et al. 2012; Farzadfard and Lu 2014), toehold circuits (Green et al. 2014), CRISPR circuits (Bikard et al. 2013; Nielsen and Voigt 2014), metabolic logic gates (Courbet et al. 2018), and metabolic perceptrons and classifiers (Pandi et al. 2019).

Thanks to the substantial efforts by the synthetic biology community, standard and modular methodologies have been established to engineer different above-mentioned devices (Kelwick et al. 2014; Marchisio 2014). Computer-assisted and bioinformatic tools are the accessory tools through which these methodologies can be generated (Nowogrodzki 2018; MacDonald et al. 2011). These approaches employ cellular components, from gene expression regulators to posttranslational level and metabolic enzymes.

The standardized and modular strategies have led the field to very advanced achievements in building sophisticated genetic networks. However, the next generation of synthetic cellular networks needs to focus on the integration of different approaches enabling hybrid analog–digital computation by the use of several types of cellular machineries (Brophy and Voigt 2014; Goñi-Moreno and Nikel 2019). The integration strategies and cross-species approaches (Kushwaha and Salis 2015; Xiao et al. 2017) empower the potential of artificial genetic networks to be applied for several applications in diverse living species and cell-free systems.

2.2 Classification of Synthetic Genetic Networks

Designed biological circuits can be categorized based on the computational approach that they lay on, digital and analog (Purcell and Lu 2014), or based on the biological functionality of the genes they employ, regulatory and metabolic (Goñi-Moreno and Nikel 2019) (top panel in Fig. 2.1).

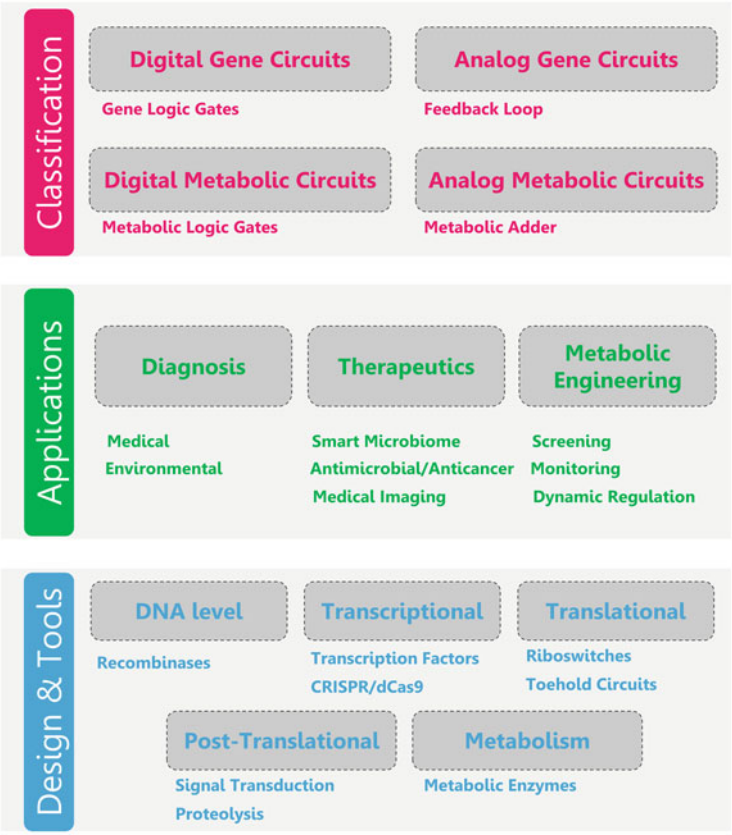


Fig. 2.1 An overview of the whole chapter in a schematic representation. This chapter is presented in three parts to cover the current progress in synthetic genetic networks. *Top panel:* Classification of the synthetic genetic networks in four classes depending on the computation approach that they rely on, digital/analog in gene expression/metabolic layer. *Middle panel:* Applications of synthetic genetic networks in diagnosis, smart therapeutics, and metabolic engineering. *Bottom panel:* Designing strategies and tools to implement synthetic genetic networks using different biological components of the cell in DNA level, transcriptional and translational, posttranslational, and metabolic components. The designed biological networks then will be implemented in eukaryotic/prokaryotic cells or in cell-free systems

2.2.1 Digital and Analog Gene Circuits

As in the electrical circuits engineering, biological gene circuits can perform digital or analog computation depending on their design (Purcell and Lu 2014). So far, the majority of the implementation of genetic networks has focused on digital computation, as it is more standardizable using well-established tools such as Verilog (Nielsen et al. 2016). The digital gene implementation follows the logic functions. For instance, if A and B both are needed to generate an output (either a

reporter in medical or environmental diagnosis or expression of a functional gene), this is an AND logic. The OR gate is the logic gate output of which is active (ON) when two or even of the conditions/inputs of A or B are "ON."

So far, several digital gene circuits have been implemented using different cell components at the level of DNA (Guiziou et al. 2018, 2019; Engelen et al. 2017; Genot et al. 2011), transcription (Nielsen et al. 2016; Kim et al. 2019; Gander et al. 2017; Buchler et al. 2003; Bradley et al. 2016), and RNA (Green et al. 2014; Deng et al. 2014; Martini et al. 2015; Kim et al. 2018; Wu et al. 2019), as well as the protein level (Gao et al. 2018; Fink et al. 2019; Razavi et al. 2014; Fernandez-Rodriguez and Voigt 2016). In cases where there are more than two inputs with complex relationships, their behavior cannot be captured as easy as for simple AND or OR gates. This is a point where computational tools can be used to introduce a complex logic circuit in which the relationship between inputs and output(s) can be computed through multilayer digital gene networks (Nielsen et al. 2016).

Since most of the synthetic biocircuits have been built in the gene expression level, the digital-like behavior (ON/OFF) in the gene expression system has compatibilized the digital computation approach. Therefore, a number of successful digital computation approaches have been introduced during the past few years. However, digital-like behavior is not the only using which cells perform computation. A considerable contribution of biological computation in living cells takes place in an analog manner where the continuous concentrations of the cellular components define the phenotype, not their presence or absence (ON/OFF) (Purcell and Lu 2014; Sauro and Kim 2013).

The substantial contribution of the analog computation in living systems brings the mindset of implementing analog gene circuits. In electronics, analog circuits consume lower energy and require fewer parts to function. In the same way, analog gene networks save cellular energy and avoid the burden (Daniel et al. 2011; Sarpeshkar 2014). This valuable advantage promotes the system orthogonality by using fewer synthetic parts.

There have been only a few studies investigating the analog computation in living systems (Daniel et al. 2011, 2013). Daniel et al. (2013) have developed synthetic analog computation in living cells using a feedback loop inspired by the feedback loop of operational amplifiers in analog electronics. In this study, a simple transcriptional circuit has been designed in a construct such that: (1) in a low-copy plasmid, the transcription factor (TF) is expressed under its cognate promoter controlled by the externally added inducer, and (2) in a high-copy plasmid, the cognate promoter expresses a fluorescent protein reporting the concentration of the ligand. This design alleviates the saturation of the TF (through the feedback loop in the low-copy plasmid that produces more TF and delays its saturation) and the saturation of the cognate promoter (through pulling the flux of transcription to the responsive promoter in the high-copy plasmid). This construction linearizes the dose response of the circuit from a digital-like behavior to an analog behavior (Daniel et al. 2013).

2.2.2 Digital and Analog Metabolic Circuits

Although the analog behavior is one of the characteristics of living cells, it is difficult to implement analog gene circuits which naturally show a digital-like behavior (ON/OFF). However, using other biological mechanisms such as metabolism is more compatible to implement analog computation (Pandi et al. 2019). In this direction, an analog metabolic computation approach has been recently established that is using metabolic enzymes to perform analog biocomputation (Pandi et al. 2019). In this study, metabolic pathways were designed using computer-aided tools (Delépine et al. 2016, 2018) and were implemented in whole-cell and cell-free systems. Multiple metabolic transducers were implemented that are metabolic pathways composed of one or more enzymes transforming a metabolite into another, a product that can be sensed using transcriptional or translational regulators (Koch et al. 2018). By combining metabolic transducers, analog adders were built in both whole-cell and cell-free systems. Cell-free systems enabled performing more complex computations by tightly controlling the amount of DNA of the circuit added to the reaction. This advantage of the cell-free system and high adjustability, along with rapid characterization and possibility of mixing multiple genes at different concentrations, enabled the development of four-input classifiers. In the classifiers, a metabolic perceptron receives four input metabolites and converts them into a common metabolite by model-computed concentrations of their associated enzyme DNA and finally reported through a gene circuit actuator. The metabolic perceptron was inspired by a perceptron algorithm invented in 1960s to mimic human neural systems in information processing and decision-making (Rosenblatt 1958). Since then, perceptrons have become the building blocks of several neural computing and deep learning algorithms (Haykin 2011).

Digital metabolic circuits are other types of biological computation using artificial networks that apply metabolic enzymes to build metabolic logic gates. A number of metabolic logic gates including AND, OR, XOR, NAND, and their combination in order to build complex circuits have been developed (Courbet et al. 2018; Katz 2017). In most of cases, dealing with cellular cofactors and coenzymes for the signal processing makes the application of digital metabolic circuits limited in whole-cell systems and biological samples. Nevertheless, depending on the case, they have a valuable potential to build synthetic genetic network.

2.3 Applications of Synthetic Genetic Networks

The following are the applications of synthetic gene networks in variety of aspects (middle panel in Fig. 2.1).

2.3.1 Diagnosis

One of the main application of synthetic genetic networks is to develop diagnostic devices (Slomovic et al. 2015). In this context, gene and metabolic circuits have been used to build various genetic networks. For instance, a simple genetic network comprising the quorum-sensing regulatory system of *Pseudomonas aeruginosa* has been engineered in the cell-free system to detect this pathogen in clinical samples (Wen et al. 2017). In a different approach, paper-based cell-free toehold circuits built using RNA switches were utilized to sense RNAs for Zika virus (Pardee et al. 2016), Ebola virus (Pardee et al. 2014), or gut microbiome bacteria in fecal samples (Takahashi et al. 2018). The CRISPR machinery also has been adapted to detect DNA and RNA of viruses and bacterial pathogens in vitro using strategies called SHERLOCK (Gootenberg et al. 2017, 2018), DETECTR (Chen et al. 2018a), and HOLMES (Li et al. 2018). In another approach, applying gene switches built by recombinases in vivo enabled detection of glucose in diabetic clinical samples (Courbet et al. 2015). Using a radically different approach, metabolic enzymes have enabled increasing the number of detectable small molecules. In this work, by plugging metabolic enzyme, a molecule is converted to another which is sensible through transcriptional regulator (Voyvodic et al. 2019). The authors have introduced a modular tool to implement and optimize cell-free biosensors and used this strategy to sense benzoic acid in beverages, as well as hippuric acid and cocaine in clinical samples (Voyvodic et al. 2019).

Biological circuits have also been used for the detection of environmental samples. In a recent study, the authors developed a strategy to build an optimized cell-based biosensor to detect toxic pollutants in environmental samples (Wan et al. 2019). They engineered multilayer amplifiers enabling a high signal-to-noise ratio detection through the transcriptional regulatory system. This promising approach provided facilities to build biosensors for arsenic and mercury with a very high fold-change response to the inducers. Thus, they were able to introduce a strategy to engineer sophisticated gene networks for in vivo diagnosis (Wan et al. 2019). In another work related to environmental diagnosis, a recent attempt used RNA output sensors activated by ligand induction (ROSALIND) in the cell-free system to detect pollutants in water (Alam et al. 2019). ROSALIND consists of three components: highly processive RNA polymerases, allosteric transcription factors, and synthetic DNA transcription templates. These elements together have provided the modular detection of a variety of water pollutants such as antibiotics, toxic small molecules, and metals (Alam et al. 2019).

2.3.2 Therapeutics

Synthetic biological networks provide a new generation of therapeutics called smart therapeutics. One of the earliest attempts was designing a synthetic mammalian circuit to maintain uric acid homeostasis (Kemmer et al. 2010). This synthetic gene

network consists of a uric acid sensor triggering the secretion of a urate oxidase enzyme which eliminates uric acid. In mice harboring this device, the synthetic circuit decreased the amount of blood urate and reduced uric acid crystal in the kidney (Kemmer et al. 2010). In a recent study, Isabella et al. (2018) provided a smart alternative for the protein-restricted diet for phenylketonuria, a genetic-metabolic disorder in metabolizing phenylalanine. For this purpose, the authors have engineered *Escherichia coli* Nissle to actuate phenylalanine metabolizing enzymes responding to anoxic conditions in the mammalian gut (Isabella et al. 2018). Designer circuits can be applied in the development of antimicrobials (Bikard et al. 2014; Bikard and Barrangou 2017), anticancers (Ding et al. 2015; Nissim et al. 2017; Liu et al. 2014; Prindle et al. 2012), microbiome editing (Ramachandran and Bikard 2019; Piraner et al. 2017), or medical imaging (Piraner et al. 2017; Farhadi et al. 2019; Lu et al. 2018; Bourdeau et al. 2018).

2.3.3 Metabolic Engineering

Utilizing synthetic gene networks for bioproduction application has rapidly grown during the last years. Genetic sensors have been applied in the field of metabolic/enzymatic engineering for (1) screening the enzymes and pathways, (2) monitoring the evolution of the products, and (3) dynamically regulating the enzymes or metabolites level (Liu et al. 2015, 2017; de Frias et al. 2018; Rogers et al. 2016; Koch et al. 2019; Venayak et al. 2015). This strategy substantially increases the speed of the design-build-test cycle in improving metabolic pathways and enzymes or exploring novel synthetic enzymes and pathways.

Synthetic gene circuits have shown an increasing potential to engineer dynamic regulation, regulatory cascades to dynamically control and improve the evolution of a product. The dynamic regulation improves the product yield either through directing metabolic fluxes into the direction of the desired product or by adjusting the expression of the enzymes and amount of intermediates as well as preventing the accumulation of a toxic intermediate (Venayak et al. 2015). One of the interests regarding metabolic engineering application is coupling cellular growth and product evolution, which can improve the production as it keeps a balance or controllable switch between growth and target production (Williams et al. 2015; Anesiadis et al. 2013; Gupta et al. 2017; He et al. 2017; Kim et al. 2017; Shong and Collins 2014). This coupling can be implemented using natural (native of the host cell) or synthetic quorum-sensing network regulating the expression of the enzymes in the metabolic pathway.

2.4 Design and Tools

The following are different cellular components providing the implementation of synthetic gene networks (bottom panel in Fig. 2.1).

2.4.1 Transcriptional Level

Undoubtedly, transcriptional regulators are the most studied tools to implement synthetic genetic networks for prokaryotic and eukaryotic applications (Nielsen et al. 2016; Khalil et al. 2012). Since transcriptional regulators are directly in contact with gene expression and DNA, and a number of these regulators are widely studied and characterized, utilizing them has become more scalable and programmable. In this direction, an enormous number of biological parts consisting of promoters, RBSs, terminators, and regulatory transcription factors have been characterized. These parts are characterized natural sequences, or they are synthetic sequences providing the orthogonality, which is of very crucial aspects in developing synthetic biological networks (Stanton et al. 2014; Chen et al. 2018b; Zong et al. 2017; Rudge et al. 2016). Moreover, the community has introduced methodologies for building, automizing, optimizing, and integrating various devices from simple gene networks to complex multilayer circuits (Nielsen et al. 2016; Zong et al. 2017; Rudge et al. 2016; Otero-Muras et al. 2016; Boada et al. 2019). Nielsen et al. have developed a tool called Cello using which complex relationships between a number of inputs could be computed through proposed circuits and the DNA sequence associated to those circuits is also generated (Nielsen et al. 2016).

Apart from transcriptional factors (including activators or repressors), CRISPR/dCas9 also have shown promising characteristics for synthesizing modular transcriptional regulators (Bikard et al. 2013; La Russa and Qi 2015; Kundert et al. 2019). The mutant version of Cas9 or other Cas nucleases which lack the nuclease activity but still maintain the specific binding through their designed gRNA can be used to target anywhere in the genome through highly specific binding of the gRNA-dCas9 complex to the target DNA (Rousset et al. 2018). By targeting desired sequences of the genome, gRNAs can simultaneously block several points in the genome acting as transcriptional repressors (Vigouroux et al. 2018). The CRISPR/dCas9 also can be fused to other proteins such as activators to regulate the activation (Dong et al. 2018; Matharu et al. 2019). There are computational and experimental tools to design such devices by tuning the level of binding through the complementarity of the gRNA and the target sequence (Vigouroux et al. 2018).

2.4.2 Translational Level

Translational regulators are components that control the translation of mRNA through the ribosome. RNA genetic switches or riboswitches are tools that regulate the gene expression in response to their input (Karagiannis et al. 2016; Robinson et al. 2016). However, some riboswitches function in the transcriptional processes such as in termination of the transportation (Serganov and Nudler 2013; Re 2017). Riboswitches consist of an aptamer (sensing) domain and an actuator (regulating) domain for binding to an input molecule and control the gene expression, respectively (Karagiannis et al. 2016; Wittmann and Suess 2012). The binding of an input

to its aptamer makes the actuator to alter the structure of the RNA, hence changing the translation process. A riboswitch can be actuated by a small molecule or another RNA sequence which in this case is called toehold circuit (Karagiannis et al. 2016; Liang et al. 2011).

Toehold circuits are RNA switches in which a short sequence of its input RNA regulates the expression of an mRNA (Green et al. 2014). The mRNA gene is designed to have a UTR sequence right upstream of the start codon that forms a secondary structure inhibiting the access of ribosomes to this mRNA (Chappell et al. 2017). At the presence of the input RNA, it opens up the secondary structure of mRNA by binding to the upstream sequence with higher affinity and exposes the RBS to ribosomes to be translated. By designing short sequences in the upstream of a reporter gene, different toehold circuits can be designed for input RNAs (a short RNA or a short sequence of a long RNA) (Pardee et al. 2014). Logic gates can be made by designing riboswitches, structure of which in their upstream is controlled by several inputs (Brophy and Voigt 2014; Rodrigo et al. 2012). Similar to toehold circuits, siRNAs also could be used to silence or inhibit mRNAs from translation (Matsuura et al. 2018; Na et al. 2013; Noh et al. 2017).

2.4.3 Others: DNA and Posttranslational Level

Gene networks can be programmably designed at DNA level by applying natural regulatory processes that occur on DNA. One of the main such tools are DNA switches enabled by recombinases (Bonnet et al. 2012; Chiu and Jiang 2017). Depending on their type, reversible or irreversible recombinases can be engineered with their specific recognition sites on DNA (Bonnet et al. 2012). A specific recombinase binds to its target site and flips a unidirectional terminator in front of a promoter to turn off/on a gene. Irreversible recombinases are tools to implement biological memory because they do not rely on the presence of the input after recombining their target sequence and turning on/off a gene (Bonnet et al. 2012; Rubens et al. 2016). These switches have also been applied to digitize or amplify the behavior of gene circuits (Bonnet et al. 2013). From simple devices to complex and multilayer gene circuits have been built using the recombinases strategy (Guiziou et al. 2018, 2019). There are multiple recombinases present in all kingdoms of life making them applicable in distinct cell hosts.

Although both prokaryotic and eukaryotic cells have signaling pathways (Parkinson 1993; Rhoads 1999), signal transduction is more a characteristic to the eukaryotic cell. The signal transduction is faster than gene circuits that function at DNA, transcriptional or translational level. This high speed is because it usually has only the outputs at gene expression level, and all the rest act in a transduction path of multiple components already expressed (Kiel et al. 2010). A recent study introduces a modular synthetic GPCR (G-protein-coupled receptor) signal transduction system that can be used to engineer GPCRs to respond to different ligands as inputs (Shaw et al. 2019). A famous example of bacterial signal transduction is the quorum

sensing of the bacteria, which is the sensory system to cellular populations (Gupta et al. 2017).

2.4.4 Cell-Free Systems as a New Platform

Cell-free systems are reliable platforms to test or implement synthetic genetic networks (Perez et al. 2016). These membrane-less and nucleic acid-free platforms are made up of cell extract plus additional elements to support the functionality of the system (Sun et al. 2013). Eukaryotic cell-free systems have shown only the ability of translation from mRNA added to the extract (Hartsough et al. 2015; Zemella et al. 2015). However, prokaryotic cell-free systems can perform both transcription and translation and thus work by adding only the coding DNA of the genes involved in the circuitry (Villarreal and Tan 2017). The cell-free system can be chemically defined and constructed in a bottom-up approach from required components called "purified recombinant elements" (PURE) system (Lavickova and Maerkl 2019). However, the PURE system is costly since everything should be provided to make the functional system. The alternative for this is TX-TL cell-free system made up by bacterial cell-lysate mixed with energy mix, amino acids, tRNAs, nucleotides, etc. (Sun et al. 2013). TX-TL systems have been applied as a chassis to build biosensors and genetic circuits, also for metabolic engineering application (Voyvodic et al. 2019; Koch et al. 2019; Dudley et al. 2015; Jiang et al. 2018).

Cell-free systems provide advantages over in vivo systems: (1) non-GMO platform to produce biological products and to build portable biosensors, (2) lower noise and higher precision as there is no growth and cellular maintenance (Chang et al. 2017), (3) rapid characterization of biological networks through quick mixing of the elements and fast expression of the circuits (Jiang et al. 2018), (4) possibility of adding linear DNA of PCR product (Sun et al. 2014; Marshall et al. 2017a; Nomoto and Tada 2018), (5) rapid cloning since there is no limitation of number of plasmids, origin of replications, and antibiotic resistance genes (Voyvodic et al. 2019), (6) higher number of genes can be used since there is less limitation of burden and resource competition (Pandi et al. 2019; Rustad et al. 2017), and (7) high tunability of the biological parts and system components as they can be altered by pipetting at any concentration (Pandi et al. 2019). Apart from above-mentioned applications, cell-free systems can also provide tools to study biological phenomena (Borkowski et al. 2018; Marshall et al. 2017b). Successful protocols have been developed to make cell-free lysate of different organism (Sun et al. 2013; Hartsough et al. 2015; Kelwick et al. 2016; Wang et al. 2014, 2018; Li et al. 2017; Moore et al. 2017, 2018; Des Soye et al. 2018; Wiegand et al. 2019). Also, optimization protocols have been shown ways in improving the functionality of cell-free systems in different condition (Caschera et al. 2011, 2018).

2.5 Perspectives

Synthetic gene networks are sophisticated tools to provide facilities in engineering biology. Since the dawn of synthetic biology, modular biological parts and methods have been increasingly equipped scientists toward a future in which cells and biological systems can be engineered for medical, environmental, and industrial applications (Brophy and Voigt 2014). The advances made so far have applied from genetic central dogma level to posttranslational, signal transduction, and metabolic enzymes in different prokaryotic cells, eukaryotic cells, and cell-free systems. Moreover, the experimental and computational approaches provide a potential perspective for the construction of next-generation synthetic biological networks. The next generation of such circuits is the integration of different tools and approaches for mix-hybrid gene circuit implementation (Brophy and Voigt 2014; Purcell and Lu 2014; Pandi et al. 2019; Goñi-Moreno and Nikel 2019; Rubens et al. 2016).

Decreasing the cost of the chemical synthesis and the sequencing of DNA provides a more affordable DNA reading and writing (sequencing and gene synthesis, respectively). Hence, the field of synthetic biology will be rapidly advancing through high-throughput experiments exploring the potential of the synthetic version of the code of life, DNA. The enormous available data of biological datasets and the future data that will be generated could be the training datasets for machine learning and deep learning exploration on these data to learn more and more about biology as well as to predict the future genetic networks (Camacho et al. 2018).

References

Alam KK, Jung JK, Verosloff MS, Clauer PR, Lee JW, Capdevila DA et al (2019) Rapid, low-cost detection of water contaminants using regulated in vitro transcription. bioRxiv:619296. https://doi.org/10.1101/619296

Anesiadis N, Kobayashi H, Cluett WR, Mahadevan R (2013) Analysis and design of a genetic circuit for dynamic metabolic engineering. ACS Synth Biol 2:442–452

Bikard D, Barrangou R (2017) Using CRISPR-Cas systems as antimicrobials. Curr Opin Microbiol 37:155–160. https://doi.org/10.1016/j.mib.2017.08.005

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437

Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X et al (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol 32:1146–1150

Boada Y, Vignoni A, Pico J (2019) Multiobjective identification of a feedback synthetic gene circuit. IEEE Trans Control Syst Technol:1–16.

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

Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D (2013) Amplifying genetic logic gates. Science 340:599–603

Borkowski O, Bricio C, Murgiano M, Rothschild-Mancinelli B, Stan G-B, Ellis T (2018) Cell-free prediction of protein expression costs for growing cells. Nat Commun 9:1457

Bourdeau RW, Lee-Gosselin A, Lakshmanan A, Farhadi A, Kumar SR, Nety SP et al (2018) Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. Nature 553:86–90

Bradley RW, Buck M, Wang B (2016) Recognizing and engineering digital-like logic gates and switches in gene regulatory networks. Curr Opin Microbiol 33:74–82

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

Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100:5136–5141

Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ (2018) Next-generation machine learning for biological networks. Cell 173:1581–1592

Caschera F, Bedau MA, Buchanan A, Cawse J, de Lucrezia D, Gazzola G et al (2011) Coping with complexity: machine learning optimization of cell-free protein synthesis. Biotechnol Bioeng 108:2218–2228

Caschera F, Karim AS, Gazzola G, d'Aquino AE, Packard NH, Jewett MC (2018) High-throughput optimization cycle of a cell-free ribosome assembly and protein synthesis system. ACS Synth Biol 7:2841–2853

Chang H-J, Voyvodic PL, Zúñiga A, Bonnet J (2017) Microbially derived biosensors for diagnosis, monitoring and epidemiology. Microb Biotechnol 10:1031–1035

Chappell J, Westbrook A, Verosloff M, Lucks JB (2017) Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. Nat Commun 8:1051

Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM et al (2018a) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360:436–439

Chen Y, Ho JML, Shis DL, Gupta C, Long J, Wagner DS et al (2018b) Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. Nat Commun 9:64

Chiu T-Y, Jiang J-HR (2017) Logic synthesis of recombinase-based genetic circuits. Sci Rep 7:12873

Courbet A, Endy D, Renard E, Molina F, Bonnet J (2015) Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. Sci Transl Med 7:289ra83

Courbet A, Amar P, Fages F, Renard E, Molina F (2018) Computer-aided biochemical programming of synthetic microreactors as diagnostic devices. Mol Syst Biol 14:e8441

Daniel R, Woo SS, Turicchia L, Sarpeshkar R (2011). Analog transistor models of bacterial genetic circuits. In 2011 IEEE biomedical circuits and systems conference (BioCAS) 2011. https://doi.org/10.1109/biocas.2011.6107795

Daniel R, Rubens JR, Sarpeshkar R, Lu TK (2013) Synthetic analog computation in living cells. Nature 497:619

de Frias UA, Pereira GKB, Guazzaroni M-E, Silva-Rocha R (2018) Boosting secondary metabolite production and discovery through the engineering of novel microbial biosensors. Biomed Res Int 2018:1

Delépine B, Libis V, Carbonell P, Faulon J-L (2016) SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Res 44:W226–W231

Delépine B, Duigou T, Carbonell P, Faulon J-L (2018) RetroPath2.0: a retrosynthesis workflow for metabolic engineers. Metab Eng 45:158–170

Deng W, Xu H, Ding W, Liang H (2014) DNA logic gate based on metallo-toehold strand displacement. PLoS One 9:e111650

Des Soye BJ, Davidson SR, Weinstock MT, Gibson DG, Jewett MC (2018) Establishing a high-yielding cell-free protein synthesis platform derived from Vibrio natriegens. ACS Synth Biol 7:2245–2255

Ding M, Li R, He R, Wang X, Yi Q, Wang W (2015) p53 activated by AND gate genetic circuit under radiation and hypoxia for targeted cancer gene therapy. Cancer Sci 106:1163–1173

Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG (2018) Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun 9:2489

Dudley QM, Karim AS, Jewett MC (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J 10:69–82

Engelen W, Meijer LHH, Somers B, de Greef TFA, Merkx M (2017) Antibody-controlled actuation of DNA-based molecular circuits. Nat Commun 8:14473
Farhadi A, Ho GH, Sawyer DP, Bourdeau RW, Shapiro MG (2019) Ultrasound imaging of gene expression in mammalian cells. bioRxiv:580647. https://doi.org/10.1101/580647
Farzadfard F, Lu TK (2014) Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science 346:1256272
Fernandez-Rodriguez J, Voigt CA (2016) Post-translational control of genetic circuits using Potyvirus proteases. Nucleic Acids Res 44:6493–6502
Fink T, Lonzarić J, Praznik A, Plaper T, Merljak E, Leben K et al (2019) Design of fast proteolysis-based signaling and logic circuits in mammalian cells. Nat Chem Biol 15:115–122
Gander MW, Vrana JD, Voje WE, Carothers JM, Klavins E (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nat Commun 8:15459
Gao XJ, Chong LS, Kim MS, Elowitz MB (2018) Programmable protein circuits in living cells. Science 361:1252–1258
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403:339–342
Genot AJ, Bath J, Turberfield AJ (2011) Reversible logic circuits made of DNA. J Am Chem Soc 133:20080–20083
Goñi-Moreno A, Nikel PI (2019) High-performance biocomputing in synthetic biology-integrated transcriptional and metabolic circuits. Front Bioeng Biotechnol 7:40
Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J et al (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356:438–442
Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 360:439–444
Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold switches: de-novo-designed regulators of gene expression. Cell 159:925–939
Guiziou S, Ulliana F, Moreau V, Leclere M, Bonnet J (2018) An automated design framework for multicellular recombinase logic. ACS Synth Biol 7:1406–1412
Guiziou S, Mayonove P, Bonnet J (2019) Hierarchical composition of reliable recombinase logic devices. Nat Commun 10:456
Gupta A, Reizman IMB, Reisch CR, Prather KLJ (2017) Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. Nat Biotechnol 35:273–279
Hartsough EM, Shah P, Larsen AC, Chaput JC (2015) Comparative analysis of eukaryotic cell-free expression systems. Bio Techniques 59:149–151
Haykin SO (2011) Neural networks and learning machines. Pearson Education, Upper Saddle River
He X, Chen Y, Liang Q, Qi Q (2017) Autoinduced AND gate controls metabolic pathway dynamically in response to microbial communities and cell physiological state. ACS Synth Biol 6:463–470
Isabella VM, Ha BN, Castillo MJ, Lubkowicz DJ, Rowe SE, Millet YA et al (2018) Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. Nat Biotechnol 36:857–864
Jiang L, Zhao J, Lian J, Xu Z (2018) Cell-free protein synthesis enabled rapid prototyping for metabolic engineering and synthetic biology. Synth Syst Biotechnol 3:90–96
Karagiannis P, Fujita Y, Saito H (2016) RNA-based gene circuits for cell regulation. Proc Jpn Acad Ser B Phys Biol Sci 92:412–422
Katz E (2017) Enzyme-based logic gates and networks with output signals analyzed by various methods. Chem Phys Chem 18:1688–1713
Kelwick R, MacDonald JT, Webb AJ, Freemont P (2014) Developments in the tools and methodologies of synthetic biology. Front Bioeng Biotechnol 2:60
Kelwick R, Webb AJ, MacDonald JT, Freemont PS (2016) Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements. Metab Eng 38:370–381

Kemmer C, Gitzinger M, Daoud-El Baba M, Djonov V, Stelling J, Fussenegger M (2010) Self-sufficient control of urate homeostasis in mice by a synthetic circuit. Nat Biotechnol 28:355–360
Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK et al (2012) A synthetic biology framework for programming eukaryotic transcription functions. Cell 150:647–658
Kiel C, Yus E, Serrano L (2010) Engineering signal transduction pathways. Cell 140:33–47. https://doi.org/10.1016/j.cell.2009.12.028
Kim E-M, Woo HM, Tian T, Yilmaz S, Javidpour P, Keasling JD et al (2017) Autonomous control of metabolic state by a quorum sensing (QS)-mediated regulator for bisabolene production in engineered E. coli. Metab Eng 44:325–336
Kim J, Yin P, Green AA (2018) Ribocomputing: cellular logic computation using RNA devices. Biochemistry 57:883–885
Kim H, Bojar D, Fussenegger M (2019) A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells. Proc Natl Acad Sci U S A 116:7214–7219
Koch M, Pandi A, Delépine B, Faulon J-L (2018) A dataset of small molecules triggering transcriptional and translational cellular responses. Data Brief 17:1374–1378
Koch M, Pandi A, Borkowski O, Cardoso Batista A, Faulon J-L (2019) Custom-made transcriptional biosensors for metabolic engineering. Curr Opin Biotechnol 59:78–84
Kundert K, Lucas JE, Watters KE, Fellmann C, Ng AH, Heineike BM et al (2019) Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. Nat Commun 10:2127
Kushwaha M, Salis HM (2015) A portable expression resource for engineering cross-species genetic circuits and pathways. Nat Commun 6. https://doi.org/10.1038/ncomms8832
La Russa MF, Qi LS (2015) The new state of the art: Cas9 for gene activation and repression. Mol Cell Biol 35:3800–3809
Lavickova B, Maerkl SJ (2019) A simple, robust, and low-cost method to produce the PURE cell-free system. ACS Synth Biol 8:455–462
Li J, Wang H, Kwon Y-C, Jewett MC (2017) Establishing a high yielding streptomyces-based cell-free protein synthesis system. Biotechnol Bioeng 114:1343–1353
Li S-Y, Cheng Q-X, Wang J-M, Li X-Y, Zhang Z-L, Gao S et al (2018) CRISPR-Cas12a-assisted nucleic acid detection. Cell Discov 4:20
Liang JC, Bloom RJ, Smolke CD (2011) Engineering biological systems with synthetic RNA molecules. Mol Cell 43:915–926. https://doi.org/10.1016/j.molcel.2011.08.023
Liu Y, Zeng Y, Liu L, Zhuang C, Fu X, Huang W et al (2014) Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. Nat Commun 5:5393
Liu D, Evans T, Zhang F (2015) Applications and advances of metabolite biosensors for metabolic engineering. Metab Eng 31:35–43
Liu Y, Liu Y, Wang M (2017) Design, optimization and application of small molecule biosensor in metabolic engineering. Front Microbiol 8:2012
Lu GJ, Farhadi A, Mukherjee A, Shapiro MG (2018) Proteins, air and water: reporter genes for ultrasound and magnetic resonance imaging. Curr Opin Chem Biol 45:57–63
MacDonald JT, Barnes C, Kitney RI, Freemont PS, G-BV S (2011) Computational design approaches and tools for synthetic biology. Integr Biol 3:97. https://doi.org/10.1039/c0ib00077a
Marchisio MA (2014) Parts & pools: a framework for modular design of synthetic gene circuits. Front Bioeng Biotechnol 2:42
Marshall R, Maxwell CS, Collins SP, Beisel CL, Noireaux V (2017a) Short DNA containing χ sites enhances DNA stability and gene expression in E. coli cell-free transcription-translation systems. Biotechnol Bioeng 114:2137–2141. https://doi.org/10.1002/bit.26333
Marshall R, Maxwell CS, Collins SP, Luo ML, Jacobsen T, Beisel C et al (2017b) Rapid and scalable characterization of CRISPR technologies using an E. coli cell-free transcription-translation system. https://doi.org/10.1101/169441
Martini L, Meyer AJ, Ellefson JW, Milligan JN, Forlin M, Ellington AD et al (2015) In vitro selection for small-molecule-triggered strand displacement and riboswitch activity. ACS Synth Biol 4:1144–1150

Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A et al (2019) CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. Science 363. https://doi.org/10.1126/science.aau0629

Matsuura S, Ono H, Kawasaki S, Kuang Y, Fujita Y, Saito H (2018) Synthetic RNA-based logic computation in mammalian cells. Nat Commun 9:4847

Moore SJ, Lai H-E, Needham H, Polizzi KM, Freemont PS (2017) Streptomyces venezuelae TX-TL - a next generation cell-free synthetic biology tool. Biotechnol J 12:1600678. https://doi.org/10.1002/biot.201600678s

Moore SJ, MacDonald JT, Wienecke S, Ishwarbhai A, Tsipa A, Aw R et al (2018) Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. Proc Natl Acad Sci 115:E4340–E4349. https://doi.org/10.1073/pnas.1715806115

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 31:170–174

Nielsen AA, Voigt CA (2014) Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. Mol Syst Biol 10:763–763. https://doi.org/10.15252/msb.20145735

Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA et al (2016) Genetic circuit design automation. Science 352:aac7341

Nissim L, Wu M-R, Pery E, Binder-Nissim A, Suzuki HI, Stupp D et al (2017) Synthetic RNA-based immunomodulatory gene circuits for cancer immunotherapy. Cell 171:1138–50.e15

Noh M, Yoo SM, Kim WJ, Lee SY (2017) Gene expression knockdown by modulating synthetic small RNA expression in Escherichia coli. Cell Syst 5:418–26.e4

Nomoto M, Tada Y (2018) Cloning-free template DNA preparation for cell-free protein synthesis via two-step PCR using versatile primer designs with short 3'-UTR. Genes Cells 23:46–53

Nowogrodzki A (2018) The automatic-design tools that are changing synthetic biology. Nature 564:291–292

Otero-Muras I, Henriques D, Banga JR (2016) SYNBADm: a tool for optimization-based automated design of synthetic gene circuits. Bioinformatics 32:3360–3362

Pandi A, Koch M, Voyvodic PL, Soudier P, Bonnet J, Kushwaha M et al (2019) Metabolic perceptrons for neural computing in biological systems. Nat Commun 10(1):3880. https://doi.org/10.1038/s41467-019-11889-0

Pardee K, Green AA, Ferrante T, Cameron DE, DaleyKeyser A, Yin P et al (2014) Paper-based synthetic gene networks. Cell 159:940–954

Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW et al (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165:1255–1266

Parkinson JS (1993) Signal transduction schemes of bacteria. Cell 73:857–871. https://doi.org/10.1016/0092-8674(93)90267-t

Perez JG, Stark JC, Jewett MC (2016) Cell-free synthetic biology: engineering beyond the cell. Cold Spring Harb Perspect Biol 8. https://doi.org/10.1101/cshperspect.a023853

Piraner DI, Abedi MH, Moser BA, Lee-Gosselin A, Shapiro MG (2017) Tunable thermal bioswitches for in vivo control of microbial therapeutics. Nat Chem Biol 13:75–80

Prindle A, Selimkhanov J, Danino T, Samayoa P, Goldberg A, Bhatia SN et al (2012) Genetic circuits in Salmonella typhimurium. ACS Synth Biol 1:458–464

Purcell O, Lu TK (2014) Synthetic analog and digital circuits for cellular computation and memory. Curr Opin Biotechnol 29:146–155

Purnick PEM, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422

Ramachandran G, Bikard D (2019) Editing the microbiome the CRISPR way. Philos Trans R Soc Lond Ser B Biol Sci 374:20180103

Razavi S, Su S, Inoue T (2014) Cellular signaling circuits interfaced with synthetic, post-translational, negating Boolean logic devices. ACS Synth Biol 3:676–685

Re A (2017) Synthetic gene expression circuits for designing precision tools in oncology. Front Cell Dev Biol 5:77

Rhoads RE (1999) Signal transduction pathways that regulate eukaryotic protein synthesis. J Biol Chem 274:30337–30340. https://doi.org/10.1074/jbc.274.43.30337

Robinson CJ, Medina-Stacey D, Wu M-C, Vincent HA, Micklefield J (2016) Rewiring riboswitches to create new genetic circuits in Bacteria. Methods Enzymol 575:319–348
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 U S A 109:15271–15276
Rogers JK, Taylor ND, Church GM (2016) Biosensor-based engineering of biosynthetic pathways. Curr Opin Biotechnol 42:84–91
Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408
Rosier BJ, de Greef TFA (2015) How to make an oscillator. elife 4. https://doi.org/10.7554/elife.12260
Rousset F, Cui L, Siouve E, Becavin C, Depardieu F, Bikard D (2018) Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLoS Genet 14:e1007749. https://doi.org/10.1371/journal.pgen.1007749
Rubens JR, Selvaggio G, Lu TK (2016) Synthetic mixed-signal computation in living cells. Nat Commun 7:11658
Rudge TJ, Brown JR, Federici F, Dalchau N, Phillips A, Ajioka JW et al (2016) Characterization of intrinsic properties of promoters. ACS Synth Biol 5:89–98
Rustad M, Eastlund A, Marshall R, Jardine P, Noireaux V (2017) Synthesis of infectious bacteriophages in an E. coli-based cell-free expression system. J Vis Exp (126). https://doi.org/10.3791/56144
Sarpeshkar R (2014) Analog synthetic biology. Philos Trans R Soc A Math Phys Eng Sci 372:20130110
Sauro HM, Kim K (2013) It's an analog world. Nature 497:572–573. https://doi.org/10.1038/nature12246
Selberg J, Gomez M, Rolandi M (2018) The potential for convergence between synthetic biology and bioelectronics. Cell Syst 7:231–244
Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152:17–24
Shaw WM, Yamauchi H, Mead J, GOF G, Bell DJ, Öling D et al (2019) Engineering a model cell for rational tuning of GPCR signaling. Cell 177:782–96.e27
Shong J, Collins CH (2014) Quorum sensing-modulated AND-gate promoters control gene expression in response to a combination of endogenous and exogenous signals. ACS Synth Biol 3:238–246
Slomovic S, Pardee K, Collins JJ (2015) Synthetic biology devices for in vitro and in vivo diagnostics. Proc Natl Acad Sci U S A 112:14429–14435
Stanton BC, Nielsen AAK, Tamsir A, Clancy K, Peterson T, Voigt CA (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. Nat Chem Biol 10:99–105
Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J Vis Exp. https://doi.org/10.3791/50762
Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM (2014) Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. ACS Synth Biol 3:387–397. https://doi.org/10.1021/sb400131a
Takahashi MK, Tan X, Dy AJ, Braff D, Akana RT, Furuta Y et al (2018) A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. Nat Commun 9:3347
Venayak N, Anesiadis N, Cluett WR, Mahadevan R (2015) Engineering metabolism through dynamic control. Curr Opin Biotechnol 34:142–152
Vigouroux A, Oldewurtel E, Cui L, Bikard D, van Teeffelen S (2018) Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. Mol Syst Biol 14:e7899
Villarreal F, Tan C (2017) Cell-free systems in the new age of synthetic biology. Front Chem Sci Eng 11:58–65. https://doi.org/10.1007/s11705-017-1610-x

Voyvodic PL, Pandi A, Koch M, Conejero I, Valjent E, Courtet P et al (2019) Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. Nat Commun 10:1697

Wan X, Volpetti F, Petrova E, French C, Maerkl SJ, Wang B (2019) Cascaded amplifying circuits enable ultrasensitive cellular sensors for toxic metals. Nat Chem Biol 15:540–548

Wang X, Zhao L, Zhao K-N (2014) An optimized yeast cell-free lysate system for in vitro translation of human virus mRNA. Methods Mol Biol 1118:219–230

Wang B, Barahona M, Buck M (2015) Amplification of small molecule-inducible gene expression via tuning of intracellular receptor densities. Nucleic Acids Res 43:1955–1964

Wang H, Li J, Jewett MC (2018) Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. Synth Biol 3:ysy003. https://doi.org/10.1093/synbio/ysy003

Wen KY, Cameron L, Chappell J, Jensen K, Bell DJ, Kelwick R et al (2017) A cell-free biosensor for detecting quorum sensing molecules in P. aeruginosa-infected respiratory samples. ACS Synth Biol 6:2293–2301. https://doi.org/10.1021/acssynbio.7b00219

Wiegand DJ, Lee HH, Ostrov N, Church GM (2019) Cell-free protein expression using the rapidly growing bacterium Vibrio natriegens. J Vis Exp (145). https://doi.org/10.3791/59495

Williams TC, Averesch NJH, Winter G, Plan MR, Vickers CE, Nielsen LK et al (2015) Quorum-sensing linked RNA interference for dynamic metabolic pathway control in Saccharomyces cerevisiae. Metab Eng 29:124–134. https://doi.org/10.1016/j.ymben.2015.03.008

Wittmann A, Suess B (2012) Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. FEBS Lett 586:2076–2083

Wu MJ, Andreasson JOL, Kladwang W, Greenleaf W, Das R (2019) Automated design of highly diverse riboswitches. bioRxiv:603001. https://doi.org/10.1101/603001

Xiao Y, Jiang W, Zhang F (2017) Developing a genetically encoded, cross-species biosensor for detecting ammonium and regulating biosynthesis of cyanophycin. ACS Synth Biol 6:1807–1815

Zemella A, Thoring L, Hoffmeister C, Kubick S (2015) Cell-free protein synthesis: pros and cons of prokaryotic and eukaryotic systems. Chembiochem 16:2420–2431

Zeng J, Teo J, Banerjee A, Chapman TW, Kim J, Sarpeshkar R (2018) A synthetic microbial operational amplifier. ACS Synth Biol 7:2007–2013

Zong Y, Zhang HM, Lyu C, Ji X, Hou J, Guo X et al (2017) Insulated transcriptional elements enable precise design of genetic circuits. Nat Commun 8:52

Current Progress and Limitations in the Design, Construction, and Characterization of Synthetic Parts

3

Vinuselvi Ponraj

Abstract

The field of synthetic biology has grown multidimensionally that we now have a large collection of interchangeable input and output modules. Design and construction of new synthetic parts are no longer a challenging task. However, the performance of a synthetic part generally has a lower accuracy than the corresponding natural system. Characterization of synthetic parts poses the actual drawback as most of the times these modules are studied in isolation and are expected to produce the same result when put together as a part of a large circuit or transferred from one chassis organism to the other. It becomes necessary to develop robust mathematical models and conquer the quantitative aspect of the synthetic parts which could then help improve the performance of the synthetic circuit. In this review, we brief the status and limitations of the design, construction, and characterization of synthetic parts and use "oscillators" as a case study to emphasize the betterment of the abstraction of the synthetic part either in isolation or in combination. The simple "oscillator circuit" was improved over a decade from being a barely visible oscillator to the one that can oscillate for up to 15 generations. The oscillator circuit forms a stand-alone example for the need for the harmony of stochastic chemistry and synthetic biology to achieve the long-standing goal of well-characterized genetic parts analogous to the electronic circuit.

Keywords

Synthetic parts · Quantitative biology · Mathematical models · Oscillator · Decoys

V. Ponraj (✉)
UMASS Amherst, Programs in Systems Biology, University of Massachusetts Medical School, Worcester, MA, USA
e-mail: Vinuselvi.Parisutham@umassmed.edu

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_3

3.1 Introduction

Realizing the presence of logical operations in the biological circuits is the biggest motivation for the birth of new field of science: "*Synthetic Biology*." The field was inspired by the promise to design well-characterized biological parts that when put together within a living cell could produce favorable and highly predictable output analogs to electronic circuits. Thanks to the efforts of the scientific community across the globe, there is a vast collection of interchangeable synthetic biology toolkits for use in wide range of organisms including both prokaryotes and eukaryotes. In addition, genomic sequencing techniques are gearing up the growth of synthetic parts with the revolutionary collection of sequence data from the depths of oceans to the heights of mountains. Genome engineering and gene cloning techniques (reviewed in Chap. 17 of this book) have now made it possible to make newer synthetic parts within few hours or assembly even the whole genome of an organism easily. Synthetic parts have taken a tremendous improvement from being based on simple DNA-binding transcription factors to those involving complex regulations mediated by sRNA and ligand-controlled riboswitches. Synthetic biology finds its grounding needs in several fields including but not limited to the production of cheaper therapeutics, biofuels, discovery of new class of antibiotics, tissue engineering, targeted drug delivery in the tissue of interest, in bioremediation, etc. (Khalil et al. 2012; Park et al. 2019; Khalil and Collins 2010).

The robust nature of the living systems combined with our limited ability to understand the interplay between gene expression pathways is a potent bottleneck for the rapid growth of biologically inspired synthetic parts. It remains a bold ambition to make independent biological parts to work analogs to electronic circuitry where individual parts' output is accurately predicted. For instance, the technical "0" in an electronic device means the complete absence of input (meaning they have a distinct ON/OFF states) whereas in a biological part "0" translates to the absence of a signal while still experiencing a basal metabolic rate (meaning they produce significant signal in OFF state). In addition to the basal metabolic rate there is a stochastic difference (both intrinsic to the cells and extrinsic from the environment) in independent cells which when put together results in unstable initial (and final) states of the biological parts translated as a technical "0." Inspired by the challenges in the theoretical predictions of the output of a synthetic part, newer mathematical frameworks have been developed to assist the characterization of synthetic parts (Brophy and Voigt 2014; Xiang et al. 2018; Xia et al. 2019). In this review, we highlight the importance of the interplay of synthetic and systems biology to improve the design and characterization of existing biological parts.

3.2 Design of Synthetic Parts

3.2.1 Input Module

A core set of three repressors (CI, TetR, and LacI) were used to design several of the initial group of synthetic circuits. Combination of inputs has increased tremendously: LacI orthologs, TetR orthologs (Stanton et al. 2014), riboswitches, light-induced promoter systems (Levskaya et al. 2005), zinc-finger transcription factors, transcription activator like effectors, several classes of constitute and inducible promoters, T7 RNAP, quorum sensing system, recombinases, etc. Several synthetic circuits have been designed and implemented using a combination of these input modules. Few of the synthetic circuits are listed in Table 3.1. The intrinsic feature of any input module is the proteins involved should be orthogonal, i.e., they cannot cross-talk with each other and with the existing biological machine.

While there are several successful attempts to develop synthetic circuits there are few potential issues to be taken care while choosing the right combination of the input modules. Multiple input module can cause acute toxicity in the cell especially if they are expressed at high copy numbers from the plasmid. It becomes too difficult to counteract the leaky level and steady state expression of multiple inputs. The large repository of input modules challenges the selection of right components for a given synthetic circuit. Components that work in larger dynamic range and with distinct OFF state expression are the high choice for most circuit design. However, there are only handful of components that satisfy these criteria. Computational tools like RBS Designer (Reeve et al. 2014) and promoter analyzer can come in handy to improve the existing array of input module to achieve larger dynamic range, lesser leaky level, and higher threshold/steady state values.

Table 3.1 Examples of common synthetic parts

Synthetic parts	Features	Reference
Pulse generator	Feed forward motif causes spatiotemporal gene expression in bacterial community	Basu et al. (2004)
Bistable switches	Genetic switches with two stable steady states	Lebar et al. (2014)
Oscillators	Circular negative feedback loops function as a biological clock	Elowitz and Leibler (2000)
Logic gates	AND, NOT, OR, and almost any logical operation	Stanton et al. (2014)
Edge detection	Used to detect boundaries of an object in an image	Tabor et al. (2009)
Cancer classifier circuit	Selectively triggers apoptosis in HeLa cells using microRNAs as input module	Xie et al. (2011)
Whole cell biosensors	Visualize and diagnose cancer metastasis	Danino et al. (2015)
Drug-induced kill switches	T cell therapies	Budde et al. (2013)

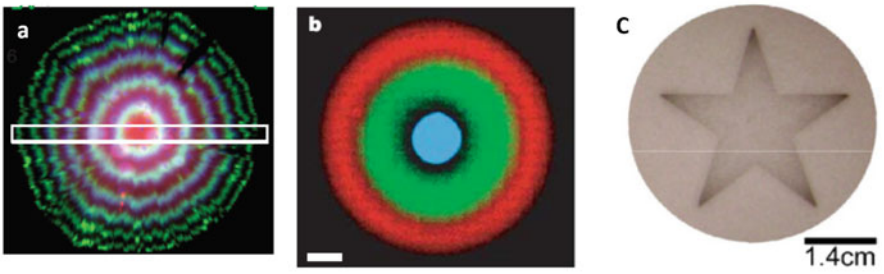


Fig. 3.1 Examples of achievements in different synthetic circuits. A. Synchronized oscillation seen in a colony of E. coli (Potvin-Trottier et al. 2016). B. Pattern formation in *E. coli* (Basu et al. 2005). C. Edge detection in *E. coli* (Tabor et al. 2009)

Several well-defined synthetic parts are in use today and are named based on the inspiration from the analogous electronic devices. Genetic switches work based on the principle of logical NOR gate with two stable steady states and a genetic memory of the current state. Oscillators use circular negative feedback loops to produce an alternating sinusoidal output. Oscillatory circuits are so common form of natural circuits with circadian rhythm patterns being the well-known example. Few of the examples commonly used synthetic parts and its characteristic features are listed in Table 3.1 and in Fig. 3.1.

3.2.2 Output Reporters

Important aspect of the design of a synthetic part is the presence of a readable output. Widely available readable output includes fluorescence reporters (Suel 2011), luminescence, enzymatic assays, and cell fitness. Fluorescence outputs are a popular choice because it is available in different colors and works well in most organisms. Drawbacks in using fluorescent output are as follows: fluorescence signal across different labs is incomparable because it depends largely on the optical property of the instrument (microscope, plate readers, FACS) used to read the signal, cross talk between fluorescence protein is highly troublesome when more than one fluorescent reporter is used in the same circuit, fluorescence is sensitive to changes in oxygen level, pH, and temperature. Luminescence on the other hand is robust reporter but involves ATP as source of energy and is not an attractive choice among scientist. Enzymatic assays cannot be used to obtain a dynamic profile of the performance of the synthetic part. Most synthetic parts do not impose fitness defects in the organism of choice and hence, cellular fitness is a less common output readout. Surprisingly limited availability of the reporter choice is a biggest limitation in obtaining characteristic details of complex synthetic parts.

3.3 Measurement Techniques

The sensitivity, robustness, and multiplexing of the measurement techniques have made significant progress in the recent years. To begin with only bulk measurements of the output readout were possible which failed to address the actual stochasticity in the data. Advancements in fluorescence microscopy and fluorescence activated cell sorting (FACS) helped look at individual cell's behavior. With the advent of microfluidics and mother machine (Okumus et al. 2018; Potvin-Trottier et al. 2018) it is now possible to monitor the same cell for several generations and even ask cells to process dynamic stimulus. We are starting to realize that most outputs of the bulk measurements are the underpredictions of the stochastic cell-to-cell variations rather than the actual underperformance of the new devised synthetic part. The faster growth of the field of stochastic chemistry and biological noise helps in deeper ability to model the response and behavior of the gene expression. Harmony between the characterization of the synthetic parts and the stochastic chemistry is important to achieve a repository of well-defined genetic parts.

3.4 Case Study: Progress and Limitations in Oscillator Circuit

A simple oscillatory circuit is built by three genes inhibiting each other's production in a single loop (A⊣B⊣C⊣A). TetR, CI, and LacI are expressed on plasmids to achieve oscillation in *E. coli*. However, in the initial attempt a noisy oscillation that lasted for only a few generations was observed (Elowitz and Leibler 2000). The same system was proven to be an effective harmonic oscillator in cell-free extracts (Niederholtmeyer et al. 2015). In another attempt oscillation was tried using the light-inducible promoter system. Long lifetime of the fluorescent proteins resulted in step-wise increase in the output signal instead of the expected sinusoidal harmonic oscillator (Lee et al. 2013). The circuit constructed by Elowitz and Leibler 2000 was later revisited and tested on a microfluidic platform and was shown to be oscillating more uniformly rather than the reported noisy oscillation (Danino et al. 2010). Soon after the brainstorming inputs from stochastic theory it was realized that the three input modules and the output modules would perform better when expressed from a single low copy plasmid rather than two separate plasmids for the input and the output modules as in the original design. Having more than one plasmid type in a cell will raise a stochastic competition between the two plasmid types resulting in greater cell-to-cell variation. Combining the input and output modules in a single plasmid did improve the performance of the oscillators but instead of the expected harmonic oscillatory response, a relaxed oscillation was observed. Again, based on the stochastic theory this could be simple because one of the three proteins (TetR) had a low threshold value and hence the other two proteins (LacI and CI) overrule the circuit for a while until the third protein was sufficiently accumulated. To overcome this TetR threshold was increased by simply adding competing binding sites for TetR on a plasmid theoretically raising the threshold

value of TetR (Berg et al. 2000). Such an improvement helps in synchronizing the entire system and harmonic oscillation was observed for several generations (Potvin-Trottier et al. 2016). Thus, the implementation of mathematical knowledge in improving the choice and design of synthetic part helped achieve the long-standing goal of synthetic biology: a well characterized harmonic oscillator which is now finding huge application in tracking the number of generations bacteria could take in uncommon environments like the human gut and can be used for programmed cell-death in whole cell biosensors after certain generations.

3.5 Conclusions

Relatively newer field of synthetic biology was once in a quest for the collection of large repositories of standard biological parts. No sooner in the growth phase of this field, it became necessary to just collect random functional biological parts and the definition of well-characterizing them stayed in the text. Now it is high time to gear back and start characterizing the individual synthetic circuits in greater details to move forward in the field of synthetic biology. At least with the oscillators we realized the low copy of any component results in randomizing the behavior and increasing the threshold helps achieve regularity. Several unique rules of this sense must be made using mathematical frameworks to improve the design and construction of synthetic parts.

Acknowledgement We would like to thank Dr. Vijai Singh for his valuable comments.

Conflict of Interest The author declares no conflict of interest.

References

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A 101(17):6355–6360

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434(7037):1130–1134

Berg OG, Paulsson J, Ehrenberg M (2000) Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. Biophys J 79(3):1228–1236

Brophy JA, Voigt CA (2014) Principles of genetic circuit design. Nat Methods 11(5):508–520

Budde LE, Berger C, Lin Y, Wang J, Lin X, Frayo SE et al (2013) Combining a CD20 chimeric antigen receptor and an inducible caspase 9 suicide switch to improve the efficacy and safety of T cell adoptive immunotherapy for lymphoma. PLoS One 8(12):e82742

Danino T, Mondragon-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. Nature 463(7279):326–330

Danino T, Prindle A, Kwong GA, Skalak M, Li H, Allen K et al (2015) Programmable probiotics for detection of cancer in urine. Sci Transl Med 7(289):289ra84

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338

Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11(5):367–379

Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK et al (2012) A synthetic biology framework for programming eukaryotic transcription functions. Cell 150(3):647–658

Lebar T, Bezeljak U, Golob A, Jerala M, Kadunc L, Pirs B et al (2014) A bistable genetic switch based on designable DNA-binding domains. Nat Commun 5:5007

Lee JM, Lee J, Kim T, Lee SK (2013) Switchable gene expression in Escherichia coli using a miniaturized photobioreactor. PLoS One 8(1):e52382

Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M et al (2005) Synthetic biology: engineering Escherichia coli to see light. Nature 438(7067):441–442

Niederholtmeyer H, Sun ZZ, Hori Y, Yeung E, Verpoorte A, Murray RM et al (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. eLife 4:e09771

Okumus B, Baker CJ, Arias-Castro JC, Lai GC, Leoncini E, Bakshi S et al (2018) Single-cell microscopy of suspension cultures using a microfluidics-assisted cell screening platform. Nat Protoc 13(1):170–194

Park M, Patel N, Keung AJ, Khalil AS (2019) Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. Cell 176(1–2):238.e20

Potvin-Trottier L, Lord ND, Vinnicombe G, Paulsson J (2016) Synchronous long-term oscillations in a synthetic gene circuit. Nature 538(7626):514–517

Potvin-Trottier L, Luro S, Paulsson J (2018) Microfluidics and single-cell microscopy to study stochastic processes in bacteria. Curr Opin Microbiol 43:186–192

Reeve B, Hargest T, Gilbert C, Ellis T (2014) Predicting translation initiation rates for designing synthetic biology. Front Bioeng Biotechnol 2:1

Stanton BC, Nielsen AA, Tamsir A, Clancy K, Peterson T, Voigt CA (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. Nat Chem Biol 10(2):99–105

Suel G (2011) Use of fluorescence microscopy to analyze genetic circuit dynamics. Methods Enzymol 497:275–293

Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM et al (2009) A synthetic genetic edge detection program. Cell 137(7):1272–1281

Xia PF, Ling H, Foo JL, Chang MW (2019) Synthetic genetic circuits for programmable biological functionalities. Biotechnol Adv. 37:107393

Xiang Y, Dalchau N, Wang B (2018) Scaling up genetic circuit design for cellular computing: advances and prospects. Nat Comput 17(4):833–853

Xie Z, Wroblewska L, Prochazka L, Weiss R, Benenson Y (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells. Science 333(6047):1307–1311

Recent Progress in DNA Parts Standardization and Characterization

4

Shalini S. Deb and Shamlan M. S. Reshamwala

Abstract

Synthetic Biology, which is the 'engineering of biology', depends on well-characterized and standard genetic elements that can be assembled together to construct complex, multi-component genetic circuits that function in a context-independent and predictable fashion. Here, we describe widely used standards employed for constructing DNA parts, and also discuss key assembly methods that can be used to build genetic devices starting from standard parts. Methods used to characterize parts and devices are discussed, and the finally, the need for 'functional standards' is outlined.

Keywords

Part standards · Assembly methods · Genetic devices · Synthetic biology

4.1 Introduction

The implementation of synthetic biology, which is an amalgam of molecular biology and engineering principles, depends on well-characterized genetic elements (e.g., coding sequences (CDS), promoters, ribosome-binding sites (RBS) and transcription terminators) that can be assembled together in different combinations to form genetic circuits that can then be tested. In line with the engineering paradigm that informs the discipline of synthetic biology, there have been attempts to define standards that genetic elements (or 'parts') can conform to, allowing them to be easily joined to each other like nuts and bolts.

S. S. Deb (✉) · S. M. S. Reshamwala
DBT-ICT Centre for Energy Biosciences, Institute of Chemical Technology, Mumbai, India
e-mail: ss.deb@ictmumbai.edu.in; ss.reshamwala@ictmumbai.edu.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_4

This chapter will review the different standards that have been proposed for the construction of genetic parts, and their relative advantages and disadvantages. As standardization is aimed at ease of assembly, the various assembly techniques that are employed to join parts are also discussed. Approaches used to characterize parts and devices are described. The chapter ends by proposing the need for a new standard that will enable functional comparison of genetic parts, a feature that is currently poor addressed.

4.2 The Need for Standardization of Genetic Parts

Traditional molecular cloning workflows rely on generating sites for restriction endonucleases on either ends of DNA fragments, usually by means of polymerase chain reaction (PCR)-mediated primer extension. This strategy is constrained by the availability of suitable restriction sites in the vector backbone, and of course, the sequence of the DNA fragment to be cloned. While this is usually not an issue for most routine cloning exercises, cloning more than one gene in a single vector (to generate a synthetic operon, for example) or cloning large DNA fragments can be severely hampered due to the increased probability of finding restriction sites within the DNA to be cloned.

One way to circumvent this issue is to modify the DNA sequence to remove restriction sequences that are found within DNA fragments to be cloned. Site-directed mutagenesis can be employed to mutate such sites, or the DNA fragment can be chemically synthesized with the offending sites suitably altered. Both these strategies have their associated disadvantages, which have led researchers to come up with cloning strategies that, to varying degrees, circumvent the problems that traditional cloning strategies pose.

In the synthetic biology approach to molecular cloning, DNA fragments are treated as ‘parts’ or ‘modules’ that can be joined to each other in a standard fashion. Genetic elements are flanked by DNA sequences that are compatible with ends of other DNA sequences, making it possible to stitch together parts in a defined and precise manner. Such treatment of DNA parts opens up possibilities that are difficult, if not impossible, to imagine with traditional cloning strategies.

Standardized structural and functional composition of biological parts is one of the goals of the discipline of biological engineering (Knight 2005). Such parts lend themselves to modular assembly, which allows parts to be shuffled together to generate different genetic constructs that are to be assessed for functionality. Moreover, there can be situations where a single gene or a set of genes may have to be expressed in multiple host organisms, both prokaryotic and eukaryotic, to determine the most suitable host for production of a biomolecule of interest. Standard parts can then be easily swapped and replaced by genetic elements specific to different host organisms, simplifying the process of vector construction.

Parallel to the efforts of part standardization, which, as discussed above, aims at enabling modular construction and reuse of basic DNA parts in order to build systems of increasing complexity, is the development of assembly methods that can

support the combinatorial assembly of these standard parts. The motivation behind developing standard assembly methods is to avoid the tedious and repetitive processes involved in assembling constructs, and instead employ simpler standardized processes.

4.3 Standards Proposed for Genetic Parts and Associated Assembly Methods

While many different standards have been proposed, this section describes a few widely used ones and explains their utility. The assembly methods associated with each standard are also described.

The key feature of the standards discussed below is the idempotent nature of parts assembly, where each new part generated from basic parts retains its original structure and can be further used for higher-order assemblies with other basic parts.

4.3.1 NOMAD

'Nucleic acid ordered module assembly with directionality' (NOMAD) was one of the first standards to be proposed to permit the creation of families of DNA constructs from a pool of DNA fragments (Rebatchouk et al. 1996). NOMAD consists of two elements: DNA fragments with standardized ends, termed 'modules', and 'assembly vectors' that can be used for sequential and directional insertion of any number of modules. The modules thus inserted can also be released from the vectors, individually or in blocks.

The DNA fragment to be inserted in the vector is flanked by recognition sites for the restriction enzyme StyI. The module thus created generates sticky ends on digestion with StyI, and can be ligated into an assembly vector that has been digested with any one of the type IIS restriction endonucleases BsmBI, BsmAI or BsaI. The vector has been designed in such a way that digestion with any of these three endonucleases gives rise to StyI-compatible sticky ends (Fig. 4.1). To add another module (which has the same StyI sticky ends), the vector can be digested with either BsmBI (for insertion upstream of the first module) or BsaI (for insertion downstream of the first module). This process can be repeated to add more modules, and the final 'composite module' can be released either by simultaneous digestion with BsmBI and BsaI or by digestion with BsmAI alone.

4.3.2 BioBrick

The BioBrick 'RFC 10' standard is by far the most popular synthetic biology standard in use, perhaps owing to the requirement of employing this standard for creating parts for the annual International Genetically Engineered Machine (iGEM) competition. This standard was first proposed by Thomas Knight (2003).

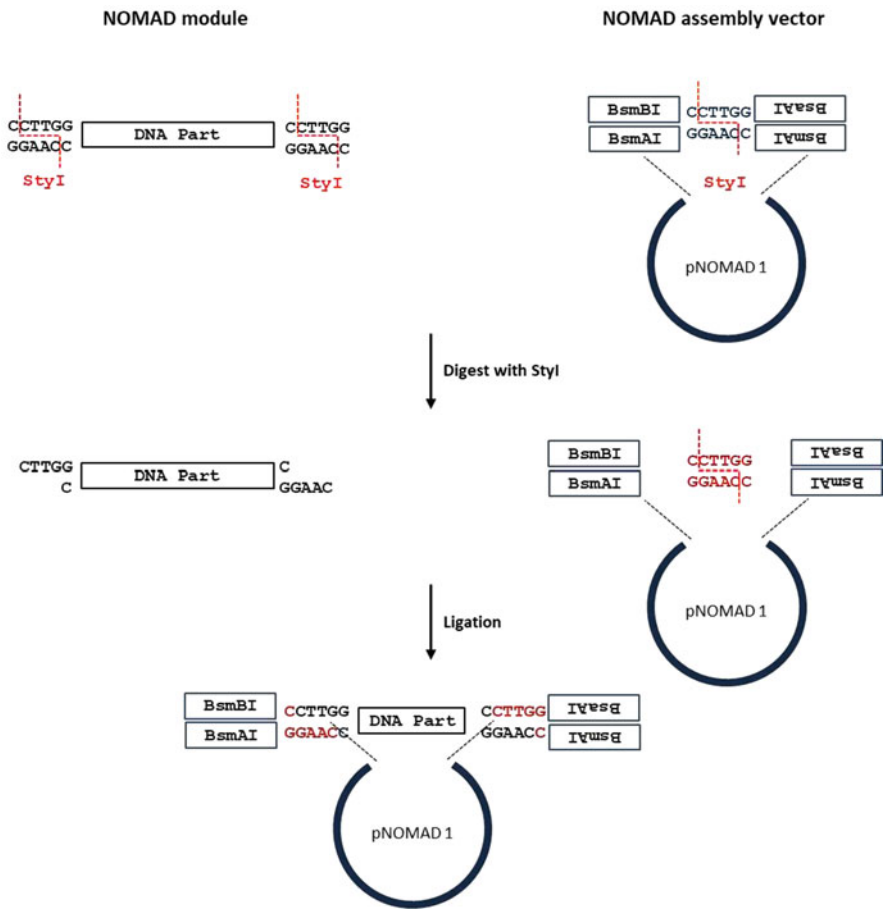


Fig. 4.1 NOMAD standard and assembly method

A BioBrick consists of a 'prefix' containing recognition sites for the restriction endonucleases EcoRI and XbaI, and a 'suffix' containing SpeI and PstI recognition sites (Fig. 4.2). NotI sites are also built into the prefix and suffix sequences, but they are not used in the BioBrick assembly process. As XbaI and SpeI generate complementary sticky ends, parts can be joined together unidirectionally into a vector that has been digested by EcoRI and PstI. The composite part can be released from the vector by digestion with EcoRI and SpeI (or XbaI and PstI), and joined upstream of another part that has been digested with XbaI and PstI (or downstream of a part digested by EcoRI and SpeI).

Ligation of XbaI and SpeI sites produces an 8 bp 'scar' that is not recognized by either of these enzymes. The presence of the scar sequence usually does not pose a problem, especially when non-coding DNA fragments are joined together. However, this strategy cannot be used to make fusion proteins, as the scar contains a TAG stop codon (see Fig. 4.2).

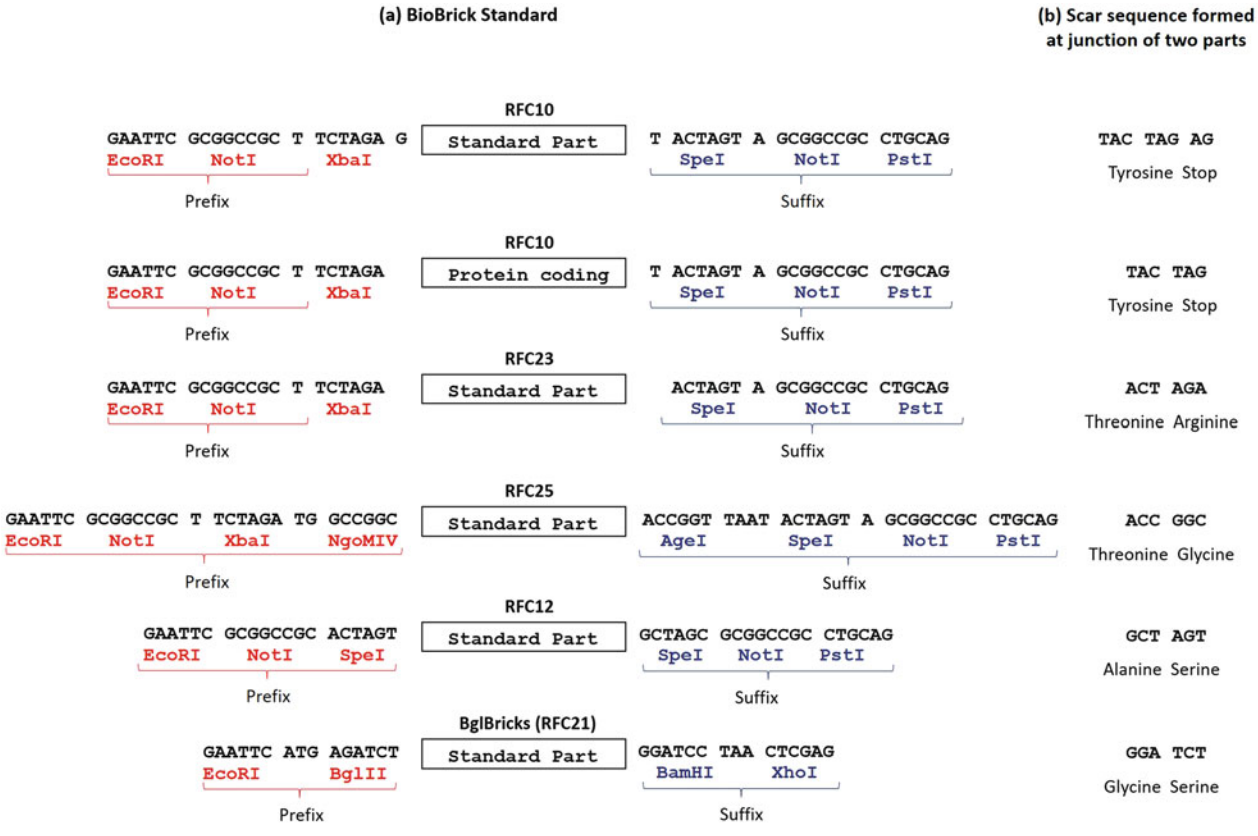


Fig. 4.2 BioBrick prefix and suffix sequences and scar sequences generated after part assembly

To circumvent this problem, a modified standard, RFC 25, has been proposed (Müller et al. 2009). In this standard, the prefix consists of, apart from sites for EcoRI, NotI and XbaI, a site for NgoMIV, while the prefix consists of a site for AgeI followed by sites for SpeI, NotI and PstI (Fig. 4.2). NgoMIV and AgeI generate compatible sticky ends, and the scar thus produced does not contain a stop codon, but rather codes for threonine and glycine. This standard can therefore be used to generate fusion proteins. Moreover, threonine and glycine are commonly used in synthetic linker peptides.

Another solution (RFC 23) proposed to create fusion proteins using the BioBrick standard (Phillips and Silver 2006) is to modify the prefix and suffix in such a manner that an in-frame 6 bp scar is produced, and that does not contain a stop codon (Fig. 4.2). This benign scar (coding for threonine and arginine) allows for in-frame assembly of fusion proteins. However, arginine in the N-terminal position can be destabilizing and result in faster protein degradation.

Another solution (RFC 12) proposed to create fusion proteins (Knight 2008) is to modify the prefix and suffix, such that a 6 bp scar coding for alanine and serine is produced (Fig. 4.2). Unlike RFC 23 and 25, this standard is incompatible with RFC 10.

As mentioned above, the RFC 10 BioBrick standard has been widely disseminated, and a repository of standard parts (http://parts.igem.org/) has been created that contains a growing number of BioBrick parts that are constructed and submitted by the iGEM community.

Standard Assembly was the first assembly technique used to join two BioBrick parts together to form a composite part (Knight 2003). In this technique, both parts (cloned in individual plasmids) are digested with appropriate enzymes, gel purified and ligated to form a composite part, with one part remaining in its plasmid (Fig. 4.3). This method has two main advantages: both part-containing plasmids can carry the same antibiotic resistance marker gene, and higher number of correct transformants are obtained as it requires only two correct ligations to form a circular plasmid. However, repetitive gel electrophoresis and extraction procedures limit the ability of this technique to be used in high-throughput assembly formats, as well as make it a cumbersome procedure.

3A assembly (Fig. 4.4) was developed to address the limitations of the Standard Assembly technique. Plasmids carrying the upstream and downstream parts, and the destination vector, are digested using the restriction enzymes whose recognition sequences are present in the parts' prefixes and suffixes. The three plasmids have different antibiotic resistance genes (from which the name '3A' or 'three antibiotics' is derived), enabling screening and selection of the correct ligation product without the need to purify DNA fragments (Shetty et al. 2011). Less efficient ligations and fewer transformants are disadvantages of this assembly method.

Amplified Insert Assembly was proposed by Speer and Richard (2011) to address the disadvantages of 3A assembly. The insert part is PCR amplified, hence the name of this assembly method. Common flanking primers are used to amplify insert parts from all BioBrick plasmids, thereby eliminating the requirement of new primer sets for amplification of new parts. After amplification, standard restriction sites are

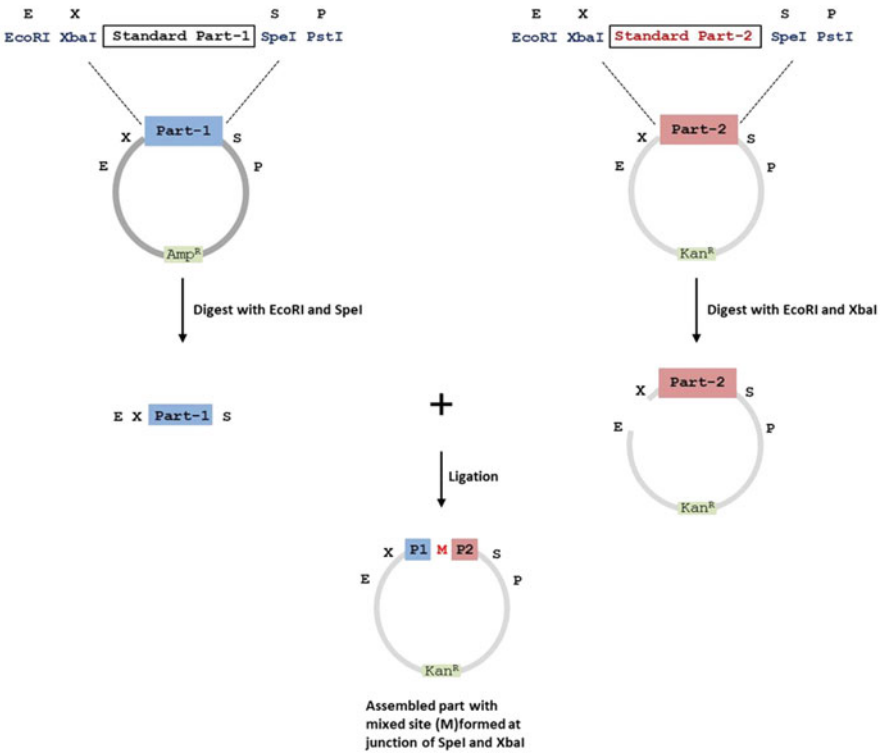


Fig. 4.3 Standard assembly methodology to join BioBrick parts

maintained in the amplified product. The PCR step adds approximately 300 bp to the amplified insert which makes small parts (such as RBSs, terminators) amenable to purification using column-based methods. Vector and insert backgrounds are avoided by the use of restriction enzymes (DpnI is used for insert, and phosphatase for phosphorylation of vector).

4.3.3 BglBrick

The BglBrick standard was described by Anderson et al. (2010) to allow the iterative assembly of standardized biological parts. Here, DNA parts are flanked by unique BglII and BamHI restriction sites at their 5′ and 3′ ends, respectively. Parts can be joined in specific order by digesting the first part with BamHI, and the second one with BglII. The sticky ends produced by both these enzymes are complementary to each other, allowing the two parts to be joined together to generate a composite part separated by a 6 bp scar that codes for glycine and serine, a commonly used linker in several host systems (Fig. 4.2). The composite part constructed by this method

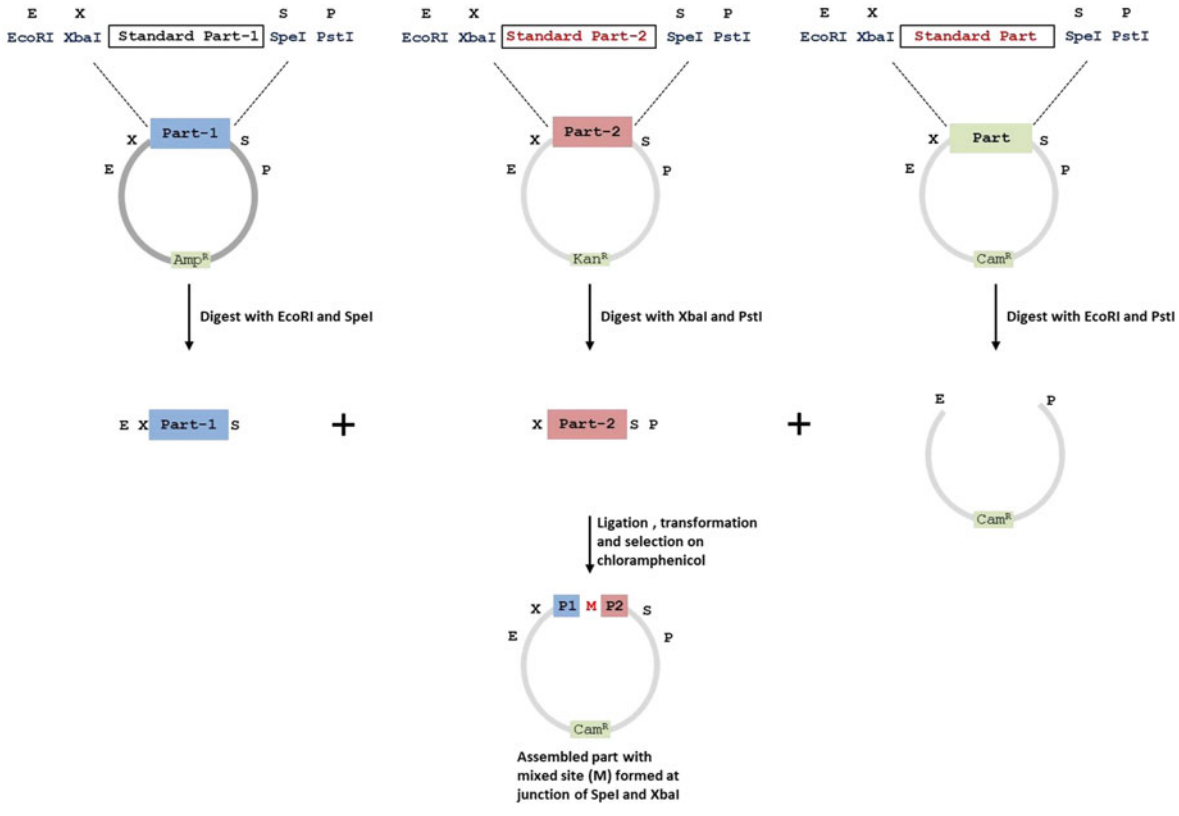


Fig. 4.4 3A assembly methodology to join BioBrick parts

is now flanked by BglII and BamHI sites, generating a new BglBrick that can be subjected to another round of assembly.

4.3.4 Golden Gate and Golden Braid

Unlike the assembly methods discussed above, the Golden Gate assembly methodology (Engler et al. 2008) allows multipartite assembly of parts in a single step.

Golden Gate assembly relies on the activity of type IIS restriction endonucleases and exploits their ability to create 5′ and 3′ overhangs by cutting outside their recognition site. Parts (cloned in a donor vector) are flanked on either side by BsaI recognition sites (Fig. 4.5). The destination vector has a *lacZ* gene flanked by BsaI sites, enabling blue-white screening of transformants. Upon addition of BsaI and ligase to a single reaction tube consisting of the donor and recipient vector, BsaI cuts outside of its recognition site (which are lost upon digestion), generating four fragments that can ligate to form only one stable recombinant plasmid that can be selected based on blue-white screening and antibiotic selection. Directional cloning is achieved based on the selection of cleavage sites. Using this method, it is possible to assemble multiple parts (entry clones), devoid of internal BsaI sites, into compatible destination vectors.

Golden Braid (Sarrion-Perdigones et al. 2011) is an adaptation of the Golden Gate methodology that can be used to assemble multiple standard parts in a single destination vector. Moreover, unlike Golden Gate, Golden Braid parts, once assembled, can be reused to generate higher-order devices.

In Golden Braid, parts are flanked by BsaI recognition sites and sticky end-generating sequences that are specific to the type of part. This feature permits directional assembly of promoters, coding sequences and transcription terminators into a vector designated level α destination plasmid. This destination vector has additional BsmBI sequences on either side of the insert, facilitating release of the insert and cloning into a level Ω destination vector that has an inverted orientation of BsmI and BsaI recognition sequences (Fig. 4.6). The relative positions of restriction sites in the vectors introduce a double-loop ('braid') topology in the cloning strategy, allowing composite parts to be created with each iterative assembly process.

4.3.5 Modular Cloning

Modular Cloning (MoClo) assembly (Weber et al. 2011), based on the Golden Gate assembly method, allows the assembly of complete transcription units from standardized modules. Each module is divided into several levels. Level 0 module consists of promoters, 5′ untranslated regions (UTRs), signal peptides, coding sequence and terminators. Each module is flanked by appropriate fusion sites that enable directional assembly. The fusion sites are specific to the function of the parts being assembled. For example, AATG is the fusion site at the start codon. Similarly, different fusion sites are used to construct fusion and cytosolic proteins.

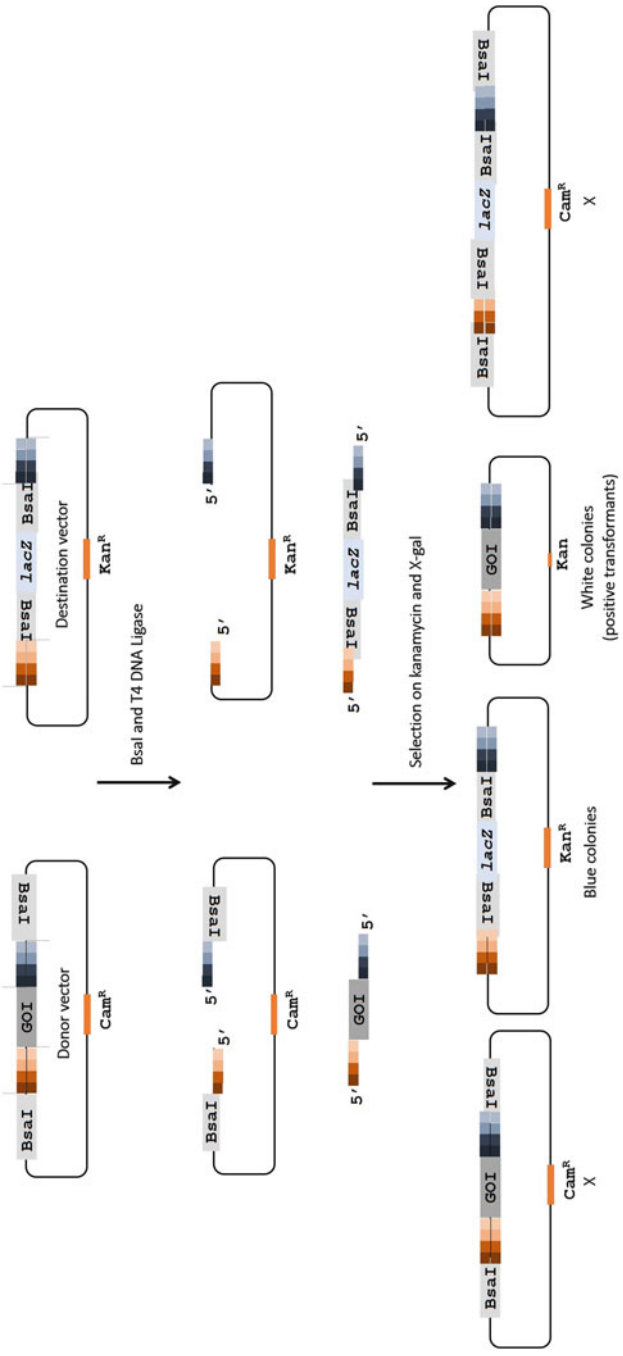


Fig. 4.5 Golden Gate assembly

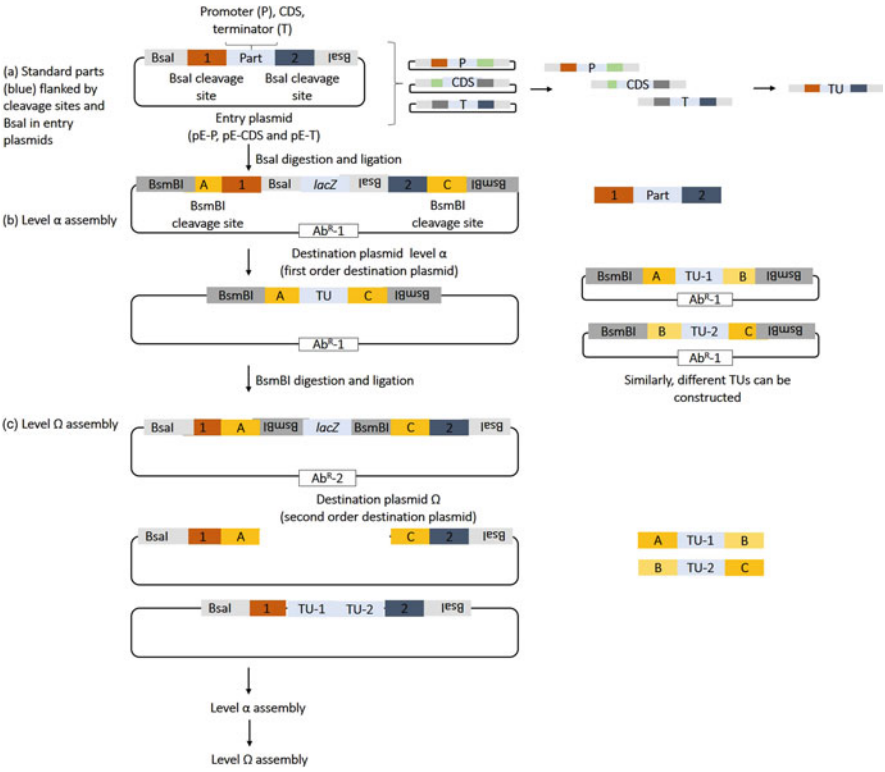


Fig. 4.6 Golden Braid assembly. (**a**) Standard parts (blue) flanked by cleavage sites and BsaI in entry plasmids. (**b**) Level α assembly. (**c**) Level Ω assembly

Level 0 modules are supported by level 0 destination vectors that contain an antibiotic marker gene and *lacZ* for screening (Fig. 4.7). The *lacZ* gene is flanked by two type IIS restriction enzymes, BsaI and BpiI, in inverse direction, such that upon digestion with these enzymes appropriate fusion sites compatible to level 0 module are created and upon assembly the BpiI and *lacZ* gene are cleaved off, retaining the part with BsaI recognition site. A level 1 module is formed by the assembly of different level 0 modules into a level 0 destination vector. The level 1 destination vector contains a different antibiotic selection marker. Using this strategy, an assembly of 11 transcription units, made up of 44 basic modules, in three Golden Gate reactions has been demonstrated. Key applications of this assembly method include the construction of multi-gene pathways and gene libraries.

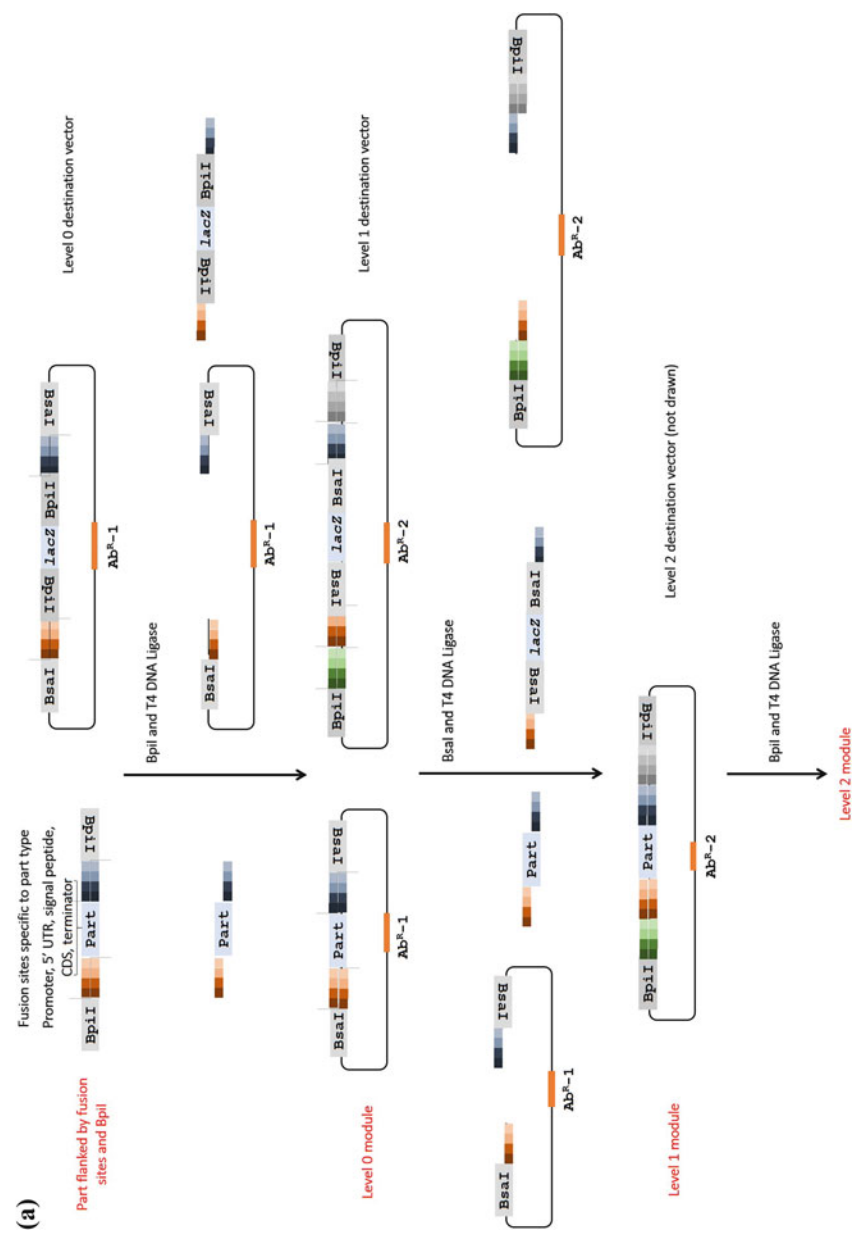


Fig. 4.7 (**a**) MoClo assembly; (**b**) individual parts are flanked by unique fusion sites

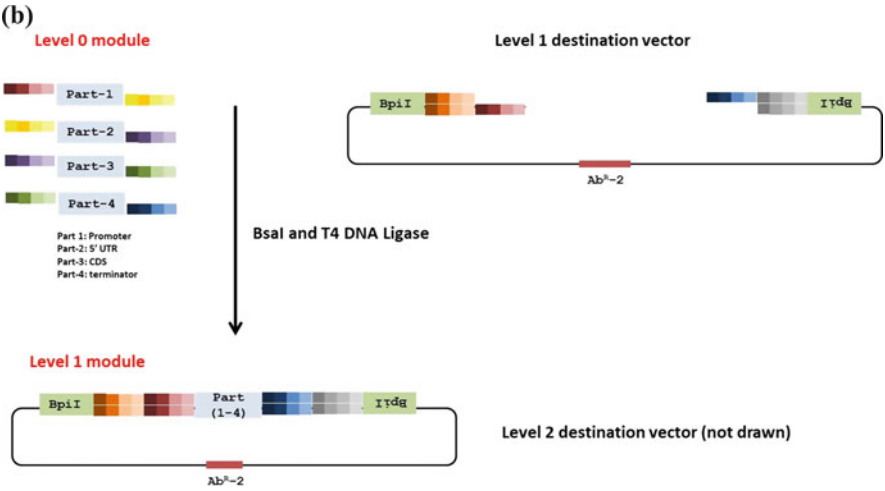


Fig. 4.7 (continued)

4.3.6 Scarless Golden Gate

Golden gate assembly utilizes type IIS endonucleases such as BsaI, which cut outside of their recognition site and form specific fusion sites that enable directional assembly of DNA parts.

Parts being assembled using the Golden Gate method should be free of internal BsaI sites, thereby either reducing flexibility or necessitating an additional step to remove internal recognition sites. Moreover, a 4 bp scar sequence is formed between assembled parts, which can be problematic when assembling transcriptional units. HamediRad et al. (2019) presented a scarless Golden Gate assembly method to address these limitations.

The authors evaluated linkers used in Golden Gate assembly based on their affinity towards themselves and other linkers. Linkers with high affinity towards themselves lead to highly efficient Golden Gate assembly reactions and are desirable, whereas greater affinity towards other linkers leads to misligation and reduced efficiency. By evaluating the data derived from next-generation sequencing, they were able to analyse the efficiency of various linkers and identified a set of 200 optimized linkers. To confirm the efficiency of the optimized linkers identified by their study, the authors assembled a six-gene zeaxanthin pathway of 10 kb. A web-based application, iBioCAD (http://ibiocad.igb.illinois.edu/), has been designed to enable efficient Golden Gate assembly reactions. As a proof-of-concept, the authors used this application and a set of optimized linkers to carry out the removal of an internal BsaI recognition site within the ampicillin resistance gene and carried out scarless assembly in a vector, in a single reaction.

4.3.7 iBrick

The restriction endonucleases employed by many standards discussed above recognize short DNA sequences, which are relatively widespread in natural DNA sequences. Such recognition sequences have to be removed from DNA sequences to construct standard parts. To circumvent this issue, Liu et al. (2014) proposed a new standard, iBrick, that employs two homing endonucleases, I-SceI and PI-PspI, instead of type II restriction endonucleases. The sequences recognized by I-SceI and PI-PspI are >18 bp and are extremely rare.

In the iBrick standard, parts are prefixed with the I-SceI recognition site and suffixed with that of PI-PspI. Digestion of these sites produces compatible non-palindromic cohesive ends, which are used for part assembly. The scar sequence is 21 bp long, and when translated in frame, produces a seven amino acids long linker peptide. The presence of three closely spaced leucine residues in the linker may impact protein solubility.

4.3.8 Start-Stop Assembly

The Start-Stop assembly method was proposed by Taylor et al. (2018) to address gaps in DNA assembly methods; modular assembly methods result in scar formation, while non-modular methods (such as those based on sequence homology) are sequence-dependent, resulting in the parts to be altered on a case-to-case basis. Scar sequences adjacent to coding sequences that are left behind as a result of assembly with type IIS enzymes (4 bp fusion sites) impact protein expression in many ways. To prevent this from occurring, the authors utilized 3 bp start and stop codons as overhangs which can be created using type IIS endonuclease such as SapI.

Similar to Golden Gate assembly and its variants, Start-Stop assembly defines discrete assembly levels. Level 0 constitutes parts such as promoter, RBS/UTR, CDS and terminators. Each of these parts is flanked by discrete fusion sites (described as alpha, beta, gamma, delta and epsilon). For example, a CDS is flanked by gamma and delta fusion sites, ATG and TAA, which are the most widely used start and stop codons, respectively. Similarly, appropriate fusion sites are assigned to promoter, UTR/RBS and terminator parts. Each part in this assembly is flanked by appropriate fusion sites and SapI and BsaI recognition sites, which can be generated either by PCR or by synthesis.

Level 1 consists of a transcription unit, which is made up of level 0 parts. In contrast to other assembly methods, a single vector (defined as level 0 storage vector) is used for the storage of different parts. Assembly from level 0 parts is done using a level 1 vector containing appropriate fusion sites. For level 2 and 3 assembly into level 2 and 3 vectors, BsaI and BbsI endonucleases are used, respectively. Similar to level 0 vectors, a single level 2 and 3 vector is used with defined fusion sites. The parts should be free from recognition sites for SapI and BsaI or BbsI. The authors evaluated the fidelity and bias of assembly method

by assembling a transcription unit, comprising of promoter(s), RBS(s), CDS (*yfp*) and terminator(s). To demonstrate the functionality and efficiency of the pathway, the authors assembled the astaxanthin pathway as five libraries in *E. coli*. Key applications include multi-fragment assembly for metabolic pathway engineering applications leading to identification of optimal part combinations.

4.4 Part Assembly Methods Not Associated with Standards

The assembly of standard parts described above relies on restriction-digestion and ligation reactions, which, as discussed above, necessarily create scar sequences between parts that are assembled together. Scar sequences are especially problematic between RBS and CDS parts, as the length of the intervening sequence influences expression. Recombination-based assembly of parts has therefore received attention from the Synthetic Biology community, as it allows the generation of constructs without intervening scar sequences (Gateway cloning is an exception in this regard; see below).

Recombination-based assembly requires homologous sequences to be present on ends to be joined together. Parts need not be engineered to remove restriction sites, as the only requirement for joining parts together is the presence of complementary overlap sequences. This greatly simplifies the process of part assembly (see Alnahhas et al. 2014 for an interesting discussion); however, recombination-based assembly methods require that each part to be assembled has to be amplified using unique primers, and parts cannot be excised from a construct or exchanged with another part, unless the parts have standard prefixes and suffixes that themselves act as overlap sequences (Sleight and Sauro 2013).

Several methods of DNA assembly based on sequence homology between overlapping fragments have been proposed as alternatives to the restriction-ligation based methods. Some of these are overlap extension PCR (OE-PCR), sequence and ligation-independent cloning (SLIC), circular polymerase extension cloning (CPEC) and Gibson assembly.

4.4.1 Overlap Extension PCR

Overlap extension PCR (OE-PCR) is one of the oldest and simplest methods of assembling DNA parts by carrying out two PCRs (Ho et al. 1989). In the first PCR, primers are used to generate inserts that contain regions of homology at the ends at which they are to be joined. A second PCR using these newly created inserts and appropriate flanking primers then leads to the generation of an assembled DNA fragment (Fig. 4.8). Applications of OE-PCR include the construction of fusion proteins and site-directed mutagenesis.

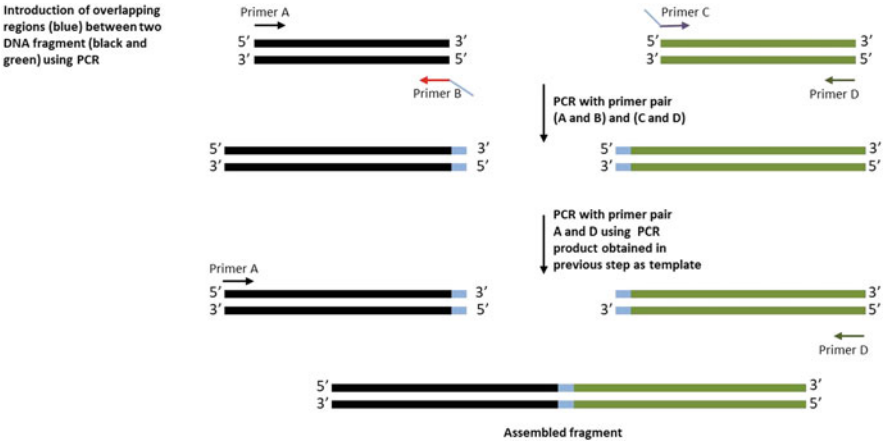


Fig. 4.8 Overlap extension PCR

4.4.2 Sequence and Ligation-Independent Cloning

Sequence and Ligation-Independent Cloning (SLIC), proposed by Li and Elledge (2007), is a method for DNA assembly that is based on in vitro homologous recombination that exploits the cell's inherent repair system. It is an advancement over the ligation-independent cloning (LIC) method for DNA assembly that was proposed by Aslanidis and de Jong (1990), as it avoids the stringent dependence on appropriate DNA sequence that LIC suffered from.

Homologous recombination in *E. coli* is mediated by either a recombinase RecA-dependent pathway or a RecA-independent pathway. Generation of homologous recombination intermediates when introduced into *E. coli* can promote DNA repair and lead to the joining of DNA fragments. Homologous recombination intermediates are generated by digesting the vector and insert (containing overlapping regions of homology) with exonucleases such as T4 DNA polymerase to produce single-stranded DNA (ssDNA) with 5′ overhangs. The insert is obtained using PCR and the vector is treated with the required restriction enzymes, following which both insert and vector are treated with exonuclease to generate ssDNA with 5′ overhangs (Fig. 4.9). The resulting fragments are assembled in vitro and transformed into *E. coli* to generate recombinant DNA of interest. A 30 bp overlap is sufficient to promote annealing and repair. Multiple fragment assembly can be carried out with overhangs between 20 and 40 bp. Key advantages of this methodology are that it is scarless, does not suffer from limitations of traditional restriction-ligation based methods and also does not impose the requirement of specific sequences as in the case of methods based on site-specific recombination.

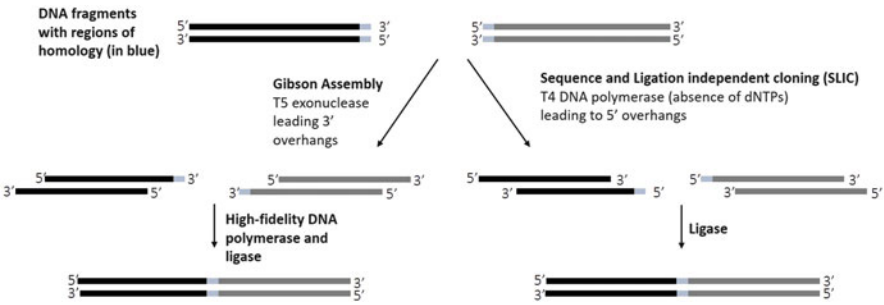


Fig. 4.9 SLIC and Gibson assembly

4.4.3 Circular Polymerase Extension Cloning

Similar to other assembly methods based on sequence homology between joining fragments, Circular Polymerase Extension Cloning (CPEC) also allows for scarless DNA assembly (Quan and Tian 2009). However, unlike other methods that entail the creation of single-stranded overhangs, CPEC exploits polymerase chain extension to assemble fragments into complete circular plasmids.

In this method, insert fragments are generated by PCR that share regions of homology to the vector (or inserts, in the case of multi-insert assembly) that is built in using appropriate primers. Upon denaturation, the insert and vector having overlapping sequences anneal and extend to form complete plasmids (Fig. 4.10). Using this method, the authors were able to demonstrate the assembly of single and multiple DNA fragments by optimizing the number of CPEC cycles. A multi-gene pathway for polyhydroxybutyrate (PHB) production consisting of more than four fragments was assembled at 100% efficiency in *E. coli* to form a functional pathway by performing 20 cycles of CPEC. The CPEC offers a simple, convenient, efficient and economical method of assembling DNA parts. High melting temperatures in overlapping sequences between DNA fragments ensure accurate annealing and lower chances of non-specific annealing.

4.4.4 Gibson Assembly

Gibson et al. (2009) described an assembly method that allows seamless, parallel and directional assembly of multiple DNA fragments in a single isothermal in vitro reaction. Part assembly is based on the presence of overlapping regions of homology that are incorporated using PCR. This method relies on the use of T5 exonuclease to chew back the 5′ end of the DNA fragments leading to the generation of 3′ complementary overhangs that are annealed and joined at 50 °C using a high fidelity Phusion DNA polymerase and Taq DNA ligase (Fig. 4.9).

Gibson assembly provides a simple and efficient methodology to assemble multiple fragments in a single reaction, with fragments as large as 300 kb. Using

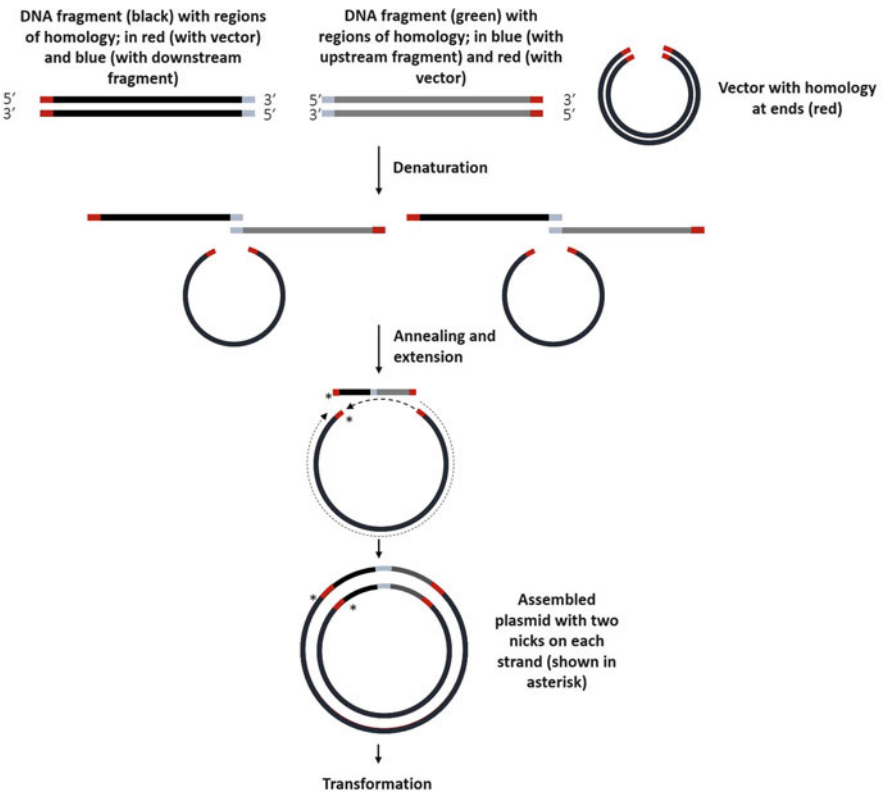


Fig. 4.10 Circular Polymerase Extension Cloning (CPEC)

this method, the authors were able to assemble multiple fragments of up to 318 kb, each containing a 40 bp overlap. The main advantage of this assembly method is that it allows highly efficient and ordered assembly of multiple DNA fragments in a single in vitro reaction.

4.4.5 Modular Overlap-Directed Assembly with Linkers

Modular Overlap-Directed Assembly with Linkers (MODAL) addresses the limitation of the lack of modularity (Casini et al. 2014), which is common to assembly methods that are dependent on overlapping sequences. Widely adopted assembly methods, including Gibson assembly, SLIC and OE-PCR among others, depend on the presence of overlapping sequences between fragments, which guide assembly reactions. Therefore, each assembly reaction is dependent on the regions of homology between the fragments being assembled, and every different assembly

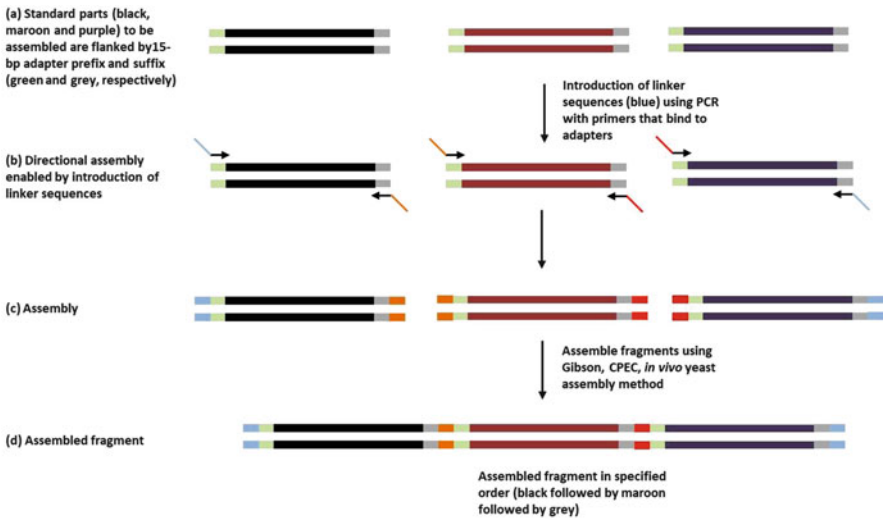


Fig. 4.11 Modular Overlap-Directed Assembly with Linkers (MODAL)

would require changing these sequences as per the assembly requirements. This limits the basic tenet of synthetic biology which aims towards standardization.

The authors devised a strategy to standardize homology-based assembly methods such as Gibson, CPEC and in vivo yeast recombination. This could be achieved by introducing a 45 bp overlap region (called linker) flanking a basic DNA part. Prefix and suffix adapter sequences are first added to the DNA part using either primers or by part synthesis, following which, a 45 bp linker is added to this part using universal primers (Fig. 4.11). These generated parts can then be assembled using any of the homology-based assembly methods such as Gibson assembly, CPEC or yeast recombination. Using this strategy, the authors were able to assemble up to five parts with efficiency ranging from 75 to 100%. Assembly using these linkers led to a 75-bp scar sequence at the junction of two parts, which could be problematic when assembling transcription units comprising of promoters, RBS and open-reading frames (ORFs). It was also shown that the presence of synthetic linker DNA did not lead to context-dependent effects. A freely available online tool, R2oDNA Designer (http://www.r2odna.com/), was developed to aid in the construction of orthogonal linker sequences having pre-determined and optimal length, guanine-cytosine (GC) content and melting temperature. Applications of this strategy include assembly as well as generation of libraries with mutation introduced in specific parts.

4.4.6 Gateway Cloning

Gateway cloning or recombinational cloning (Hartley et al. 2000) is a widely used assembly method based on site-specific recombination of λ phage. Integration

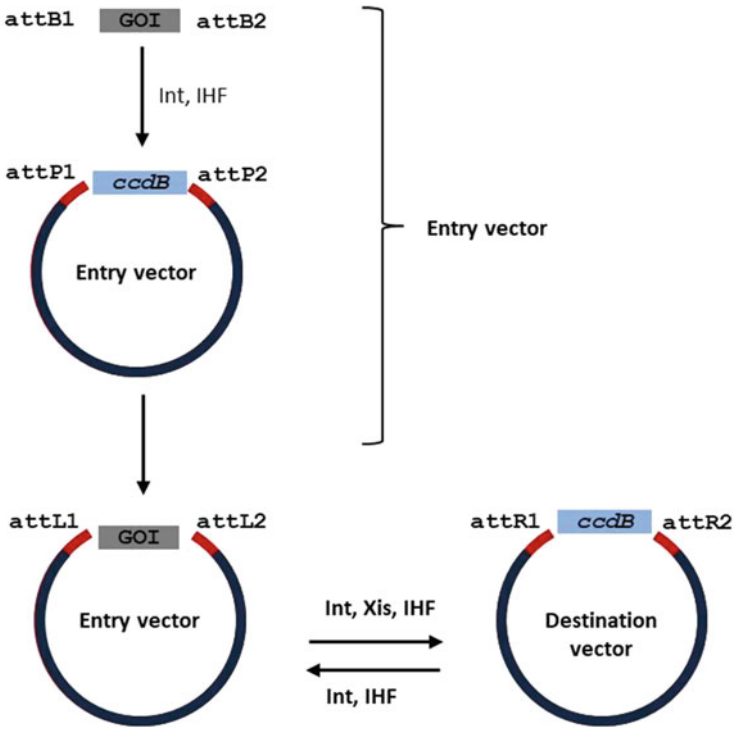


Fig. 4.12 Gateway cloning

events by λ phage in *E. coli* are carried out due to the presence of attachment sites *attB* and *attP* present in bacteria and phage, respectively. Integration is mediated by the phage and bacterial encoded integration proteins integrase (Int) and integration host factor (IHF). This process is reversible and requires the presence of another enzyme, excisionase (Xis). Gateway cloning is based on this system where the integration and excision events take place in vitro.

In this system, two vectors are used, the entry clone and the destination vector. First, the DNA fragment flanked by recombination sites *attB1* and *attB2* is integrated into the entry vector, which has the counterselectable marker *ccdB* flanked by *attP1* and *attP2* recombination sites (Fig. 4.12). This in vitro recombination is mediated by integrase and integration host factor (Int and IHF). This fragment integrated in the entry clone (now flanked by the newly created *attL1* and *attL2* sites) is subcloned into a destination vector which contains *ccdB* flanked by *attR1*and *attR2* sites, by incubation of both vectors in the presence of integrase, excisionase and integration host factor (Int, Xis and IHF). The screening of correct clones is carried out by the selection for antibiotic resistance and loss of *ccdB*.

Using this methodology, it is possible to carry out simultaneous and ordered cloning of gene of interest into multiple destination vectors. One of the key features

of this system is that it is highly efficient and specific with no crosstalk. The recombination sites can be mutated to generate variants that retain functionality but lead to specific recombination sites that recombine with high specificity and accuracy. The Gateway cloning method has been widely adopted by the scientific community and has been used in several applications such as protein expression and optimization.

4.5 Part Characterization

DNA parts, once constructed, are expected to be used in a modular fashion. This is possible only when parts are characterized according to standard metrics and are demonstrated to behave in a consistent and predictable manner in different contexts. To make part activity independent of context and consistent between circuits, individual parts are insulated by means of terminator sequences or innocuous random sequences with no predicted biological activity (Carr et al. 2017; Davis et al. 2011; Torella et al. 2014).

Part characterization is aided by the use of automated high-throughput robotic systems (Linshiz et al. 2016), which reduces errors due to variations that creep in due to manual handling (Decoene et al. 2018; also see Beal et al. 2016). However, the use of automation does not eliminate all variations, as experimental settings differ between manufacturers, which are at times not accounted for or reported (Chavez et al. 2017).

Characterization parameters developed for DNA parts and devices are discussed in the following sections.

4.5.1 Characterization of Genetic Parts

Promoter strength characterization is usually carried out by mutating a promoter sequence and testing the ability of the various mutant versions to drive expression of a reporter gene. This approach depends on screening mutant promoter libraries, rather than in silico prediction of promoter properties (Boyle and Silver 2012).

Perhaps, the simplest approach to measuring promoter activity is exemplified by a study by Keren et al. (2013) to measure promoter activity in yeast. Promoters under study were integrated into the genome upstream of a yellow fluorescent reporter (YFP), and promoter activity was defined as the YFP production rate per optical density (OD) unit per second in the window of maximal growth. However, such measurements cannot capture global variation due to the metabolic load of an introduced synthetic gene circuit, differences in growth conditions or the initial state of inoculated cells (Rudge et al. 2016).

Promoter activity varies across environmental conditions and is also dependent on the reporter gene being expressed from the promoter under study. To address these concerns, Kelly et al. (2009) proposed a Relative Promoter Unit (RPU) that can be used to report promoter activity in compatible units. The constitutive promoter

BBa_J23101 is used as an in vivo reference standard, and the promoter activity is to be reported with reference to this standard. The RPU is defined as the ratio of the rate of synthesis of a reporter protein from the test promoter per cell in the exponential phase to that of the reference standard promoter.

$$\mathrm{RPU} = \frac{\left(\frac{dF\Phi}{dt}\right)/\mathrm{ABS}\Phi}{\left(\frac{dFJ23101}{dt}\right)/\mathrm{ABSJ23101}}$$

For a valid comparison to be made between the promoter to be tested and the reference promoter, both promoters should have identical transcription initiation sites and identical sequences downstream of the initiation site. This is to ensure that the messenger RNA (mRNA) expressed by the test promoter is identical to the mRNA produced by the reference promoter, and that the mRNA degradation rate and translation rate of the reporter mRNA are similar (Kelly et al. 2009).

Though the RPU proposal was originally described for promoters cloned in plasmids, this approach has also been used for promoters integrated in the genome, as the conditions of the RPU approach were assumed to be valid (Zucca et al. 2013).

In another study using the RPU approach, the medium-strength J23118 promoter was used as the reference promoter (Pasotti et al. 2012).

Radeck et al. (2013) used a mathematical model for the Lux system to estimate the time dependence of promoter activity that results in a given luciferase activity. For simplicity, the authors assumed that one of the Lux proteins is rate-limiting for light production, and that luminescence is proportional to this protein species. Promoter activity was calculated as a time-dependent transcription rate, α, of the Lux operon. *luxABCDE* mRNA degradation and translation rates, as well as degradation rate of the Lux proteins, were empirically determined, and the model solved for α.

Rudge et al. (2016) proposed to quantitatively measure transcription from a given promoter that is consistent under a range of conditions by developing a method for in vivo characterization of promoters based on the dual-channel measurement of reporter proteins. In this approach, two fluorescent reporter genes are cloned under two different promoters (test and the BBa_J23101 standard reference) in a single plasmid to eliminate 'extrinsic' variables that affect promoter activity, so that only 'intrinsic' variables remain to be measured.

The authors demonstrated that promoter function is a product of intrinsic factors (the affinity of a promoter for available resources and its maximal attainable activity) that are not dependent on growth conditions, and extrinsic factors (mRNA degradation, translation, fluorescent protein maturation and cell growth). The authors developed a promoter characterization method based on a mathematical model for cell growth and reporter gene expression, and which exploits multiple in vivo measurements to compensate for variation due to extrinsic factors. They derived a quantitative promoter characteristic (ρ) that provides a robust measure of the intrinsic properties of a promoter, relative to the control.

Espah Borujeni et al. (2014) developed a biophysical model using thermodynamic first principles and a four-parameter free energy model to predict translation initiation rates for 136 synthetic 5′ UTRs. Importantly, the authors demonstrated that the ribosome's interactions with sites upstream of Shine-Dalgarno sequences can modulate translation initiation rates by 100-fold. The model is used in the widely used RBS Calculator software (available at https://salislab.net/software/forward).

The prediction of the efficiency of 148 *E. coli* Rho-independent transcription terminators was carried out by d'Aubenton Carafa et al. (1990). Termination efficiency was predicted as a function of the total length of the terminator, the free energy of the stem loop structure and the number of thymine residues in the stretch of 15 nt after the stem loop.

Chen et al. (2013) studied the efficiency of 227 annotated *E. coli* terminators, as well as 265 synthetic terminators. The authors developed a biophysical model to capture how the terminator sequence affects its strength based on free energy considerations (including that of binding between the U-tract and template DNA, as well as those of hairpin folding and closure).

4.5.2 Characterization of Synthetic Circuits

The minimal information necessary for characterization of biological devices are (1) the maximum and minimum levels of gene expression at steady state, (2) the change in expression of the 'output' gene when 'input' is varied and (3) the signal-to-noise ratio (Beal et al. 2018). Experimentally, this translates into large numbers of single-cell measurements, measurement of output signal across the full range of input signals, copy-number effect of the device and statistical distribution of single-cell output/input levels to determine variations (Beal et al. 2012).

Quantitative description of synthetic devices to better support composition and abstraction is a requirement for higher-order genetic circuit design. Canton et al. (2008) provide an example of a description of a genetically encoded receiver that uses 3-oxohexanoyl-L-homoserine lactone as the input, the output being the flow of RNA polymerase (RNAP) along DNA (polymerases per second, or PoPS). Static and dynamic performance, input compatibility and reliability of the device were determined, and a 'datasheet' created similar to those used in engineering disciplines.

The prediction of genetic circuit behaviour is required to guide the design of complex circuits. Davidsohn et al. (2015) developed a method for the prediction of genetic regulatory network behaviour from characterizations of its constituent parts. This method, dubbed Empirical Quantitative Incremental Prediction (EQuIP), models gene expression as a function of regulatory inputs, circuit copy number and time, based on high-precision experimental observations. The authors used their method to model and predict the behaviour of synthetic genetic circuits in mammalian cells.

The physiological state of the host cell, which includes available RNA polymerases and ribosomes, as well as cell volume, impacts gene expression (Klumpp et

al. 2009). Braniff et al. (2019) used a model to predict host physiological properties from an observed exponential growth rate, and to distinguish intrinsic parameters of genetic constructs from extrinsic parameters of the host's physiological state. This was implemented by using optimal experimental design algorithms to characterize a genetic component from simulated data.

Huynh and Tagkopoulos (2016) applied a modelling framework to describe part behaviour of 118 circuits and 165 genetic parts of *E. coli* obtained from 135 publications. The architecture and dynamics of each published circuit were obtained from published values, and a model was built for each circuit using a set of functions to describe regulatory and other biological behaviour. A database, PAMDB (http://pamdb.com/), was created, which contains information related to model parameters and simulations, apart from curated data and metadata of the published gene circuits.

Mutalik et al. (2013) constructed a combinatorial library of frequently used transcription and translation genetic elements in *E. coli* and measured their individual amounts of mRNA and protein for all element combinations. The authors assumed that output of the genetic devices is a function of promoter, 5′ UTR and gene of interest elements, as well as their interactions with each other. Full factorial analysis of variance (ANOVA) was performed to predict outputs according to the identity of each element and element–element interactions. The mathematical framework developed could used to score intrinsic activities of genetic elements across different contexts, with variation in part activity serving as a quantitative score of part quality and reliability.

Gorochowski et al. (2017) developed RNA-seq methods to measure the performance of promoters, insulators and terminators in a genetic circuit. The authors defined the activity of a promoter as the increase in the flux of RNAPs that occurs between the beginning and end nucleotides of a promoter. The activity of a terminator was calculated as a change in flux of RNAPs, which either dissociate from the DNA or read-through the terminator. The activity of a sensor was calculated as the activity of the output promoter in the presence and absence of the inducing signal.

Polycistronic gene circuits are commonly employed in synthetic biology, but it is not known how do adjacent coding sequences (and non-coding RBSs) affect each other's expression. Wu et al. (2018) constructed synthetic operons with a reporter gene flanked by different 'adjacent transcriptional regions' with varying GC content and sizes, and studied gene expression to build a model that accounts for these parameters. The authors used the metric to design and construct synthetic gene circuits that could tune protein expression levels over a 300-fold range.

4.6 Conclusion

The use of standardized and characterized biological parts is essential for rational, design-oriented synthetic biology (Exley et al. 2019). Development of rigorous characterization and manufacturing protocols for biological parts, along with

efficient assembly methodologies, are prerequisites for realizing the promises of synthetic biology (Arkin 2008).

Standards proposed for DNA parts permit assembly (and disassembly) of these parts in a modular fashion. These standards can therefore be considered to be 'assembly standards'. These standards have been adopted by the Synthetic Biology community, and the number of parts submitted to the Registry of Standard Biological Parts every year by teams participating in the iGEM competition attests to the utility of these standards.

With the advent and widespread use of recombination-based assembly methods, standards that allow parts to function in a predictable fashion and to be functionally compared with other similar parts need to be defined. This need is well recognized; for example, for promoters to be compared to each other, the transcription initiation site and the sequence downstream of the initiation site must be identical (Kelly et al. 2009). Definition and adoption of such 'functional standards' is the way forward for a truly engineering approach to biology.

References

Alnahhas RN, Slater B, Huang Y et al (2014) The case for decoupling assembly and submission standards to maintain a more flexible registry of biological parts. J Biol Eng 8:28

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

Arkin A (2008) Setting the standard in synthetic biology. Nat Biotechnol 26:771–774

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

Beal J, Weiss R, Yaman F, et al (2012) A method for fast, high-precision characterization of synthetic biology devices. https://dspace.mit.edu/handle/1721.1/69973. Accessed 26 May 2019

Beal J, Haddock-Angelli T, Gershater M et al (2016) Reproducibility of fluorescent expression from engineered biological constructs in *E. coli*. PLoS One 11:e0150182

Beal J, Haddock-Angelli T, Farny N et al (2018) Time to get serious about measurement in synthetic biology. Trends Biotechnol 36:869–871

Boyle PM, Silver PA (2012) Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab Eng 14:223–232

Braniff N, Scott M, Ingalls B (2019) Component characterization in a growth-dependent physiological context: optimal experimental design. PRO 7:52

Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol 26:787–793

Carr SB, Beal J, Densmore DM (2017) Reducing DNA context dependence in bacterial promoters. PLoS One 12:e0176013

Casini A, MacDonald JT, De Jonghe J et al (2014) One-pot DNA construction for synthetic biology: the modular overlap-directed assembly with linkers (MODAL) strategy. Nucleic Acids Res 42:e7

Chavez M, Ho J, Tan C (2017) Reproducibility of high-throughput plate-reader experiments in synthetic biology. ACS Synth Biol 6:375–380

Chen YJ, Liu P, Nielsen AA et al (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 10:659–664

d'Aubenton Carafa Y, Brody E, Thermes C (1990) Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. J Mol Biol 216:835–858

Davidsohn N, Beal J, Kiani S et al (2015) Accurate predictions of genetic circuit behavior from part characterization and modular composition. ACS Synth Biol 4:673–681
Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 39:1131–1141
Decoene T, De Paepe B, Maertens J et al (2018) Standardization in synthetic biology: an engineering discipline coming of age. Crit Rev Biotechnol 38:647–656
Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647
Espah Borujeni A, Channarasappa AS, Salis HM (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res 42:2646–2659
Exley K, Reynolds CR, Suckling L et al (2019) Utilising datasheets for the informed automated design and build of a synthetic metabolic pathway. J Biol Eng 13:8
Gibson DG, Young L, Chuang RY et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
Gorochowski TE, Espah Borujeni A, Park Y et al (2017) Genetic circuit characterization and debugging using RNA-seq. Mol Syst Biol 13:952
HamediRad M, Weisberg S, Chao R et al (2019) Highly efficient single-pot scarless Golden Gate assembly. ACS Synth Biol 8:1047–1054
Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome Res 10:1788–1795
Ho SN, Hunt HD, Horton RM et al (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
Huynh L, Tagkopoulos I (2016) A parts database with consensus parameter estimation for synthetic circuit design. ACS Synth Biol 5:1412–1420
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
Keren L, Zackay O, Lotan-Pompan M et al (2013) Promoters maintain their relative activity levels under different growth conditions. Mol Syst Biol 9:701
Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. Cell 139:1366–1375
Knight T (2003) Idempotent vector design for standard assembly of Biobricks. https://dspace.mit.edu/handle/1721.1/21168. Accessed May 26 2019
Knight TF (2005) Engineering novel life. Mol Syst Biol 1:2005.0020
Knight T (2008) Draft standard for Biobrick BB-2 biological parts. https://dspace.mit.edu/handle/1721.1/45139. Accessed 26 May 2019
Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
Linshiz G, Jensen E, Stawski N et al (2016) End-to-end automated microfluidic platform for synthetic biology: from design to functional analysis. J Biol Eng 10:3
Liu JK, Chen WH, Ren SX et al (2014) iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. PLoS One 9:e110852
Müller KM, Arndt KM, Grünberg R (2009) Fusion protein (Freiburg) Biobrick assembly standard. https://dspace.mit.edu/handle/1721.1/45140. Accessed 26 May 2019
Mutalik VK, Guimaraes JC, Cambray G et al (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods 10:347–353
Pasotti L, Politi N, Zucca S et al (2012) Bottom-up engineering of biological systems through standard bricks: a modularity study on basic parts and devices. PLoS One 7:e39407
Phillips I, Silver P (2006) A new Biobrick assembly strategy designed for facile protein engineering. https://dspace.mit.edu/handle/1721.1/32535. Accessed 26 May 2019
Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441
Radeck J, Kraft K, Bartels J et al (2013) The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. J Biol Eng 7:29

Rebatchouk D, Daraselia N, Narita JO (1996) NOMAD: a versatile strategy for in vitro DNA manipulation applied to promoter analysis and vector design. Proc Natl Acad Sci U S A 93:10891–10896

Rudge TJ, Brown JR, Federici F et al (2016) Characterization of intrinsic properties of promoters. ACS Synth Biol 5:89–98

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

Shetty R, Lizarazo M, Rettberg R et al (2011) Assembly of BioBrick standard biological parts using three antibiotic assembly. Methods Enzymol 498:311–326

Sleight SC, Sauro HM (2013) Randomized BioBrick assembly: a novel DNA assembly method for randomizing and optimizing genetic circuits and metabolic pathways. ACS Synth Biol 2:506–518

Speer MA, Richard TL (2011) Amplified insert assembly: an optimized approach to standard assembly of BioBrickTM genetic circuits. J Biol Eng 5:17

Taylor GM, Mordaka PM, Heap JT (2018) Start-stop assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. Nucleic Acids Res 47:e17

Torella JP, Boehm CR, Lienert F et al (2014) Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. Nucleic Acids Res 42:681–689

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

Wu F, Zhang Q, Wang X (2018) Design of adjacent transcriptional regions to tune gene expression and facilitate circuit construction. Cell Syst 6:206–215.e6

Zucca S, Pasotti L, Politi N et al (2013) A standard vector for the chromosomal integration and characterization of BioBrick parts in *Escherichia coli*. J Biol Eng 7:12

Current Status and Challenges of DNA Sequencing

5

Indra Mani

Abstract

This chapter analyzes the DNA sequencing as its core technology impact to establish a relatively new scientific discipline as genomics to metagenomics and transcriptomics to metatranscriptomics. Due to the development of technology from conventional Sanger method to next-generation sequencing (NGS) techniques, it has transformed the genomics. NGS has improved the expense, scale of characterization of genome, and has significantly expanded the possibility of an investigation. Enormously comparable sequencing has demonstrated innovative bioinformatics tools, changing the assumption of genomics to address life issues at a genome- level. Emerging and established evidence of various genomic databases have demonstrated that the NGS has played a great role in the establishment of the whole-genome sequence (WGS) and whole exome sequence (WES). Currently, sequencing allows disease investigation like prognosis (mutational status), diagnosis (DNA and RNA-based), treatment (therapeutic identification), and molecular phylogeny. Preferably, sequencing should be precise, rapid, and cover long read. The elucidation of genomes from NGS platforms needs to be programmed and completely interconnected. In addition, the curation of data is very much necessary and helpful in an accurate annotation of the genome. Though retaining uniformity and precision in genome annotation is a challenging issue. Without annotation of the genome, sequencing data is meaningless. Consequently, appropriate reasonable investigation requires a constant and broad opinion of biological data.

Keywords

Next-generation sequencing (NGS) · Sanger sequencing · DNA · Genome · PCR

I. Mani (✉)
Department of Microbiology, Gargi College, University of Delhi, New Delhi, India

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_5

5.1 Introduction

The invention of DNA sequencing has been about 42 years, and the technologies are continuing to advance in the upcoming eras. Due to the continuous development of DNA sequencing technology and understanding of the genome, it has established comparatively a novel scientific discipline as genomics. For studies of genomics, purification of DNA is an essential step to further manipulation, including sequencing, restriction digestion, ligation, transformation, mutagenesis, and construction of probes (for hybridization). The utilization of different cloning vectors (M13mp or pUC) in *Escherichia coli* (*E. coli*) to increase the copy number (amplification) of DNA is very useful (Saiki et al. 1988). In addition, bacteriophage such as M13 is also very potential for these purposes (Messing 2001). A technique for the semi-automation of DNA sequence evaluation has been established, during an initial improvement of DNA sequencing. For each of the reactions particular for the bases A, C, G, and T, a different colored fluorophore has been used (Smith et al. 1986).

The whole-genome random sequencing and assembly have been utilized to determine the entire genome sequence (580,070 base pairs) of the *Mycoplasma genitalium*. It was the minimum known genome of any free-living organism. After comparative evaluation of this genome with *Haemophilus influenzae*, it has been suggested that the variances in genome content reflected as great differences in metabolic capacity and physiology of these two microbes (Fraser et al. 1995; Fleischmann et al. 1995). Enormously comparable sequencing has been established as novel and changing the hypothesis of genomics helps to understand the biology of organisms at a genome level. Due to the rapid development of genomics, it has revolutionized the biomedical research and clinical medicine (Shendure et al. 2017). Remarkably, DNA sequencing has been utilized in like prognosis (mutational status), diagnosis (DNA and RNA-based), treatment (therapeutic identification), and molecular phylogeny (Marian 2011; Koboldt et al. 2013; Smith 2017; Hu et al. 2019; Won et al. 2019).

Recently, for bacteria, algae, fungi, and protozoa, a DNA barcode has been developed. Remarkably, it was associated with error-free and rapid species identification that supported in understanding the microbial species which involve in a particular disease and microbial diversity (Chakraborty et al. 2014). Similarly, next-generation sequencing has facilitated the clinical metagenomic (culture-independent approach) and plays a vital role in research laboratories to clinical applications. This developing technology helps to vary diagnosis and treatments of communicable and noncommunicable (like cancer, diabetes, heart attack, etc.) diseases (Chiu and Miller 2019). In the oncology discipline, methylated-DNA sequencing tools are yielding great amounts of methylome records from cancer samples, from which cancer-associated differentially methylated CpG sites (cDMCs) have constantly recognized and filed. The addition of as numerous cDMCs as likely helps advance the precision of cancer examination and occasionally identify cancer subtypes. Nevertheless, the absence of a well-known technique for the analysis of 100 s of cDMCs normally hinders their vigorous practice in treatment (Jeon et al.

2017). Moreover, high-throughput sequencing (HTS) is progressively essential in defining cancer diagnoses, with subsequent prognostic and therapeutic implications (Guillermin et al. 2018). Presently, NGS plays a significant role to understand the gene expression in normal and cancer cells (Craig et al. 2016). It was possible because of the development of NGS technologies.

Recently, Oxford Nanopore sequence technology, MinION, has been developed, and it has improved the other technical requirements and is portable for onsite sequencing. This technology has been applicable in various fields like clinical medicine, environmental sphere, and biosecurity (Runtuwene et al. 2019). In contrast, for quantitative gene expression analyses, the RNA-Seq technology has been used as a widespread methodology. Further, it has also helped in the annotation of the transcriptome in gene expression study under various conditions (Blow 2009; Roberts et al. 2011; Garalde et al. 2016). Nevertheless, precise gene expression estimation needs exact genome information. Due to the development of technology from conventional Sanger method (Fig. 5.1a) to next-generation sequencing (NGS) techniques, it has significantly helped to understand genomics, transcriptomics, metagenomics, and metatranscriptomics Fig. 5.1.

5.2 Next-Generation Sequencing

Modern DNA sequencing technology is commonly known with different names, such as next-generation sequencing (NGS), deep sequencing, enormously comparable, second generation, or third generation (Fig. 5.1b, c). However, the previous Sanger sequencing technology is considered as the conventional or the first generation sequencing, which has taken a decade to complete the final draft of the human genome. Presently, NGS is mostly utilized in the research work (Behjati and Tarpey 2013). An alternate method of polymerase chain reaction (PCR) colonies, or polonies, has been used to amplify for a single template molecule. At a time, it can amplify the millions of clones (Mitra and Church 1999).

Recently, the proliferation of genome sequencing projects has encouraged a search for alternate approaches to decrease time and expense. Toward accomplishing about 100-fold expansion in throughput over Sanger sequencing technology, it has established an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing procedure enhanced for solid support and picoliter-scale capacities. The pyrosequencing technique has demonstrated the efficiency, accuracy, throughput, and robustness of the system through de novo assembly and shotgun sequencing of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy (Margulies et al. 2005). Currently, the virome analysis relies on deep sequencing, NGS data, and nucleic acid databases like the National Center for Biotechnology Information (NCBI), DNA Data Bank of Japan (DDBJ), and European Molecular Biology Laboratory Report (EMBL) databases. There are two frequently used NGS platforms: Illumina and Ion Torrent, which advocate maximum fragment lengths (300 to 400 nucleotides) for assessment separately. Recently, NGS technology has been enhanced with real-time sequencing

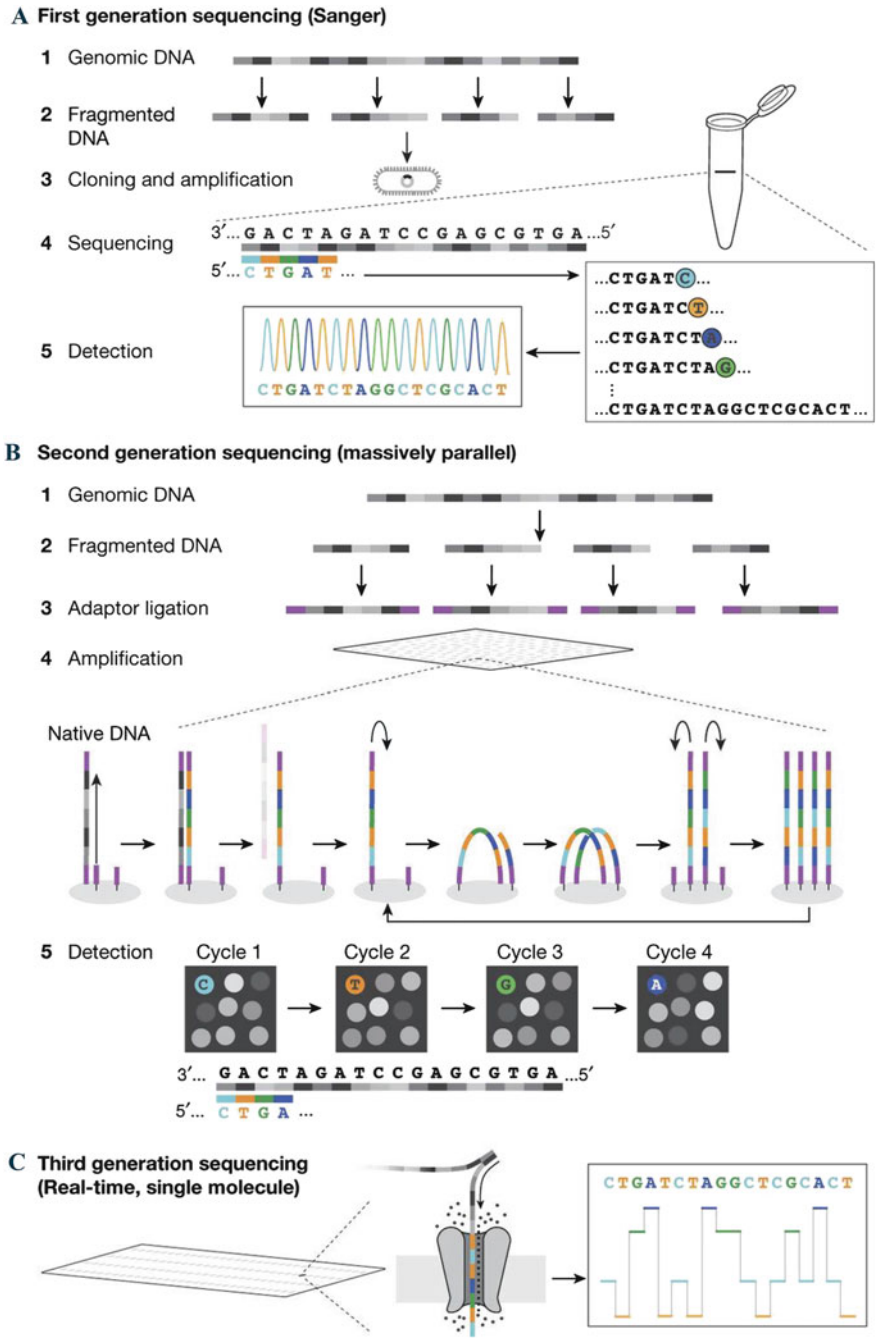


Fig. 5.1 DNA sequencing technologies. (**a**) Schematic examples of first, (**b**) second and (**c**) third generation sequencing are shown. Second generation sequencing is also referred to as next-generation sequencing (NGS) in the text (Shendure et al. 2017. Adopted with permission)

at single-molecule level (third generation) (Fig. 5.1c). It has indicated that the sequencing at single-molecule level will help in the advancement of various fields of biology, including virology, cancer biology, metagenomics, transcriptomics, and bioremediation (Ramamurthy et al. 2017; Schloss and Handelsman 2004; Bharagava et al. 2019; Jaswal et al. 2019; Costeira et al. 2019; Goto et al. 2019).

The advancement in DNA sequencing technology, NGS method, has supported to understand the microbiome diversity and also its characterization at a molecular level (Zaura 2012; Cao et al. 2017). The software developed for conventional sequencing technologies is becomes frequently incompetent to agree with nature of NGS technologies, which produce small and massively parallel reads. It has shown an amplification-free approach for analysis of the nucleotide sequences of about 0.3 million individual DNA molecules concurrently (Stein 2011; Harris et al. 2008). These findings have helped to understand the comparative analysis of human genomics.

5.3 Application of NGS in Various Fields

The NGS produces an amazing vision into metagenomes, metatranscriptomes, and metabarcodes of various organisms such as viruses, archaea, bacteria, and eukaryotes (Faure and Joly 2015; Bruno et al. 2015). The NGS tools have established the competence to sequence DNA at a unique speed, thereby allowing previously thought inconceivable scientific attainments and new biological applications. However, a significant challenge for data storage, analyses, and management solutions has been arising due to enormous data generated by the NGS (Zhang et al. 2011). Innovative bioinformatics tools are important for the effective management of NGS data. Through exploiting NGS approaches, investigators have identified and analyzed the important genes (Bai et al. 2012).

Multiplex ligation-dependent probe amplification and Sanger sequencing have been utilized for the genetic analysis of BReast CAncer genes 1 and 2 (BRCA1/2) which include the assessment for single nucleotide variants and insertion/deletion and for larger copy number variations (CNVs). Due to the introduction of NGS, it has become possible to specify the CNV information and sequence data (Schmidt et al. 2017). On the other hand, the NGS technique has been utilized to understand the molecular mechanism involved in gene regulation in hypertension. A remarkable study has identified the several genes' loci associated with different cardiovascular diseases (Costa and Franco 2017). The influence of NGS technologies on genomics will be very useful in the understanding of such type of complex diseases.

With the introduction of NGS tools, genome sequencing has become reasonable for regular genetic analyses. It has much facilitated the understanding of pathways involved in disease progression and analyses of rare genetic information of complex traits from the large datasets (Weissenkampen et al. 2019). The NGS has significantly extended our knowledge and skill to identify and characterize the gene and genetic composition from the microbial communities. Moreover, it has facilitated the analysis of microbiome with the help of metagenomics (culture-independent)

approaches (Song et al. 2013). In the metagenomics, total genetic material is extracted from total communities and processed and sequenced simultaneously.

Interestingly, NGS has revealed the metabolic interaction between gut microbiome and host. Genome-scale modeling is an evolving method which has established the various relations (microbe–microbe, host–microbe) under biological environments (Sen and Orešič 2019). However, the mechanisms of these interactions are still unclear. Gradual advancement of NGS tools has helped to understand the transcriptomics and metatranscriptomics (Tarkkonen et al. 2017). Furthermore, RNA-Seq data allow the analysis of transcriptome in the absence of a reference genome (de novo assembly) (Saggese et al. 2018). Due to large datasets generated from NGS tools, sophisticated in silico tools and skilled person are required.

Soil microbial communities are directly affected through natural environmental conditions, and functions are also fluctuating (Barboza et al. 2018). Taxonomic and functional profiles of soil samples can be analyzed by NGS approaches. On the other hand, bioremediation is usually observed as one of the effective methods to clean the environment with the help of microorganisms, instead of conventional physical and chemical methods. Due to an emerging concept of metagenomics, it can be utilized to understand the active microbial species, beneficial genes, enzymes, and bioactive molecules from the particular environmental sample. Such microbial species or specific gene can be utilized for effective bioremediation of the particular biohazardous compounds (Marco 2008; Ju and Zhang 2015; Czaplicki and Gunsch 2016; Conrads and Abdelbary 2019).

In addition, the series of omics, like metagenomics, metatranscriptomics, metaproteomics, metabolomics, and fluxomics, is also being utilized in the characterization, identification, and selection of particular strain of microbes (Schloss and Handelsman 2004; Bharagava et al. 2019). Therefore, a multi-omics approach, like metagenomics, metatranscriptomics, metaproteomics, and metabolomics, provides an excellent way to understand the metabolic pathway and microbes, which are involved in the bioremediation of particular contaminants in the environmental sites. Utilizing these approaches in the establishment of microbial consortium may be useful to provide specific microbial strain for the degradation of a particular contaminant from the environments.

5.4 Challenges of DNA Sequencing

Advancement in DNA sequencing technology like NGS and availability of international nucleotide sequence database collaboration (INSDC), and in silico tools (Softwares) significantly help to rapidly generate the genome sequences and understand the functional genomics of any organism. However, the analysis of NGS data requires sophisticated in silico tools, expensive infrastructure, and skilled person (Iacoangeli et al. 2019). The interpretation and characteristic of the NGS data have been challenged at several steps; such as sequencing errors, storage, algorithm, and statistical analyses (Zaura 2012).

In the recent past, rapid technological developments directed by academic institutions and companies are continuous to extend NGS methods from basic research to the clinical applications. However, the NGS implementation offers various process such as sequence analysis, storage, and quality control (Xuan et al. 2013; Vallenet et al. 2017). Genome-wide association studies (GWAS) have developed an applied technique to identify the genetic loci associated with disease by examining numerous markers throughout the genome. Ultimately, NGS has gained increasing popularity in the current years through its ability to analyze a much larger number of markers throughout the genome. Though NGS platforms have accomplished examining a higher number of single nucleotide polymorphisms (SNPs) associated with GWA studies (GWAS) (Alonso et al. 2015). In the case of GWAS, trivial effects started large sample sizes, usually made possible through meta-analysis by exchanging summary statistics throughout consortia. While NGS studies groupwise test for the association of multiple potentially causal alleles by every gene. Therefore, they have developed MetaSeq, a procedure for meta-analysis of genome-wide sequencing data, and it is publicly available as open source (Singh et al. 2013). The results gained by the NGS need thorough analysis, as their biological relationship is not well understood.

Numerous emerging biological applications, such as targeted exome sequencing, chromatin immunoprecipitation sequencing (ChIP-Seq), and whole transcriptome shotgun sequencing technology or RNA-Seq, have been established to accomplish various biological determinations. Exome sequencing (Mamanova et al. 2010) affects the disadvantage of the high expense of sequencing the whole genome without intronic regions, and selectively sequencing the exonic regions, which might be of further direct interest. ChIP-Seq (Johnson et al. 2007) is used to study protein–DNA/RNA interactions while RNA-Seq (Mortazavi et al. 2008) is used to exploit the NGS technologies to sequence cDNAs. The gathering of NGS reads leaves over a puzzling assignment. It is particularly true for the assembly of environmental samples that originate through metagenomics approaches possibly containing huge microbial diversity (Warnke-Sommer and Ali 2016). Nevertheless, due to lack of availability of the reference genome, it creates problems while executing the conclusion of analyzed data (Hiendleder et al. 2005). Moreover, NGS has facilitated the understanding of microbiomes associated with infection, environmental, bioremediation, and diversity of agriculturally important microorganisms. Recently, solid-state nanopore-based NGS has been developed. It has demonstrated that advancement in the technology, such as the processability, the robustness, and the large-scale integratability (Goto et al. 2019) and other sequencing platforms, is under development.

5.5 Concluding Remarks

This chapter analyzes an innovation in the DNA sequencing technology, like second and third generation sequencing techniques, which has demonstrated the efficiency, accuracy, throughput, and robustness of the system. In addition, it is

significantly helps to generate rapid and large datasets, like metagenome and metatranscriptome. Moreover, the availability of various biological databases, such as NCBI-GenBank, DDBJ, EMBL-EBI (European Bioinformatics Institute), and in silico tools (Softwares), considerably helps to understand the functional genomics of the organisms. Exploiting the NGS technology and databases in the field of microbiology, metagenomics molecular method is one of the powerful culture-independent approaches to understand the microbial diversity and mine a specific gene from any sample such as environmental, human gut, and rumen. The advantages of NGS are the high throughput, low cost, and accuracy of the data and exponentially support the NGS datasets. However, there are sequencing artifacts (low-quality and contaminated reads) that need to be tackled when using the NGS analysis. A major DNA sequencing method, with high precision, extended read length, and high throughput, would be necessary for further developments of fields.

Competing Interests There is no competing interest.

References

Alonso N, Lucas G, Hysi P (2015) Big data challenges in bone research: genome-wide association studies and next-generation sequencing. Bonekey Rep 4:635

Bai Y, Sartor M, Cavalcoli J (2012) Current status and future perspectives for sequencing livestock genomes. J Anim Sci Biotechnol 3(1):8

Barboza ADM, Pylro VS, Jacques RJS, Gubiani PI, de Quadros FLF, da Trindade JK, Triplett EW, Roesch L (2018) Seasonal dynamics alter taxonomical and functional microbial profiles in Pampa biome soils under natural grasslands. PeerJ 13(6):e4991

Behjati S, Tarpey PS (2013) What is next generation sequencing? Arch Dis Child Educ Pract Ed 98(6):236–238

Bharagava RN, Purchase D, Saxena G, Mulla SI (2019) Applications of metagenomics in microbial bioremediation of pollutants: from genomics to environmental cleanup. In: Microbial diversity in genomic era. Academic Press, London, pp 459–477

Blow N (2009) Transcriptomics: the digital generation. Nature 458(7235):239–242

Bruno F, Marinella M, Santamaria M (2015) e-DNA meta-barcoding: from NGS raw data to taxonomic profiling. Methods Mol Biol 1269:257–278

Cao Y, Fanning S, Proos S, Jordan K, Srikumar S (2017) A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. Front Microbiol 8:1829

Chakraborty C, Doss CG, Patra BC, Bandyopadhyay S (2014) DNA barcoding to map the microbial communities: current advances and future directions. Appl Microbiol Biotechnol 98(8):3425–3436

Chiu CY, Miller SA (2019) Clinical metagenomics. Nat Rev Genet 20(6):341–355

Conrads G, Abdelbary MMH (2019) Challenges of next-generation sequencing targeting anaerobes. Anaerobe 58:47

Costa A, Franco OL (2017) Impact and influence of "omics" technology on hyper tension studies. Int J Cardiol 228:1022–1034

Costeira R, Doherty R, Allen CCR, Larkin MJ, Kulakov LA (2019) Analysis of viral and bacterial communities in groundwater associated with contaminated land. Sci Total Environ 656:1413–1426

Craig DW, Nasser S, Corbett R, Chan SK, Murray L et al (2016) A somatic reference standard for cancer genome sequencing. Sci Rep 6:24607

Czaplicki LM, Gunsch CK (2016) Reflection on molecular approaches influencing state-of-the-art bioremediation design: culturing to microbial community fingerprinting to omics. J Environ Eng 142(10):03116002. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001141

Faure D, Joly D (2015) Next-generation sequencing as a powerful motor for advances in the biological and environmental sciences. Genetica 143(2):129–132

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269(5223):496–512

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270(5235):397–403

Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M et al (2016) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 15(3):201–206

Goto Y, Akahori R, Yanagi I (2019) Challenges of single-molecule DNA sequencing with solid-state nanopores. Adv Exp Med Biol 1129:131–142

Guillermin Y, Lopez J, Chabane K, Hayette S, Bardel C, Salles G, Sujobert P, Huet S (2018) What does this mutation mean? The tools and pitfalls of variant interpretation in lymphoid malignancies. Int J Mol Sci 19(4):E1251

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320(5872):106–109

Hiendleder S, Bauersachs S, Boulesteix A, Blum H, Arnold GJ, Frohlich T, Wolf E (2005) Functional genomics: tools for improving farm animal health and welfare. Rev Sci Tech 24(1):355–377

Hu P, Shao Y, Xu J, Wang T, Li Y et al (2019) Genome wide study on genetic diversity and phylogeny of five species in the genus Cervus. BMC Genomics 20(1):384

Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR et al (2019) DNAscan: personal computer compatible NGS analysis, annotation and visualisation. BMC Bioinf 20(1):213

Jaswal R, Pathak A, Edwards B III, Lewis R III, Seaman JC, Stothard P, Krivushin K, Blom J, Rupp O, Chauhan A (2019) Metagenomics-guided survey, isolation, and characterization of uranium resistant microbiota from the Savannah River Site, USA. Genes 10(5):E325

Jeon K, Min B, Park JS, Kang YK (2017) Simultaneous methylation-level assessment of hundreds of CpG sites by targeted bisulfite PCR sequencing (TBPseq). Front Genet 8:97

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830):1497–1502

Ju F, Zhang T (2015) Experimental design and bioinformatics analysis for the application of Metagenomics in environmental sciences and biotechnology. Environ Sci Technol 49(21):12628–12640

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. Cell 155(1):27–38

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. Nat Methods 7(2):111–118

Marco D (2008) Metagenomics and the niche concept. Theory Biosci 127(3):241–247

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380

Marian AJ (2011) Medical DNA sequencing. Curr Opin Cardiol 26(3):175–180

Messing J (2001) The universal primers and the shotgun DNA sequencing method. In: Graham CA, Hill AJM (eds) DNA sequencing protocols. Methods in molecular biology™, vol 167. Humana Press, Totowa

Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res 27(24):e34

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5(7):621–628

Ramamurthy M, Sankar S, Kannangai R, Nandagopal B, Sridharan G (2017) Application of viromics: a new approach to the understanding of viral infections in humans. Virus Disease 28(4):349–359

Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-seq. Bioinformatics 27(17):2325–2329

Runtuwene LR, Tuda JSB, Mongan AE, Suzuki Y (2019) On-site MinION sequencing. Adv Exp Med Biol 1129:143–150

Saggese I, Bona E, Conway M, Favero F, Ladetto M, Liò P, Manzini G, Mignone F (2018) STAble: a novel approach to de novo assembly of RNA-seq data and its application in a metabolic model network based metatranscriptomic workflow. BMC Bioinf 19(Suppl 7):184

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Ehrlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 293:487–491

Schloss PD, Handelsman J (2004) Status of the microbial census. Microbiol Mol Biol Rev 68:686–691

Schmidt AY, Hansen TVO, Ahlborn LB, Jønson L, Yde CW, Nielsen FC (2017) Next-generation sequencing-based detection of germline copy number variations in BRCA1/BRCA2: validation of a one-step diagnostic workflow. J Mol Diagn 19(6):809–816

Sen P, Orešič M (2019) Metabolic modeling of human gut microbiota on a genome scale: an overview. Metabolites 9(2):E22

Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. Nature 550(7676):345–353

Singh AP, Zafer S, Pe'er I (2013) MetaSeq: privacy preserving meta-analysis of sequencing-based association studies. Pac Symp Biocomput:356–367. https://doi.org/10.1142/9789814447973_0035

Smith M (2017) DNA sequence analysis in clinical medicine, proceeding cautiously. Front Mol Biosci 4:24

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321(6071):674–679

Song S, Jarvie T, Hattori M (2013) Our second genome-human metagenome: how next-generation sequencer changes our life through microbiology. Adv Microb Physiol 62:119–144

Stein LD (2011) An introduction to the informatics of "next-generation" sequencing. Curr Protoc Bioinformatics 36(1):11.1.1–11.1.9

Tarkkonen K, Hieta R, Kytölä V, Nykter M, Kiviranta R (2017) Comparative analysis of osteoblast gene expression profiles and Runx2 genomic occupancy of mouse and human osteoblasts in vitro. Gene 626:119–131

Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A et al (2017) MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res 45(D1):D517–D528

Warnke-Sommer J, Ali H (2016) Graph mining for next generation sequencing: leveraging the assembly graph for biological insights. BMC Genomics 17:340

Weissenkampen JD, Jiang Y, Eckert S, Jiang B, Li B, Liu DJ (2019) Methods for the analysis and interpretation for rare variants associated with complex traits. Curr Protoc Hum Genet 101(1):e83

Won YJ, Kang LH, Lee SG, Park SW, Han JI, Paik SY (2019) Molecular genomic characterization of severe fever with thrombocytopenia syndrome virus isolates from South Korea. J Microbiol 57:927. https://doi.org/10.1007/s12275-019-9174-8

Xuan J, Yu Y, Qing T, Guo L, Shi L (2013) Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 340(2):284–295

Zaura E (2012) Next-generation sequencing approaches to understanding the oral microbiome. Adv Dent Res 24(2):81–85

Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. J Genet Genomics 38(3):95–109

Biomimetic Approaches in Synthetic Biology

6

Renuka Suravajhala, Aakanksha Kumar,
Harshavardhan Reddy Burri, and Babita Malik

Abstract

Biomimetics is an advanced method to create full or partial replicates in stimulating biological system at macro and nanoscale. There are several inspired innovations that are developed to resolve the biological problems in efficient method. This chapter briefs development of engineering designs for recognition of biological phenomena accelerating the synthetic biology research. In addition, we describe the strategies for finding potential analogies in biological phenomena, including searching functional repertoire across multiple levels of organization, from the molecule to the biosphere. In addition, initial efforts at finding appropriate analogies are documented using an example.

Keywords

Redesigning · Metabolic pathways · Analogy · Function · Synthetic molecules · Remanufacturing

6.1 Introduction

The redesigning technology has emerged in the form of synthetic biology to explore and understand different mechanisms involved in living materials such as proteins, genes, complex systems, and metabolic pathways (Sismour and Benner 2005). Synthetic biology has been potentially focused by studies with respect

R. Suravajhala (✉) · B. Malik
Department of Chemistry, School of Basic Sciences, Manipal University Jaipur, Jaipur, Rajasthan, India
e-mail: renuka.181017602@muj.manipal.edu; babita.malik@jaipur.manipal.edu

A. Kumar · H. R. Burri
Bioclues.org, Hyderabad, Telangana, India

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_6

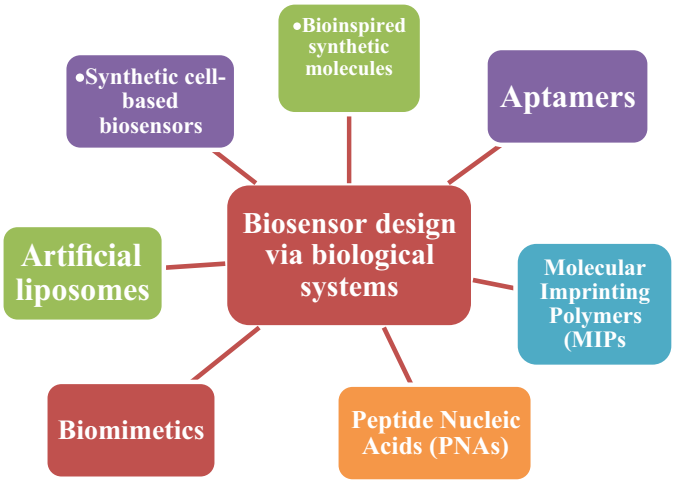


Fig. 6.1 Scheme representation of main application fields of designed biosensors

to sensitivity and stability that have a key role to improve potential biosensors. This can be considered selectively by optimizing complex engineered biological systems (Andrianantoandro et al. 2006). The recent convergence in interdisciplinary approaches of various technologies include genetics, chemistry, engineering, biology, mathematics and molecular simulations, modeling as a part of synthetic biology. They are used for new innovations which are of major concern in developing new drug materials, nanomaterials, renewable biomass and fuels (Lee and Na 2013). In the recent past, nanotechnology-based synthetic biology has been constantly explored with increased potential applications and broad spectrum in biomedical avenues, biopharma, biodefense, environmental, agri-feed (Fig. 6.1). Synthetic biology was worth US$ 1,537 million in 2011 in the global market and reached approximately US$ 16.745 million by 2018. Thus, the annual growth rate between 2010 and 2018 has increased about 41% (Bhatia and Chugh 2013). There have been several new biological approaches, redesigns of pre-existing elements developed through synthetic biology as biosensors (Marner 2009). Recent advances in technology for molecular engineered molecules via computational biology and nanotechnology are used to reassemble genes, enzymes through modifications that serve as specific molecular substrates or organisms to function efficiently (Effendi et al. 2008). In the recent past, emerging designs through combinatorial library of biomolecules that enables specificity and stability were developed via biosensors, mimetics, molecular imprinting polymers (MIPs), and aptamers (Checa et al. 2012).

Biosensors have emerged from synthetic biology concepts with a wide range of applications wherein the first synthetic biosensors were designed from genetically modified organisms to detect specific pollutants that cause adverse environmental condition. Several biosensors such as transcriptional biosensors affecting the genes transcription, translational biosensors, variation in gene expression, post-

translational biosensors via modifying signal transduction pathway drug response have been widely employed (Khalil and Collins 2010). Biosensor technology via biosynthetic materials has paved way for potential applications with a significant role in healthcare. Biosensors were also employed for abnormal cells and drug-related cancer/arterial diseased cells (Neumann and Neumann-Staubitz 2010). They have also been employed for detection of harmful chemicals, explosives relating to environmental and Agri-feed besides varied applications in converging technologies adopted in synthetic biology and biomimetic chemistry (Bensaude-Vincent 2009). In this process, artificial new entities were created for biosensor application via synthetic biology, and novel genetic materials were designed and are known as biomimetics (Fig. 6.1).

6.2 Cell-Based Synthetic Biosensors

Biosensors that are cell-based have limited applications due to short life, whereas biosensors designed from single molecule are known to be more stable with less functionality (Zhao et al. 2006). Synthetic materials based on its efficiency as artificial sensors such as mimic cells have been streamlined to detect durability and sensitivity in modified cells. In addition, synthetic biosensor efforts have been made in modeling, and simulation studies to cut down the time consumption in conventional methodologies have resolved several challenges related to selectivity in the field of biosensor. Several natural materials were redesigned for obtaining specific functionality such as synthetic metabolic pathways, minimal cells, and genetic circuits. Furthermore, studies have shown that efficient biosensor was developed by coupling gene networks (Lu et al. 2009; Siuti et al. 2013). The microorganism-based genetic circuit was developed as synthetic cellular sensors have shown protein signals. While specific protein-coding genes with regulatory circuited synthetic sensors are expressed, it releases the protein to a specific target (Fig. 6.2).

In 1970s, *Salmonella enterica* cells were the first bacterial cells that were redesigned as (*enterica serovar Typhimurium str. L2*) showing positive response towards mutagenicity tests on various chemicals (Ames et al. 1973). The *Escherichia coli* derived protein A (RecA)-LexA is known to regulate SOS response. The SOS regulates the induced genes and leads to production of two proteins, that is, LexA (a transcriptional repressor protein) and RecA that undergo autocatalytic coprotease self-cleavage of LexA. It has been known that RecA protein forms nucleoprotein filaments on single- and double-stranded DNA control damage. While a promoter-specific 20-nucleotide site is linked with LexA repressor protein-resisting gene expression, UV radiation or chemical-induced DNA damage is controlled by restricted DNA synthesis and cell division (Khalil and Collins 2010). Similarly, new programmable cell was developed from*E. coli.* For example, *Vibrio fischeri* sensing transgenic signaling pathway detected acyl-homoserine lactone (AHL) molecules (Kobayashi et al. 2004). These designed synthetic cells produce a targeted protein in specific density of cell population. Such studies in

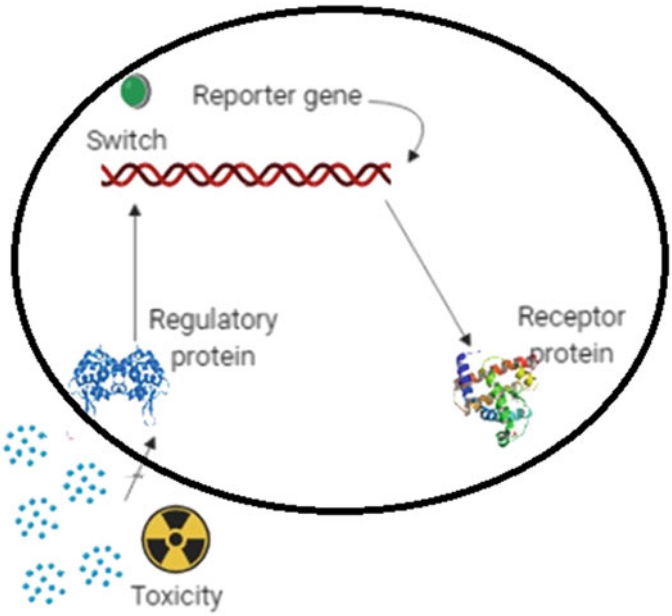


Fig. 6.2 Synthetic regulatory circuit sensor—the active reporter gene encode to a reporter protein assigned to specific response can be studied based on bioreporter cell response toxicity studies

synthetic biology proved that biosensing applications in novel gene programmable cells serve as regulatory circuits of the cell. Furthermore, a microbial biosensor developed to detect iron uptake showed different Fe II concentrations measured by reporter protein expression (Quintero et al. 2007).

To develop biosensors, major challenges such as standard protocol for production and validation are essential, providing cellular environment with intercellular variability and specific functionality is more challenging. One such challenge is seen in development of "biopixels" sensing gas-phase redox signaling between colonies synchronizing thousands of oscillations in a liquid crystal display (LCD). Thus, the ability to develop large-scale cellular coordinates of genetic biosensors serves as efficient, fast, and cost-effective detectors (Prindle et al. 2012).

Novel antimicrobial synthetic biosensors to detect infectious pathogens were developed from *E. coli* to detect *Pseudomonas aeruginosa*. The schematic representation in (Fig. 6.3) depicts the *P. aeruginosa* sensing mechanism, consists tetR promoter produces LasR (a transcriptional factor) bind to AHL 3OC12HSL forming LasR-3OC12HSL complex, binds with luxR inducible promoter in the sensor device. It has led to pyocin S5 and lysis E7 proteins production in *E. coli*, and its accumulation diffuses in the pathogen leading to cellular integrity damage through the acyl homoserine lactones (AHLs) (Saeidi et al. 2011).

Synthetic biosensors have not only used to optimize specific metabolites in microorganism's assimilation but also *Saccharomyces cerevisiae* recombinant

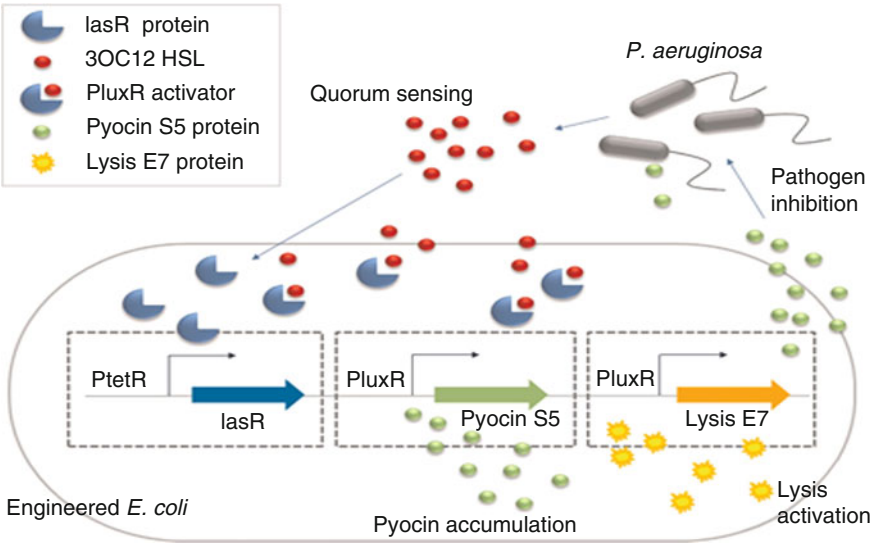


Fig. 6.3 The process to locate and destroy has been studied at various concentrations of 3OC12 HSL in *P. aeruginosa* (Source: Saeidi et al. 2011)

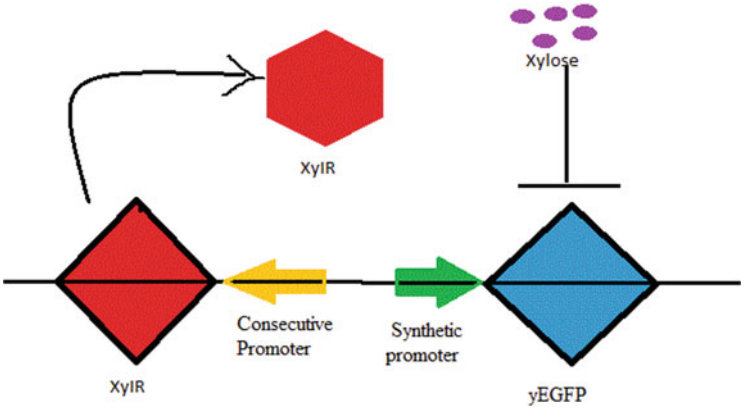


Fig. 6.4 The xylose biosensor developed from XylR repressor that binds to hybrid synthetic promoter blocks XylR operator sequence transcription was detected by xylose

strains which cannot ferment xylose but can assimilate xylose. To support this, synthetic operon of yeast was cloned with PGK1 and PFY1 promoters coding sequences up streamed in XylR, encoding the sequence the yEGFP reporter protein was known to be cloned XylR operator sequence at downstream (Teo and Chang 2015) (Fig. 6.4). Thus, translated biochemicals intracellular concentration of interest with high efficiency enabled protein synthesis to regulate and minimize stress through xylose flux pathways at cellular level (Teo and Chang 2015). Simarly,

RNA-based sensors with complete cell linked to reporter genes and activators applied to sensitize the concentration of small molecules growth or toxicity in cells were reported. The intracellular metabolite levels detected by RNA-based sensor known as riboswitches, wherein an aptamer encoded with gene-regulatory function binds to target metabolite thus paving way for conformational changes in riboswitch in biosynthetic pathways (Michener et al. 2012). On the other hand riboswitch as fluorescent biosensors is used to detect the second messenger cyclic di-AMP (cdiA) in bacteria and archaea were developed (Kellenberger et al. 2015). Furthermore, RNA-based biosensors developed from different gram-positive, negative bacteria, archaea potentially efficient in *in vivo* metabolite imaging, besides enhancing unidentified signaling domains.

Biosensors based on synthetic electrochemical designs in which optically absorbing/fluorescing compounds were detected with high efficiency, sensitivity. Electrochemical biosensor viz ElectrEcoBlu, a microbial fuel cell converts input pollution to output as measurable electrical signal. The device contain recognition element (a pollutant, i.e., XylR or DntR, that bind to chemicals such as BTEX (benzene, toluene, ethylbenzene, and xylene) and dinitrotoluene) were detected as input signals associated with S-adenosylmethionine dependent N-methyltransferase (PhzM), hydroxylase (PhzS) enzymes. Thus, in *P. aeruginosa*, phenazine-1-carboxamide (PCA) is known to be converted to pyocyanin (PYO) and cloned to *E. coli*, Pseudomonas strain a non-pathogenic bacteria (Gu et al. 2010). Such approaches lead to designing novel biosensors that are cost effective and user friendly to detect targets in real time.

6.3 Artificial Liposomes

The lyotropic structural and chemical properties of lipid membrane have emerged with wide range of synthetic lipid nanostructures as artificial cells with biosensing application. Semi-disposable microfluidic biosensor was synthesized polydimethylsiloxane (PDMS) mounted with polymethyl methacrylate (PMMA) lead to two different immobilized DNA sequences wherein, the biorecognition of unique sequences of RNA targeted pathogen was reported (Kwakye and Baeumner 2003). First, synthetic liposomes are coupled, encapsulated with nucleic sequence as a dye, and later coupled with second nucleic sequence through superparamagnetic beads. Thus, biosensors are known to have increased sensitivity and reduced analysis time. Studies also further showed biosensors for dengue fever virus developed via liposomes have comparatively 100 times reduced detection than membrane-based biosensors. Such cell-based artificial biosensor was developed mimicking complete cell helped to detect protein toxicity in listeriolysin O (LLO). LLO is a foodborne pathogen that secrets *Listeria monocytogenes* for bacterial virus present in the pores form hemolysin. The fluorescent dye based unilamellar liposomes immobilized with porous silica synthesise lead to LLO membrane insertion and pore formation (Baeumner et al. 2002). Thus, whole cell biosensors were known to show high detection rate than that of engineered sensors and showed less efficiency in severe

growth conditions. The designed LLO showed fast detection less than 30 min with high stability approx, 5 months is more efficient than that of conventional methods that take 5–7 days to identify *L. monocytogenes* in contaminated food product (Zhao et al. 2006).

6.4 Synthetic Biosensor

Synthetic biosensors are synthetic and systems biology applications of genetically engineered materials. In this, analyte binds selective transcripted inducer or repressor genes (a reporter) that encode specific protein as output signal (Choffnes et al. 2011). Thus, by modifying natural genomes, novel organismal are created from existing DNA via constructing or manipulating with new functions. The following section details how "biosensor biomimetic chemistry" has emerged from "synthetic biology," wherein mixed and matched to design artificial genetic materials that imitate biochemical pathways synthesizing antibiotics, proteins, and enzymes (Benner and Sismour 2005; Sarikaya et al. 2003).

In the recent years, new materials are explored to meet the requirements of designing biosensors. Several studies related to designing of synthetic molecules such as aptamers, biomimetics, molecular imprinting polymers, peptide nucleic acids, and ribozymes lead to biomimetic detectors in resolving the drawbacks related to sensitivity, stability, and reliability in process of designing biosensors (Fig. 6.5).

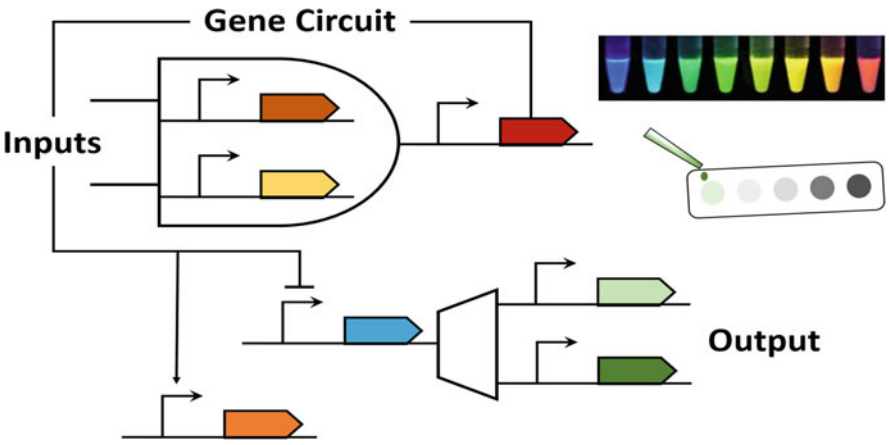


Fig. 6.5 Various synthetic biosensors and its detection (Courtesy: image http://synbiolab.org/research.htm)

6.5 Biomimetics

The designing of protein-based biosensors has several limitations that depend on the biological recognition receptors. The protein-based designing requires large amounts of production, purification with respect to specific recognizing element and fluorescent dyes labeling requires huge man hours to have high throughput analysis. Thus, denaturation, microarray formats play major role while configuring stable sensor molecules.

In this context, biomimetics serve as potential new entities in genetic engineering (Lepora et al. 2013) wherein automated chemical synthesis leads to stable 3D peptide sequences that are versatile and undergo mutagenesis efficiently to that of proteins.

As synthetic peptides-based biosensor designs are able to overcome storage, dry, dissolved, and immobilized states (Scognamiglio et al. 2013), a synthetic peptide with 18 amino acids (pTMVP) with 6-bromomethyl-2-(2-furanyl)-3-hydroxychromone is produced and functionalized. The sensitivity of fluorophore emits two-bands that detected, Fab57P, a fragment of recombinant antibody in tobacco mosaic virus coat protein (TMVP) (Enander et al. 2008). The combined ability of peptides and fluorophores can detect efficiently the immobilized phase through signaling assays. The biosensor had been developed as clinical diagnosis for hepatitis G virus infection (GBV-C) with chimeric multiple antigenic peptides (MAPs) having E2, NS4, and NS5 proteins for hepatitis C virus (HCV) humans (Gómara et al. 2010). On the other hand, a sensor developed from protein/peptide with specific antibodies in anti-delta genotype of hepatitis D virus infected humans detected by surface plasmon resonance imaging, wherein the mimetic peptides are combined with recombinant proteins which are sensitive to surface plasmon resonance imaging (SPRi), leading to efficient hepatitis viruses detection on a wide range of samples (Villiers et al. 2015).

Nanotechnology-based mimetic peptides with fluorescent dyes are used by connecting with terminal carboxyl groups of quantum dots (QDs) surface monitoring chemical and biological molecules mechanisms (Prasuhn et al. 2010). An example for such sensor was functionalized QDs labeled dye peptides via fluorescence resonance energy transfer (FRET) to monitor enzymatic activity of downstream effector: caspase 3 in apoptosis, Ca2 ions unknown mechanism in biological system (Fig. 6.6). Studies on synthetic sensor developed from dimeric Cu (II) complex ([Cu(m2-hep) (hep-H)]2_{*}2ClO4) coupled with silver nanoparticles in detecting various drug sensitivities have shown that the biomimetic sensors have broad spectrum of applications such as surface modifications, sensitivity, low cost were considered to the determination of catecholamines.

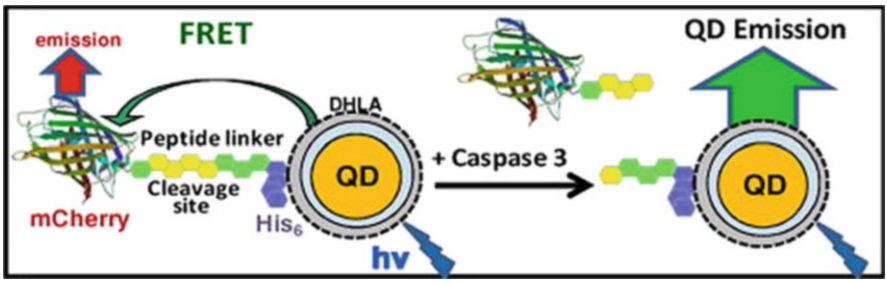


Fig. 6.6 Emission of QD proximity QD-FRET from chromophore only. Thus, split in two creates caspase activity from which FRET disappear (Dennis et al. 2012)

6.6 Molecularly Imprinting Polymers (MIPs)

The molecularly imprinting polymers (MIPs) serve as potential mimic antibodies applied in developing diagnostic tools with high stability, affinity, selectivity (Piletsky et al. 2001a). MIPs are synthetic artificial materials that recognize the binding to specific target. The functional monomers undergo polymerization around a template molecule as target molecule via covalent or non-covalent bonds. The target molecule was removed resulting in specific size, shape, and functional groups positions created as recognition sites (Fig. 6.7). While the biological activity of receptor was mimicked with synthetic molecule, MIP has enhanced specificity and stability, ligand binding studies were developed (Ye and Haupt 2004). The MIPs are resistant to high temperature, chemicals, enzymatic degradation, and low pH. These

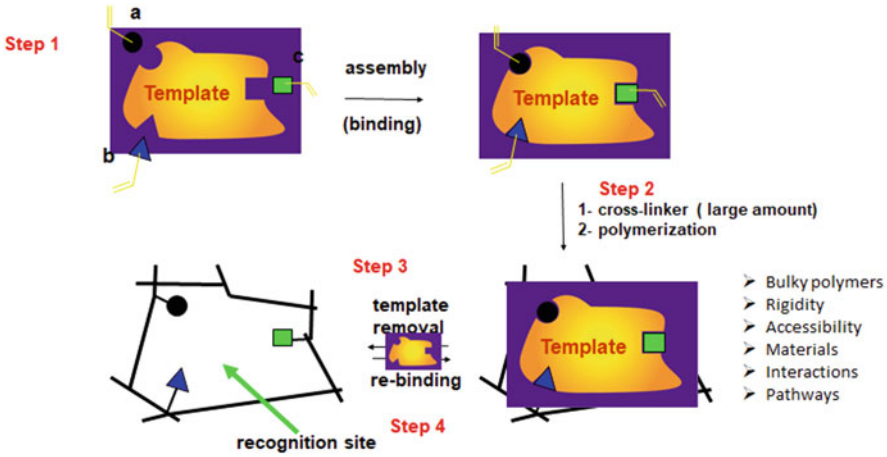


Fig. 6.7 Molecular imprinted polymers with functional monomer and template undergo covalent or non-covalent interactions lead to polymerization was followed by template removal lead to specific recognition sites

are easily soluble in aqueous solution, organic solvents, and cost effective compared to that of traditional polymerization (Nicholls et al. 2001). Thus, MIPs are efficient biosensors explored in food, pharmaceutical, environmental related studies. MIPs are also combined with optical sensors for detecting fluorescence transduction in spectroscopic studies and developed electrochemical sensors with high sensitivity to detect conductivity, potential, voltage. MIPs based sensors are developed to detect chemicals that are monoamine naphthalenes, phenols, 2,4-dinitrophenol, parathion, microcystin-LR, herbicides food contamination, and heavy metals were detected (Valero-Navarro et al. 2009; Sergeyeva et al. 2010; Liu et al. 2012; Marx et al. 2004; Chianella et al. 2003; Pardieu et al. 2009; Liang et al. 2009). The MIPs based sensors to monitor serotonin levels in human blood plasma, creatinine, theophylline were studied (Peeters et al. 2012; Subrahmanyam et al. 2001; Kindschy and Alocilja 2007). Thus, the high stability and non-degradability of MIPs serve as alternate for conventional antibodies or receptors (Piletsky et al. 2001a, b). Furthermore the MIPs based aniline and thiophene derivatives are developed to study environmental and drug screening analysis. Studies related to cardiac biomarker were developed from non-imprinted polymer (NIP) reported with high affinity that leads to acute myocardial infarction (Karimian et al. 2013). Similarly, the MIPs based thin gold chip was developed to detect histamine levels.

6.7 Aptamers

Aptamers are short single-stranded nucleic acids or artificial peptide based molecules. Aptamers have strong specific binding affinity of target molecules like proteins and other amino acids and hence are referred as chemical antibodies (Suravajhala et al. 2014). These are selected from chemically synthesized combinatorial library of molecules that contain random sequences, depending on the affinity to target protein and are site-specific modified to insert reporters and highten their stability, specificity, and affinity (Mascini 2008). These are emerged as antibodies alternatives with high specificity and affinity, high shelf life, reduced size, low cost (Kanwar et al. 2015). Thus, aptamers are potential biomimetic components to reassemble as specific biosensors in recognizing a wide range of biomolecules such as small molecules, peptides, proteins, viruses, or cells that have unique binding constants ranging in between picomolar and nanomolar (Tombelli et al. 2007). The potential bioanalytical applications have been explored in clinical diagnostics, therapeutics and have been widened to signal transduction systemic studies. The optical sensitivity of the aptamers was done by various insert labels methods to label aptamer with quencher and a fluorophore resulted in "light on-off" sensors. A fluorescence variation is formed from the conformation of quencher and fluorophore when the ligand binds modifying the engineered molecule (Fig. 6.8).

Similarly, sensor designed based on fluorescent was developed. Wherein a DNA aptasensor for IgE and vasopressin, and a RNA aptasensor for tobramycin and vasopressin were attached with fluorescein and dabcyl as fluorophore/quencher pair that detected target induced strand releasing biosensor was developed (Pei 2011).

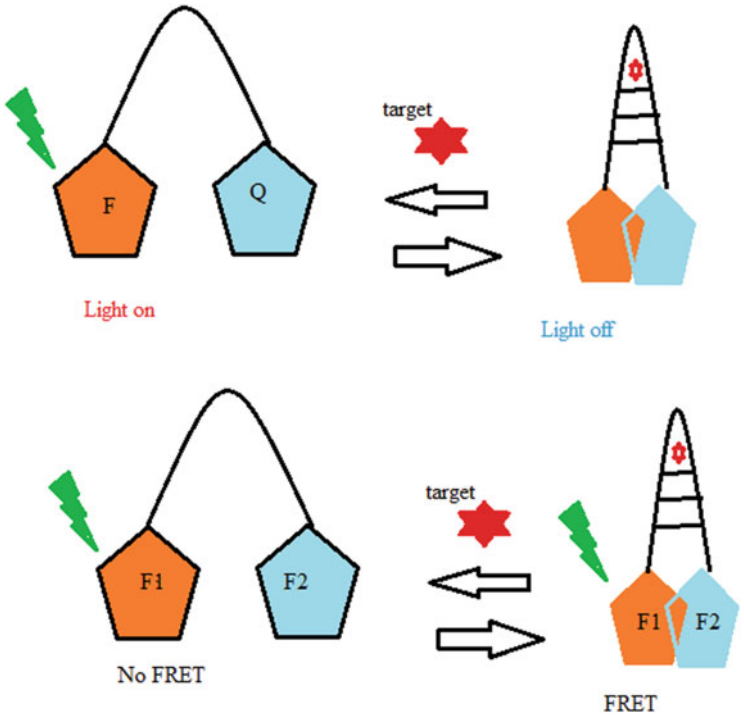


Fig. 6.8 Quenching aptamers to develop optical sensors and aptamers-based sensor via Forster resonance energy transfer (FRET)

These fluorescent aptasensors are known to integrate with many configurations that can lead to designing of powerful molecular sensor tools.

Fluorophore and quenching method are expensive and can lead to false signals due to ligands or solvents interference. Wherein the fluorescent molecules of FRET interact with fluorophores donor/acceptor at the aptamer terminal and the targets that are involved were determined by radiometric analysis (Fig. 6.7 bottom). A FRET based biosensor was designed by reconfiguring DNA-based aptamer active sites, in which the synthetic molecule with designed aptamer was hybridized with an oligonucleotide and doubly labeled with a fluorescent probes pair. A folded (high FRET) and an unfolded (low FRET) conformation depends on the absence or presence of the target molecule was studied (Buranachai et al. 2012).

6.7.1 Peptide Nucleic Acids (PNAs)

Macromolecules such as DNA and RNA are known to interacting with other nucleic acids, proteins and small molecules are known to perform key regulatory roles in the cell processes. Wherein the development of microarray, lab-on-chip for gene

expression pattern, detection of pathogens leads to development of DNA/RNA-based biosensors. Peptide nucleic acids (PNA) are artificially synthesized molecules that mimic DNA or RNA that are used to develop biosensors, antisense and antigen agents, and molecular probes (Peter and Michael 1999). The mimics DNA by replacing their deoxyribose phosphate backbone of nucleobases were linked with different peptide polymers. Wherein, PNAs have flexible and uncharged polyamide backbone, which gives them high affinity and specificity, which is utilized to hybridize with complementary DNAs or RNAs in a sequence-dependent manner. This makes PNA probes better than nucleic acids to be used as molecular probes in electrochemical biosensors. Thus high sensitivity, affinity, fast hybridization kinetics, shorter probe length as compared to DNA, independent by iconic strength and resistance to enzymatic degradation has resulted in several robust PNA-dependent procedures. Studies related to antigen, optimize PCR reactions, *in situ* studies on genomic mutation *via* labeling chromosomes, anticancer drugs efficiency detecting sensors designing and its development studies are reported (Franck et al. 2004) (Fig. 6.9).

Electrochemical biosensor to develop cervical cancer biomarker to detect the human papilloma virus (HPV) type 16 DNA was reported (Jampasa et al. 2014). Wherein anthraquinone was linked with pyrrolidinyl PNA (acpcPNA) and was immobilized on modified chitosan carbon electrode. Thus, redox active anthraquinone (AQ) was labeled with A 14mer acpcPNA probe to detect targets at 14 nucleotides of HPV_ 16 L1 gene. The HPV type 16 DNA was detected selectively. Thus, the studies reported that PNAs together with graphene oxide (GO) lead to the FRET sensor development that has enhanced PNA specificity and stability (Lee et al. 2014).

The PNA probe with graphene quencher was developed as dsDNA sensors, wherein, the dsDNA sensors undergo hybridization with targeted sequence emit fluorescent signals, high selectivity in protein serum sequences and avoiding denaturation. Similarly, PNAs immobilized on nano-graphene oxide (NGO) were developed as biosensor to detect microRNA (miRNA) in post-transcriptional reg-

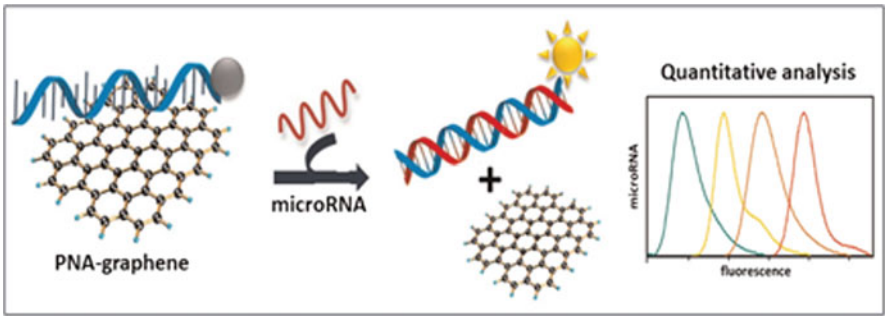


Fig. 6.9 The optical biosensors developed by nano-graphene oxide (NGO) based immobilized PNA regulates gene expression and detect microRNA (miRNA) (Ryoo et al. 2013) (Open access)

ulation processes (Ryoo et al. 2013) (Fig. 6.9). The abnormal expression levels of miRNAs make them important biomarkers linked with diseases like diabetes and cancer.

6.7.2 Ribozymes

Ribozymes are catalytic enzymes RNA molecules made up of nucleic acids that act as molecular scissors that can cleave other RNA molecules. They can repress few genes like those belonging to pathogenic viruses or cancer promoting genes. Ribozymes play vital role in construction of biomolecules as switches for synthetic chemical sensors (David and Nanette 2016). They are now discovered extensively in nature, present in plants, bacteria, viruses, and lower eukaryotes.

To construct artificial ribozyme RNA molecules, labeling chemistry and directed evolution are combined to create molecular switches which can regulate and influence biochemical processes like mRNA splicing, structural assembly, tRNA dispensing, and chaperone facilitated folding.

6.8 Future Perspectives

In the recent past, biosensors had remarkable potential in clinical diagnostics as biosensors are used to detect single or multi-array. Synthetic biology has not only paved emerging ways to develop new bioderived materials but also lead to design various sensors based on biocircuits, biopolymers, bio-machines with an impact related to environmental, pharmaceutical, and biomedical studies. However, synthetic biology with nanotechnology-based biosensors recognizes biological elements that are still needed to reach a commercial success (Scognamiglio 2013). Several studies on bioreceptor molecules that can serve as alternative biosensor development have been dealt with studies on small molecule sensors resulted in detections of miRNA, gene expression, and viral RNAs (Hartig and Famulok 2008). Several redesigned aptazymes as riboswitches are known to be predictive models to designing biosensors (Chen and Ellington 2009; Ogawa 2011) was adopted for detecting alkaloids, other drugs like tetracycline, sulforhodamine B. Several studies related to biosensors based on ribozyme for the detection of glutamate (Okumoto et al. 2005), adenosine (Sun et al. 2010), hepatitis C virus (HCV) (Levesque et al. 2010), L-histidine (KonG et al. 2011) were reported.

Presently directed evolution and computational technology are being used to convert natural and artificial ribozymes to allosteric ribozymes coined as aptazymes through aptamer insertion which catalyze ribozyme activity when bound with ligand (Isaacs et al. 2006). While gene expression in yeast and bacteria using synthetic RNA is being engineered, it needs to be determined how to extend such systems to mammalian structures.

6.9 Conclusion

Biomimetic concept generation has been used by engineers and inventors throughout history. We have elaborated challenges and limitations of such systems by documenting an example each for function, design for remanufacture, therapeutics, and other aspects of biotechnology sector. The redesigned advanced systems production has indeed allowed us to bring important technological breakthroughs from pharmaceutical to biomedical diagnosis. The efficiency to improve dynamic profiling in biotechnology has enabled gene expression to accelarate research advances.

Acknowledgements The authors wish to thank Dr. Prashanth Suravajhala for providing the suggestions, encouragement, and fruitful discussion during preparation of this chapter.

Competing Interests None declared.

References

Ames BN, Durston WE, Yamasaki E, Lee FD (1973) Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. Proc Natl Acad Sci U S A 70(8):2281–2285

Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2. 2006.0028

Baeumner AJ, Schlesinger NA, Slutzki NS, Romano J, Lee EM, Montagna RA (2002) Biosensor for dengue virus detection: sensitive, rapid, and serotype specific. Anal Chem 74(6):1442–1448

Benner SA, Sismour AM (2005) Synthetic biology. Nat Genet 6:533–543

Bensaude-Vincent B (2009) Biomimetic chemistry and synthetic biology: a two-way traffic across the borders. Hyle Int J Philos Chem 15(1):31–46

Bhatia P, Chugh A (2013) Synthetic biology based biosensors and the emerging governance issues. Curr Synth Syst Biol 1(1):1–7

Buranachai C, Thavarungkul P, Kanatharana P (2012) A novel reconfigurable optical biosensor based on DNA aptamers and a DNA molecular beacon. J Fluoresc 22(6):1617–1625

Checa SK, Zurbriggen MD, Soncini FC (2012) Bacterial signaling systems as platforms for rational design of new generations of biosensors. Curr Opin Biotechnol 23:766–772

Chen X, Ellington AD (2009) Design principles for ligand-sensing, conformation-switching ribozyme. PLoS Comput Biol 5(12):e1000620

Chianella I, Piletsky SA, Tothill IE, Chen B, Turner APF (2003) MIP-based solid phase extraction cartridges combined with MIP-based sensors for the detection of microcystin-LR. Biosens Bioelectron 18:119–127

Choffnes ER, Relman DA, Pray L, IOM (Institute of Medicine) (2011) The science and applications of synthetic and systems biology. The National Academies Press, Washington, DC

David PC, Nanette JP (2016) Biotechnology, 2nd edn. Elsevier, Amsterdam

Dennis AM, Rhee WJ, Sotto D, Dublin SN, Bao G (2012) Quantum dot-fluorescent protein FRET probes for sensing intracellular pH. ACS Nano 6(4):2917–2924

Effendi L, Nielsen D, Solomon K, Prather KJ (2008) Engineering microbes with synthetic biology frameworks. Trends Biotechnol 26(12):674–681

Enander K, Choulier L, Olsson AL, Yushchenko DA, Kanmert D, Klymchenko AS, Demchenko AP, Mely Y, Altschuh D (2008) A peptide-based ratiometric biosensor construct for direct fluorescence detection of a protein analyte. Bioconjug Chem 19:1864–1870

Franck P, Paulasova P (2004) The peptide nucleic acids, efficient tools for molecular diagnosis. Int J Mol Med 13(4):521–525
Gómara MJ, Fernández L, Pérez T, Ercilla G, Haro I (2010) Assessment of synthetic chimeric multiple antigenic peptides for diagnosis of GB virus C infection. Anal Biochem 396:51–58
Gu X, Trybiło M, Ramsay S, Jensen M, Fulton R, Rosser S, Gilbert D (2010) Engineering a novel self-powering electrochemical biosensor. Syst Synth Biol 4:203–214
Hartig JS, Famulok M (2008) Screening of molecular interactions using reporter hammerhead ribozymes. Methods Mol Biol 429:251–263
Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. Nat Biotechnol 24:545–554
Jampasa S, Wonsawat W, Rodthongkum N, Siangproh W, Yanatatsaneejit P, Vilaivan CO (2014) Electrochemical detection of human papillomavirus DNA type 16 using a pyrrolidinyl peptide nucleic acid probe immobilized on screen-printed carbon electrodes. Biosens Bioelectron 54:428–434
Kanwar JR, Roy K, Maremanda NG, Subramanian K, Veedu RN, Bawa R, Kanwar RK (2015) Nucleic acid-based aptamers: applications, development and clinical trials. Curr Med Chem 22(21):2539–2557
Karimian N, Vagin M, Zavar MHA, Chamsaz M, Turner APF, Tiwari A (2013) Label-free biosensing: advanced materials, devices and applications. Biosens Bioelectron 50:492–498
Kellenberger CA, Chen C, Whiteley AT, Portnoy DA, Hammond MC (2015) RNA-based fluorescent biosensors for live cell imaging of second messenger cyclic di-AMP. J Am Chem Soc 137:6432–6435
Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Genet 11:367–379
Kindschy L, Alocilja E (2007) Development of a molecularly imprinted biomimetic electrode. Sensors 7(8):1630–1642
Kobayashi H, Kærn M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ (2004) Programmable cells: interfacing natural and engineered gene networks. PNAS 101(22):8414–8419
Kong RM, Zhang XB, Chen Z, Meng HM, Song ZL, Tan W, Yu RQ (2011) Unimolecular catalytic DNA biosensor for amplified detection of L-histidine via an enzymatic recycling cleavage strategy. Anal Chem 83(20):7603–7607
Kwakye S, Baeumner A (2003) A microfluidic biosensor based on nucleic acid sequence recognition. Anal Bioanal Chem 376:1062–1068
Lee GN, Na J (2013) Future of microbial polyesters. Microb Cell Fact 12:54
Lee J, Park IS, Jung E, Lee Y, Min DH (2014) Biosens Bioelectron 62:140–144
Lepora NF, Verschure P, Prescott TJ (2013) The state of the art in biomimetics. Bioinspir Biomim 8:1–11
Levesque MV, Levesque D, Briere FP, Perreault JP (2010) Investigating a new generation of ribozymes to target HCV. PLoS One 5(3):e9627
Liang R, Zhang R, Qin W (2009) Biomedical nanomaterials. Sens Act B Chem 141(2):544–550
Liu Y, Zhu L, Zhang Y, Tang H (2012) Sens Act B Chem 171:1151–1158
Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. Nat Biotechnol 27(12):1139–1150
Marner WD (2009) Practical application of synthetic biology principles. Biotechnol J 4:1406–1419
Marx S, Zaltsman A, Turyan I, Mandler D (2004) Parathion sensor based on molecularly imprinted sol−gel films. Anal Chem 76:120–126
Mascini M (2008) Analytical applications of aptamers. J Biol Chem 36(4):273–282
Michener K, 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
Neumann H, Neumann-Staubitz P (2010) Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. Appl Microbiol Biotechnol 87:75–86
Nicholls IA, Adbo K, Andersson HS, Andersson PO, Ankarloo J, Hedin-Dahlström J, Jokela P, Karlsson JP, Olofsson L, Rosengren J, Shoravi S, Svenson J, Wikman S (2001) Can we rationally design molecularly imprinted polymers? Anal Chim Acta 435:9–18

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

Okumoto S, Looger LL, Micheva KD, Reimer RJ, Smith SJ, Frommer WB (2005) Detection of glutamate release from neurons by genetically encoded surface-displayed FRET nanosensors. PNAS 102(24):8740–8745

Pardieu E, Cheap H, Vedrine C, Lazerges M, Lattach Y, Garnier F, Remita S, Pernelle C (2009) Molecularly imprinted conducting polymer based electrochemical sensor for detection of atrazine. Anal Chim Acta 649:236–245

Peeters M, Troost FJ, van Grinsven B, Horemans F, Alenus J, Murib MS,… Wagner P (2012) MIP-based biomimetic sensor for the electronic detection of serotonin in human blood plasma. Sensor and Actuat B-Chem 171:602–610

Pei R (2011) Aptamer-based biosensors through target-induced strand release. In: International Conference on Nanotechnology and Biosensor IPCBEE, 2. IACSIT Press, Singapore

Peter EN, Michael E (1999) An introduction to peptide nucleic acid. Curr Iss Mol Biol 1(2):89–104

Piletsky SA, Piletska EV, Bossi A, Karim K, Lowe P, Turner AP (2001a) Substitution of antibodies and receptors with molecularly imprinted polymers in enzyme-linked and fluorescent assays. Biosens Bioelectron 16(9–12):701–707

Piletsky SA, Alcock S, Turner AP (2001b) Molecular imprinting: at the edge of the third millennium. Trends Biotechnol 19(1):9–12

Prasuhn DE, Feltz A, Blanco-Canosa JB, Susumu K, Stewart MH, Mei BC, Yakovlev AV, Louko C, Mallet JM, Oheim M, Dawson PE, Medintz IL (2010) Quantum dot peptide biosensors for monitoring Caspase 3 proteolysis and calcium ions. ACS Nano 4(9):5487–5497

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

Quintero A, Garcia S, Guevara C, Rincon C, Ospina C, Guevara P, Cuero R (2007) A microbial biosensor device for iron detection under UV irradiation. IET Synth Biol 1(1–2):71–73

Ryoo SR, Lee J, Yeo J, Na HK, Kim YK, Jang H, Min DH (2013) Quantitative and multiplexed microRNA sensing in living cells based on peptide nucleic acid and nano graphene oxide (PANGO). ACS Nano 7(7):5882–5891

Saeidi N, Wong CK, Lo TM, Nguyen HX, Ling H, Leong SSJ, Poh CL, Chang MW (2011) Engineering microbes to sense and eradicate Pseudomonas aeruginosa, a human pathogen. Mol Syst Biol 7:521–532

Sarikaya M, Tamerler C, Jen AKY, Schulten K, Baneyx F (2003) Molecular biomimetics: nanotechnology through biology. Nat Mater 2:577–585

Scognamiglio V (2013) Nanotechnology in glucose monitoring: advances and challenges in the last 10 years. Biosens Bioelectron 47:12–25

Scognamiglio V, Stano P, Polticelli F, Antonacci A, Lambreva MD, Pochetti G, Giardi MT, Rea G (2013) Design and biophysical characterization of atrazine-sensing peptides mimicking the *Chlamydomonas reinhardtii* plastoquinone binding niche. Chem Phys 15:13108–13115

Sergeyeva TA, Slinchenko OA, Gorbach LA, Matyushov VF, Brovko OO, Pi-letsky SA, Sergeeva LM, Elska GV (2010) Catalytic molecularly imprinted polymer membranes: development of the biomimetic sensor for phenols detection. Anal. Chim. Acta 659:274–279

Sismour AM, Benner SA (2005) Synthetic biology. Expert Opin Biol Ther 5(11):1409–1414

Siuti P, Yazbek J, Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. Nat Biotechnol 31(5):448–453

Subrahmanyam S, Piletsky SA, Piletska EV, Chen B, Karim K, Turner APF (2001) 'Bite-and-Switch' approach using computationally designed molecularly imprinted polymers for sensing of creatinine. Biosens Bioelectron 16:631–637

Sun C, Liu X, Feng K, Jiang J, Shen G, Yu R (2010) An aptazyme-based electrochemical biosensor for the detection of adenosine. Anal Chim Acta 669:87–93

Suravajhala P, Burri HVR, Heiskanen A (2014) Combining aptamers and in silico interaction studies to decipher the function of hypothetical proteins. Eur Chem Bull 3(8):809–810

Teo WS, Chang MW (2015) Bacterial XylRs and synthetic promoters function as genetically encoded xylose biosensors in *Saccharomyces cerevisiae*. Biotechnol J 10:315–322

Tombelli S, Minunni M, Mascini M (2007) Aptamers-based assays for diagnostics, environmental and food analysis. Biomol Eng 24:191–200

Valero-Navarro A, Salinas-Castillo A, Fernández-Sánchez JF, Segura-Carretero A, Mallavia R, Fernández-Gutiérrez A (2009) The development of a MIP-optosensor for the detection of monoamine naphthalenes in drinking water. Biosens Bioelectron 24:2305–2311

Villiers M-B, Cortay J-C, Cortès S, Bloquel B, Brichler S, Brakha C, Kay A, Falah N, Zoulim F, Marquette C, Marche PN, Dény P (2015) Protein-peptide arrays for detection of specific anti-hepatitis D virus (HDV) genotype 1, 6, and 8 antibodies among HDV-infected patients by surface plasmon resonance imaging. J Clin Microbiol 53(4):1164–1171

Ye L, Haupt K (2004) Molecularly imprinted polymers as antibody and receptor mimics for assays, sensors and drug discovery. Anal Bioanal Chem 378(8):1887–1897

Zhao J, Jedlicka SS, Lannu JD, Bhunia AK, Rickus JL (2006) Liposome-doped nanocomposites as artificial-cell-based biosensors: detection of listeriolysin. Biotechnol Prog 22(1):32–37

Design Principles of Synthetic Biological Oscillators

7

Ashish Panghalia and Vikram Singh

Abstract

Oscillations, having indispensible roles in the dynamics of various cellular and subcellular processes, are ubiquitous in biological systems. Construction of novel genetic oscillators is one of the key aspects of synthetic biology research. In this chapter, we have examined the underlying design principles of a variety of synthetic oscillators. We have first reviewed the oscillations in natural systems and then specified the basic characteristics of genetic oscillators. Computational methods used for the in silico studies of biological oscillators are described, and a list of software and tools being used is given. We have further discussed about the ten synthetic oscillators and presented a brief summary of the computational and experimental strategies involved in their realizations. This chapter is concluded by providing the hands-on exercises for numerical simulations of two synthetic oscillators.

Keywords

Genetic oscillators · Repressilator · Feedback · Deterministic simulation · Stochastic simulation · Synthetic biology

7.1 Introduction

Synthetic biology is a rapidly expanding field of research in biological sciences that aims to establish engineering rules for the forward synthesis of cellular functions (Purcell et al. 2010). Construction of the new systems, devices, and biological

A. Panghalia · V. Singh (✉)
Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Dharamshala, Himachal Pradesh, India
e-mail: vikramsingh@cuhimachal.ac.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_7

parts and redesigning of an existing biological system are the main thrust areas of synthetic biology. The credit of the introduction of synthetic biology research field is given to the two landmark publications: (i) genetic toggle switch (Gardner et al. 2000) and (ii) repressilator (Elowitz and Leibler 2000) that are the classic examples of artificial gene regulatory circuits, published in the beginning of the year 2000. Synthetic biology includes the combination of in silico and in vitro and sometimes in vivo analyses integrating the concepts of biology, mathematics, and engineering toward empowering the development of novel approaches and systematic frameworks for exploring the phenotypic behavior of a given regulatory network (Stricker et al. 2008).

Oscillations, i.e. the periodic waveform representing quantitative changes in any cyclic process (Winfree 2001), are one of the essential characteristics of all life forms and are ubiquitous in living systems. Most common routines of our life, e.g. periodic heartbeat and breathing, follow oscillatory behavior. Oscillations are observed at every time scale in biological systems, such as calcium dynamics in seconds, glycolytic oscillations in minutes, circadian clock in hours, and hormonal oscillations in days and months (Maini 1996). These oscillations are derived from several regulatory mechanisms that control the variety of dynamics occurring in living systems. The interval between two reference points in a wave signifies the time period of oscillation, which constitutes its most important feature (Winfree 2001). From the most primordial bacteria to the advanced life forms, rhythmic regulatory systems play a critical role in the emergence, evolution, and existence of life on earth. In order to capture these oscillations and understanding the science behind them, it was the need of era to develop theoretical models of these oscillatory networks. Synthetic biology aims to design the natural systems that can mimic the biological processes in order to get quantitative as well as qualitative understanding and deeper insights about the rate-limiting steps of various regulatory processes. In this chapter, we discuss biological oscillators and their existence in various forms of nature, computational and molecular approaches being employed for designing and studying these oscillations, tools and software available for exploring this concept, and the recent updates on the current state of research work in this field of science.

7.2 Oscillations in Nature

Oscillations are distributed throughout nature in one or the other form. They can exist on years' scale in the form of galactic motion or on microscopic scale as neutrino movement inside an element (Shaviv et al. 2014; Maltoni et al. 2004). The cycle of seasonal variation sheds the rays of periodicity in natural environment. Importance of these oscillations can be related to the very fundamental phenomenon of our daily life, day, and dawn (Kruse and Jülicher 2005). The cycle of day and night gives the hint that periodicity is an essential element of our life. The formation of a complete multicellular body from a single-cell "zygote" is a result of multiple cell division cycles. Heart, a vital organ in mammals, also follows rhythmic behavior generating a particular pattern of oscillations while supplying blood to various

Table 7.1 Natural biological oscillators

Biological rhythm	Approximate time period
Neural oscillations	Millisecond
Cardiac rhythms	Second
Calcium and glycolytic oscillations	Minute
Mitotic and hormonal oscillations	Hour
Circadian oscillations	Day
Menstruation cycles	Month
Circannual cycle	Years

body parts. Several regulatory mechanisms of the body, like circadian rhythms and menstrual cycle, are also common examples of biological oscillations (Karpman and Sadovskaya 1964). The external stimuli are also important in generating frequencies in some of the biological processes. The membrane ion current is an example of such biological frequencies that occur via resonance of stereocilia present over the hair cell of inner ear (Fettiplace 2017). A variety of natural oscillators with their time scale are listed in Table 7.1.

7.3 Underlying Principles of Genetic Oscillations

Oscillation can be defined as a repetitive variation of a measure with respect to time, between two different states. Oscillations are the cyclic changes in a measurable quantity that exhibit a waveform of a relatively constant period (Winfree 2001). Oscillations are ubiquitous in nature; e.g., nervous system, heartbeat, respiration, circadian rhythms, and menstruation are all following a cyclic pattern (Cardon and Iberall 1970). In many of the cellular processes, oscillations play a vital role; such as axonemes of cilia in a cytoskeleton structure, oscillations in the concentration of Min protein play an important role in site selection for the division in some bacteria, and circadian clocks also find their basis on the genetic oscillators. The dynamic nature of cellular interacting components and their collective behavior to achieve self-organization can lead to the emergence of oscillations (Kruse and Jülicher 2005). The temporal or spatial oscillations are majorly controlled via five elements, i.e. (i) time delay, (ii) nonlinearity in regulation, (iii) negative feedback, (iv) positive feedback, and (v) noise (Cao et al. 2016).

(i) Time delays are explained as the amount of time associated between the sensing of some disturbance and the production of a response against it. These time delays for different processes vary in a wide range. For example, various hormonal and neural activities control the rate of heartbeat and blood pressure. Every activity has an effective delay of time for response (Glass et al. 1988).
(ii) The living systems are nonlinear in nature, which means there is no equivalent response to the external and individual inputs (Janson 2012). Nonlinearity

in the dynamic systems gives rise to the spontaneous oscillations, which are also known as "active oscillations." Most of the biological systems follow nonlinearity as they allow a continuous exchange of energy between the system and its surrounding environment, thus falling under this category. Biological systems generally exhibit spontaneous oscillations (Janson 2012).

(iii) Negative feedback is observed when the output of a reaction affects its own progress in such a way so as to reduce the output. Although these negative feedback circuits are the fundamental requirement for generating oscillations, in many cases these are unable to produce oscillations on their own (Cao et al. 2016). In order to generate persistent oscillations, negative feedback loops may have to be combined with time delays or to be coupled with positive feedback loops (Cao et al. 2016).

(iv) The positive feedback circuit is simply a closed circuit in which output enhances the input value and vice versa (Zeigler et al. 2019) and thereby causing an increase in perturbations as a result of small disturbance on a circuit (Zeigler et al. 2019). These positive feedback loops in combination with negative feedback can generate several dynamical behaviors, like excitable pulses and bistability (Cao et al. 2016).

(v) Noise can be simply defined as random fluctuations (Tsimring 2014) and can affect a circuit in both, positive and negative, ways. An independent negative feedback loop is more prone to noise compared to an integrated negative–positive feedback loop (Bratsun et al. 2005). Sometimes, noise is also considered as a source of oscillations when any active feedback loop is missing (Simakov and Pérez-Mercader 2013). These elements play a vital role in modulating the oscillatory dynamics of a biological circuit, as they can enhance or damp the oscillations based on network architecture.

7.4 Computational Approaches to Study Biological Oscillations

In order to develop in silico models for studying the biological oscillations, the underlying biological reactions have to be developed in the form of differential equations that in turn have to be simulated to observe the emerging dynamical behavior (Singh 2015). Mathematical models can mimic the biological systems, under a given set of conditions for a defined time period. There are majorly two approaches to solve these mathematical models, i.e., deterministic and stochastic approaches. In deterministic method, the outcome of the models is completely dependent on the initial conditions and the given parameter values. Deterministic models mostly solve differential equations without introducing any random fluctuations. The output of a deterministic system is fixed if it is provided with particular input values, irrespective of the number of times it is iterated. Thus, deterministic methods model the changes in the state of a system with time as a continuous process, which can be predicted by solving coupled ODEs. In case of stochastic models, randomness is introduced, which tries to give a tint of natural systems in

mathematical models (Chandran et al. 2008). Also, deterministic models are found to have an increase in error rate with reduction in rate of reaction or lowering the protein concentration. Natural systems, generally, operate under the influence of noise resulting from random events. Therefore, in order to monitor the effect of noise on mathematical models describing a biological system, randomness is to be incorporated and such methods are known as stochastic methods (Hahl and Kremling 2016). These stochastic simulation methods, also known as Monte Carlo methods, are probability-based models. The output values of stochastic models may vary with number of iterations irrespective of same input for all simulation runs. These stochastic models consider uncertainty as an important factor in their calculations. The deterministic approach is usually taken into consideration when we expect a determined output for a specific input, but in case we have to study a system with inherent noise, most preferred method is stochastic modeling. In the following, we briefly describe the RK4 method and Gillespie's algorithms that are the two well-studied methods for deterministic and stochastic approach to solving mathematical models, respectively.

7.4.1 Runge–Kutta Method

This method is one of the well-studied methods of numerical simulations given by two German mathematicians, Carl Runge and Wilhelm Kutta, in the year 1900. The fourth order of the Runge–Kutta method also known as RK4 is generally considered as a simple and reliable numerical solution for differential equations by introducing a definite time step for a given time period. It involves the calculation of four values of slopes against the one slope calculation strategy of Euler method for solving ordinary differential equations over a range of time using a fixed value of time step. In this method, the ordinary differential equations (ODEs) are integrated by a trial time step that helps in reducing the error terms of lower order. For solving differential equation using RK4, we need to write the reaction rates in the form of ODEs for each species y under consideration with the known initial values of y, i.e. y_0.

First-order ODEs can be constructed by further assuming the M chemical reactions, in which the N chemical species are interacting, as continuous processes.

$$\frac{dy_1}{dt} = f_1\left(y_1, \ldots y_n\right)$$

$$\frac{dy_2}{dt} = f_2\left(y_1, \ldots y_n\right)$$

$$\frac{dy_n}{dt} = f_n\left(y_1, \ldots y_n\right)$$

These equations are known as coupled rate-reaction equations, where f_i determines the rate constant of the ith chemical reaction.

In the following, we provide the basic RK4 algorithm for a single rate equation that may be extended for the coupled rate equations. Suppose, for a particular species y, $\frac{dy}{dt} = f(t, y)$ with value of y (at time t_0) $= y_0$ is known. This method is used to calculate the approximate value of y at given time point t, by calculating four values of slopes, as following.

$$k_1 = h * f(t_n, y_n)$$

$$k_2 = h * f\left(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)$$

$$k_3 = h * f\left(t_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right)$$

$$k_4 = h * f(t_n + h, y_n + k_3)$$

where $n = 0, 1, 2, 3, \ldots \left(\frac{t - t_0}{h}\right)$ and h is the step size.

$$y_{n+1} = y_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O\left(h^5\right)$$

This calculates the value of y at next time step (y_{n+1}) from the value of y at the previous one (y_n). k_1 gives the value of increment on the basis of value of the function at the beginning, i.e. t_n. k_2 gives the value of increment on the basis of slope k_1 at the midpoint using $t_n + \frac{h}{2}$. k_3 gives the value of increment on the basis of slope k_2 at the midpoint using $t_n + \frac{h}{2}$. k_4 gives the value of increment on the basis of slope k_3 in the end using $t_n + h$.

7.4.2 Gillespie's Algorithm (Stochastic Method)

Biological processes involve the dynamics of large number of molecular events with intrinsic noise. The dynamics of such processes involve concentration of all molecules involved at nanomolar concentrations. Thus, a probabilistic model of such systems gives a closer and accurate insight about dynamics of molecular systems. D. T. Gillespie in 1977 proposed a stochastic method for simulating the effects of inherent intrinsic noise in the dynamics of systems with low molecular concentrations (Gillespie 1977). Unlike deterministic method, this algorithm is based on a single ordinary differential equation known as "Master equation" and is proposed as a solution of the "Master equation." This algorithm is derived on the basis of Monte Carlo approach to solve the dynamics of a system mathematically over a certain time period, but the time increment is not defined by definite steps.

This method attempts to solve the problem in which a fixed volume V contains the mixture of N number of different chemical species that are interacting with each other with the M number of different chemical reactions. The molecule number of each species at an initial time is given as X_i, and we have to find the population of this species at a given later time. To simulate the set of reactions that were represented as the coupled differential equations in the previous section, Gillespie's method provides the answers of two basic questions: (i) When the next reaction will occur, and (ii) Which reaction will it be? The basic methodology of the Gillespie's algorithm is described in the following schema.

- Enter the initial value of $X_i(t_0)$ for each X_i where $i = 1, 2, 3,..., N$ and $t_0 = 0$.
- For each reaction, enter the values of reaction probabilities c_ν, where $(\nu = 1, 2, .., M)$.
- For each reaction, define the molecular combination functions h_ν, where $(\nu = 1, 2, .., M)$.
- Compute the values of reaction propensities $a_\nu = h_\nu c_\nu$ where $(\nu = 1, 2, \ldots, M)$.
- Also calculate $a_0 = \sum_{\nu=1}^{M} a_\nu$.
- Using a uniform random number generator, get two random numbers r_1 and r_2.
- Define $\tau = \left(\frac{1}{a_0}\right) \ln \left(\frac{1}{r1}\right)$.
- Define an integer μ such that, $\sum_{\nu=1}^{\mu-1} a_\nu < r2a_0 \leq \sum_{\nu=1}^{\mu} a_\nu$.
- Update all the X_i according to the μ.
- Set $t = t + \tau$ and $n = n + 1$.

Value of τ gives the time interval after which a reaction will occur and μ indicates which reaction it will be. The process is to be iterated till the final time up to which simulation is to be performed.

7.5 Tools and Software to Simulate Biological Oscillations

A large number of biological systems have been studied by designing their mathematical models that can mimic the underlying dynamics of the natural system under study. Synthetic biological oscillators are, in general, designed first using a mathematical modeling approach, and the obtained results are then validated using in vitro as well as in vivo experiments. These mathematical models are composed of reaction rate equations based on the biochemical processes responsible for the emerging dynamics in the biological system under study. These equations can be solved stochastically or by converting them into ordinary differential equations (ODEs). There are various tools and software available, provided by paid and free repositories, which can be used to simulate the dynamics of these mathematical models of biological systems and to further analyze their biological significance. Few of these software and tools are listed in Table 7.2.

Table 7.2 Various tools and software for generating and analyzing the mathematical models

Name	Availability	Weblink
ASCEND	Free	http://ascend4.org/Main_Page
GNU Octave	Free	https://www.gnu.org/software/octave/
Scilab	Free	https://www.scilab.org/
Wolfram Mathematica	Free	http://www.wolfram.com/mathematica/
Simmer	Free	https://arxiv.org/abs/1705.09746
COPASI	Free	http://copasi.org/
SAGE	Paid	http://www.sageofathens.com/
XPPAUT	Free	http://www.math.pitt.edu/~bard/xpp/xpp.html
MATLAB	Paid	https://www.mathworks.com/products/matlab.html
CellDesigner	Free	http://www.celldesigner.org/
TinkerCell	Free	http://www.tinkercell.com/

7.6 A Brief Overview of the Various Synthetic Biological Oscillators

Repressilator is the first-ever-designed synthetic biological system in which oscillations were observed in the constituent protein concentrations (Elowitz and Leibler 2000). Before the realization of repressilator, there were few theoretical studies examining the onset of oscillations in biological systems that provided the basis for understanding the underlying basic concepts for successful realization of a synthetic oscillator, such as synthetic genetic oscillator by Goodwin in 1965 (Goodwin 1965), genetic feedback repressor solved using Boolean modeling by Fraser in 1974 (Fraser and Tiwari 1974), molecular basis of periodic cellular rhythms by Goldbeter in 1997 (Goldbeter 1997), and two-gene oscillator by Smolen in 1998 (Smolen et al. 1998). In approximately 20 years of post repressilator synthetic biology research, significant work is done in the development and understanding of synthetic biological oscillators. In the following, we provide a brief summary of ten synthetic oscillators.

7.6.1 Goodwin Oscillator

Goodwin oscillator (Goodwin 1965) is known as the first synthetic genetic oscillator that was proposed with a theoretical model and studied computationally over 50 years ago. In this model, there is only one gene that represses the expression of itself as shown in Fig. 7.1a.

L_i represents the gene locus transcribing mRNA (denoted by X_i) that in combination with the ribosome forms polysome (R) that helps in the production of its proteins (Y_i). The protein formed is assumed as an enzyme that helps in the metabolic transformation and gives rise to the metabolic species M_i that in turn

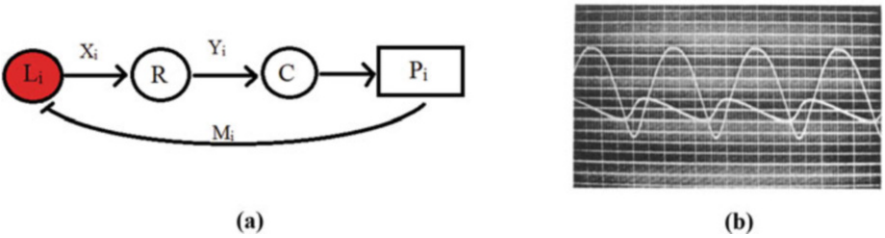


Fig. 7.1 (**a**) Topology of single-gene oscillator, which represses itself. (**b**) Simulation results of the Goodwin oscillator. Figure (**b**) reproduced with permission from Goodwin (1965) © (1965) Elsevier Ltd

lowers down the expression of gene (L_i) by repressing it. Following equations give the dynamics of this system.

$$\frac{dX_i}{dt} = \frac{a_i}{A_i + k_i Y_i} - b_i$$

$$\frac{dY_i}{dt} = \alpha_i X_i - \beta_i$$

Here, X_i represents the concentration of mRNA and Y_i represents the concentration of protein. The oscillations produced from the numerical simulations of these equations are shown in Fig. 7.1b.

7.6.2 Smolen's Oscillator

Smolen's oscillator (Smolen et al. 1998) is a simple kinetic model that examines the dynamic activity of the gene regulatory systems including autoregulation, phosphorylation, and dimerization of transcription factors along with noise and feedback. There are two genes in this oscillator. The first gene (A) promotes the transcription of second gene (B) and also its own transcription, and the second gene represses the transcription of first gene as well as its own transcription as shown in Fig. 7.2a.

In this model, signal-transduction pathway was considered in which stimuli leads in the formation of the homodimer and the phosphorylation of TFs, those bind with the specific responsive-element DNA sequences (TF-REs). Dimer of phosphorylated TF-A protein binds to TF-RE and activates the transcription of *tf-a* and *tf-r* genes which forms the TF-A and TF-R proteins. The rate constants for degradation of TF-A and TF-R are represented by $k_{1,d}$, and $k_{2,d}$ respectively. TF-R protein represses the transcription of the *tf-r* gene by binding to TF-RE, whereas it represses the transcription of the *tf-r* gene by competitively inhibiting the binding of TF-A to TF-RE as shown in Fig. 7.2b.

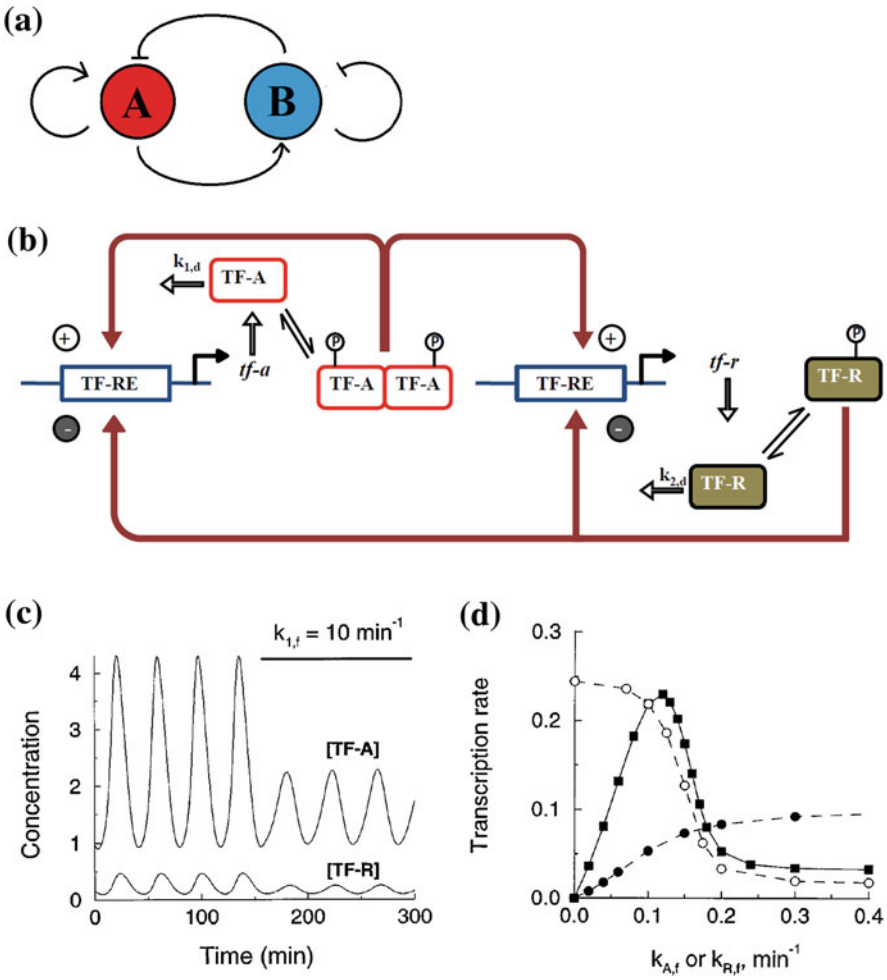


Fig. 7.2 (**a**) Topology of Smolen's oscillator. (**b**) Schematic representation of the Smolen's oscillator. (**c**) Oscillations obtained by the numerical simulations of the Smolen's oscillator. (**d**) Oscillations obtained in the experimental conditions. Figures (**b**), (**c**), (**d**) are adopted from Smolen et al. (1998)

7.6.2.1 ODEs for the In Silico Implementation of Smolen's Oscillator

Here, $K_{R,d}$ represents the dissociation constant of TF-R from TF-REs. $k_{1,f}$, and $k_{2,f}$ are the maximal synthesis rates. $k_{1,d}$ and $k_{2,d}$ are the rate constants of degradation, and $K_{2,d}$ represents the dissociation constant of TF-A from TF-REs.

$$\frac{d[TF-A]}{dt} = \frac{k_{1,f}[TF-A]^2}{[TF-A]^2+K_{1,d}\left(1+[TF-R]/K_{R,d}\right)} - k_{1,d}\,[TF-A] + r_{1,\text{bas}}$$

$$\frac{d\left[TF-R\right]}{dt}=\frac{k_{2,f}[TF-A]^2}{[TF-A]^2+K_{2,d}\left(1+[TF-R]/K_{R,d}\right)}-k_{2,d}\left[TF-R\right]$$

$r_{1,\text{bas}}$ gives the basal rate due to which an activator is synthesized at a negligible concentration of dimer. $k_{1,d}$ gives the rate constant of the first-order degradation of TF-A. Oscillations were readily generated by this model as shown in Fig. 7.2c.

7.6.3 Repressilator

The theoretical framework of repressilator was developed in 1974, using discrete Boolean simulations as an extension of the Goodwin's oscillator (Fraser and Tiwari 1974). It may be explained as a regulatory network of three genes, in which each gene represses the expression of its successor gene in a cyclic manner (Elowitz and Leibler 2000). LacI protein from *Escherichia coli* inhibits the transcription of the TetR gene. The expression of cI gene from λ phage is inhibited by the second repressor protein named TetR of Tn10 transposon. Finally, cI which is the third repressor protein inhibits the transcription of LacI gene and completes the cycle shown in Fig. 7.3a, b.

7.6.3.1 ODEs for the In Silico Implementation of Repressilator

Following six-coupled first-order differential equations determine the dynamics of this system. Here, p_i represents the concentration of repressor proteins and m_i represents the concentration of their corresponding mRNAs. Each of these is considered as the continuous dynamic variables (i is LacI, TetR, and cI, respectively).

$$\frac{dm_i}{dt}=-m_i+\frac{\alpha}{\left(1+p_j^n\right)}+\alpha_0$$

$$\frac{dp_i}{dt}=-\beta\left(p_i-m_i\right)\quad\begin{pmatrix}i=l<\text{cI, tetR, cI}\\ j=\text{cI, lacI, tetR}\end{pmatrix}$$

Where, α_0 gives the protein copy number produced by a given promoter type in each cell when the saturated amount of repressor is present and $\alpha+\alpha_0$ is the protein copies when the repressor is absent. The ratio of decay rates of protein to mRNA is denoted by β and n is the Hill coefficient. Using these six coupled differential equations corresponding to three mRNAs and their repressor proteins, oscillations were obtained by in silico ODEs simulation as shown in Fig. 7.3c that were also successfully reproduced in the in vitro experiments shown in Fig. 7.3d.

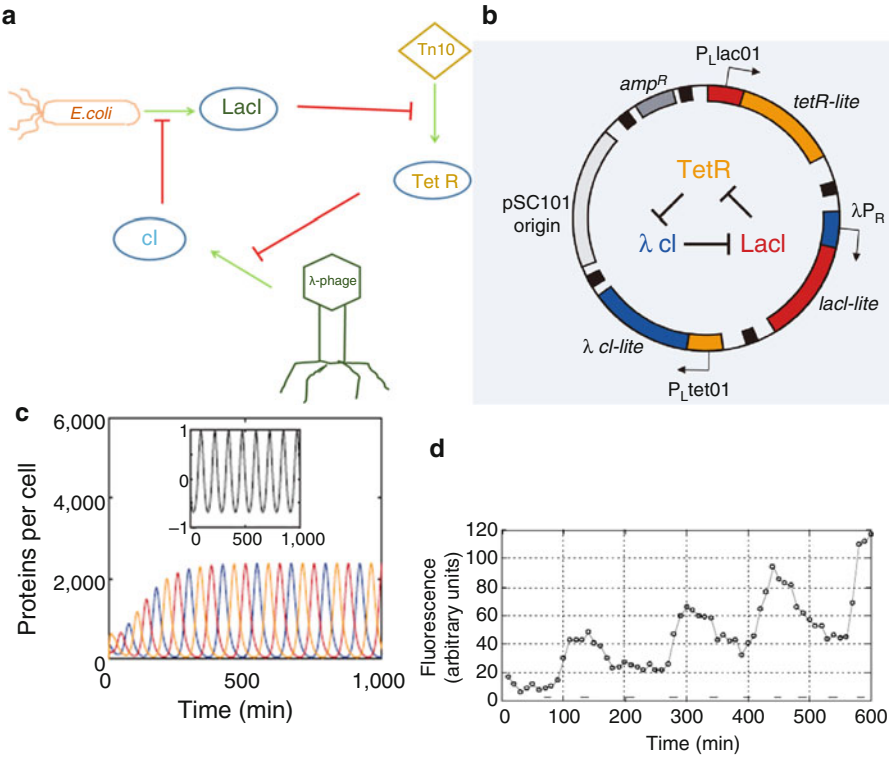


Fig. 7.3 (**a**) Topology of repressilator in which the every gene represses the transcription of its successor. (**b**) Repressilator network. (**c**) Simulations of ODEs of repressilator genes. (**d**) Oscillation shown in living bacteria. Figures (**b**), (**c**), (**d**) are reproduced with permission from (Elowitz and Leibler 2000) © (2000) Macmillan Magazines Ltd

7.6.4 Genetic Relaxation Oscillator

The topology of this oscillator consists of various nodes that were linked with the activating and repressing links Chaos (Hasty et al. 2001). The transcription of gene A and gene B is upregulated by gene A at the lower concentration of its protein. The transcription of the first gene (A) and second gene (B) is regulated by gene A via variable promoter, and the transcription of gene A is repressed by gene B via degradation as shown in Fig. 7.4a.

There are two plasmids which are used to control the repressor (X) and RcsA (Y) production by using the P_{RM} promoter as shown in Fig. 7.4b. Both the plasmids are turned on by the activity of repressor at P_{RM}. The concentration of RcsA starts increasing as the promoter gets activated, which leads to the reduction of the concentration of repressor.

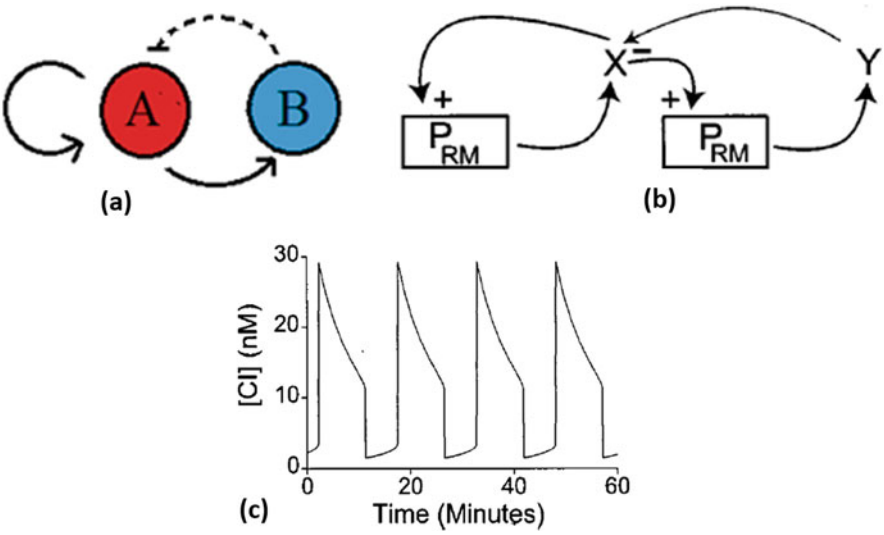


Fig. 7.4 (**a**) Topology of amplified variable link oscillator. (**b**) Working of amplified variable link oscillator. (**c**) Simulation generated for amplified variable link oscillator. Figures (**b**), (**c**) are reproduced with permission from Hasty et al. (2001) © (2001) American Institute of Physics

7.6.4.1 ODEs for the In Silico Implementation of Genetic Relaxation Oscillator

In the given equations, x and y represent the quantity and m_x and m_y are the copy numbers of the plasmid of cI and RcsA, respectively.

$$\frac{dx}{dt} = m_x f(x) - \gamma_x x - \gamma_{xy} xy$$

$$\frac{dy}{dt} = m_y f(x) - \widetilde{a}_y y$$

γ_x and γ_y represent the degradation of cI and RcsA and γ_{xy} shows the degradation rate of cI by RcsA. Simulations of the ODEs give the oscillations as shown in Fig. 7.4c.

7.6.5 Amplified Negative Feedback Oscillator

The previously studied oscillators, like the Goodwin oscillator and repressilator, were based on the negative feedback cycle only. The amplified negative feedback oscillator (Atkinson et al. 2003) was proposed to showcase the usage of positive feedback loops in the expression of oscillations. It comprises two genes (A and B): one gene (A) activates the self-transcription via positive feedback and also promotes

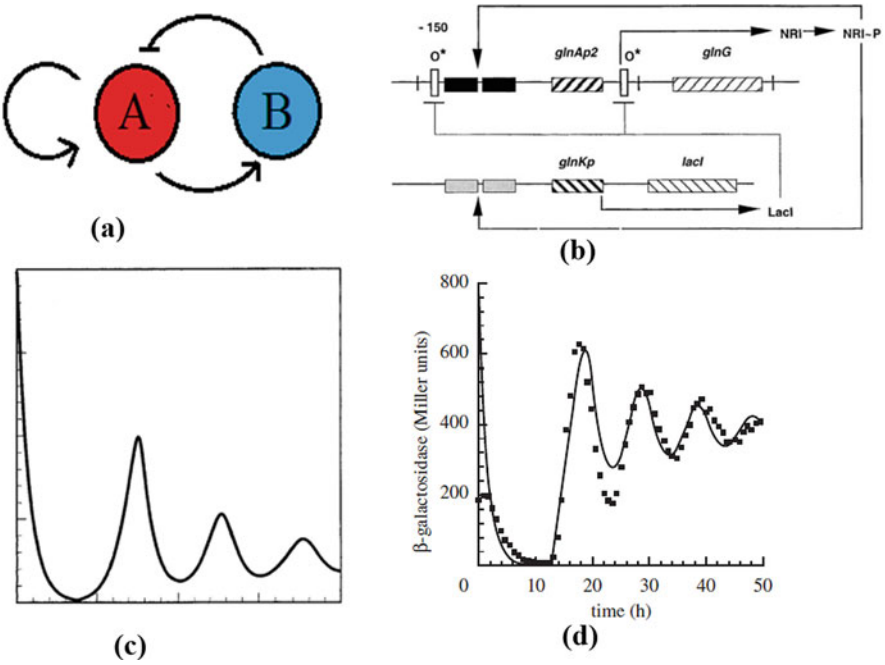


Fig. 7.5 (**a**) Topology of amplified negative feedback oscillator. (**b**) Schematic working of amplified negative feedback oscillator. (**c**) Simulation generated for amplified negative feedback oscillator. (**d**) Dynamics obtained in the experimental conditions. Figures (**b**), (**c**), (**d**) are reproduced with permission from Atkinson et al. (2003) © (2003) Cell Press

the transcription of the second gene (B). Simultaneously, the transcription of the first gene is repressed by the second gene as shown in Fig. 7.5a.

7.6.5.1 ODEs for the In Silico Implementation of Amplified Negative Feedback Oscillator

The concentrations of various components of a network are represented by x_i, and the concentrations of proteins and mRNAs are represented by the even and odd-numbered variables, respectively.

$$\frac{dx_{(2k-1)}}{dt} = \beta_{(2k-1)}\left(f_{(2k-1)} - x_{(2k-1)}\right)$$

$$\frac{dx_{(2k)}}{dt} = \beta_{2k}\left(x_{2k-1} - x_{2k}\right) \text{ for } k = 1, 2, \tag{3}$$

Here, f_i (transcriptional rate of three phases) and β_i (numerous rates which describe constants) are described by the two equations. There are two modules in this network: one is NRI that encodes the activator and second is LacI that encodes the repressor. Encoding NRI is fused with glnG to the control region on the promoter of glnA as shown in Fig. 7.5b to construct the module which acts as an activator to the oscillator. The regulation of promoter was done in the presence of two phosphorylated NRI-binding sites (NRIp) which are upstream to the promoter and form enhancer, and two LacI operators one of which is present downstream of promoter and other is upstream of enhancer. The oscillations are obtained by in silico ODEs simulation as shown in Fig. 7.5c.

7.6.6 Metabolator

Metabolator (Fung et al. 2005) is reported as the first oscillator in which along with the genes, metabolites also constitute the main components of its circuitry. It consists of two genes (A and B), gene B produces an enzyme which can covert one metabolic pool named M2 to another pool named M1 and the transcription of this gene B is also activated by M2. At the same time, another enzyme is produced by the gene A which enables the conversion of the M1 pool to the M2 pool, and the M2 represses the transcription of this gene A. Moreover, there is an influx and efflux from the M1 and M2 pools, respectively, as shown in Fig. 7.6c. The topology of this oscillator (metabolator) is schematically represented in Fig. 7.6b as a gene regulatory network.

Gene A and B repress themselves by activating the M2 and also activate each other as shown in Fig. 7.6b. This concept was implemented by modeling with ODEs and Chemical Langevin Equations (CLEs) which represent a GFP reporter (Fung et al. 2005). Due to an increase and decrease in the M2 pool, the two genes repress their activity as a result of which these two genes regulate the activity of each other.

7.6.6.1 ODEs for the In Silico Implementation of Metabolator

$$\frac{d\ AcCoA}{dt} = V_{\text{Acs}} - V_{\text{Pta}} - V_{\text{gly}} - V_{\text{TCA}}$$

$$\frac{d\ AcP}{dt} = V_{\text{Pta}} - V_{\text{Ack}}$$

$$\frac{d\ OAc^{-}}{dt} = V_{\text{Ack}} - V_{\text{AcE}} - V_{\text{Acs}}$$

$$\frac{d\ HOAc}{dt} = V_{\text{AcE}} - V_{\text{out}}$$

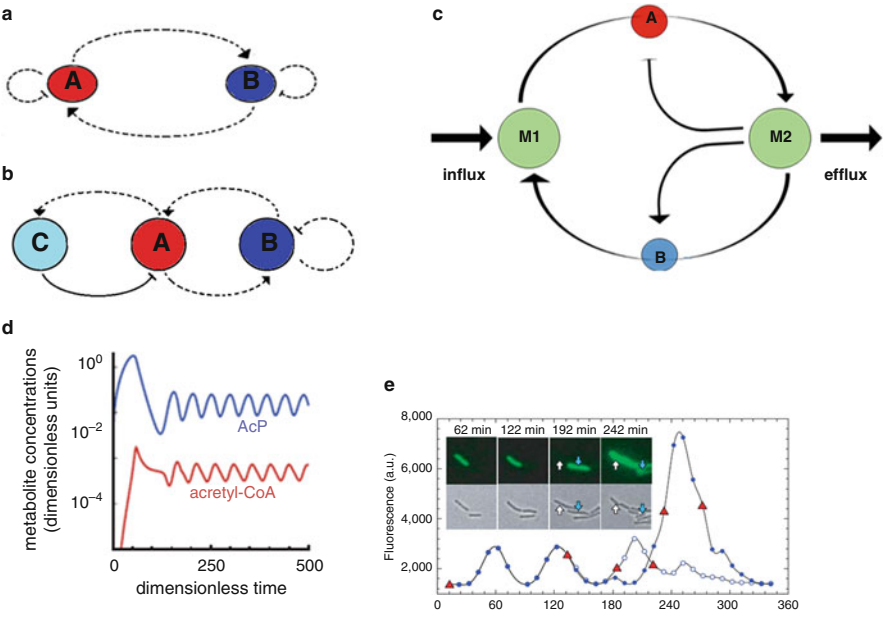


Fig. 7.6 (**a**) Schematic GRN diagram and (**b**) topology of metabolator. (**c**) Conceptual view of metabolator. (**d**) In silico simulations in high glycolytic flux. (**e**) Fluorescence trajectory of single cell. Figures (**d**), (**e**) are reproduced with permission from Fung et al. (2005) © (2005) Nature Publishing Group

The concentration of three key proteins was determined by:

$$\frac{d\ \text{LacI}}{dt} = R_{\text{LacI}} - R_{d,\text{LacI}}$$

$$\frac{d\ \text{Pta}}{dt} = R_{\text{Pta}} - R_{d,\text{Pta}}$$

$$\frac{d\ \text{Acs}}{dt} = R_{\text{Acs}} - R_{d,\text{Acs}}$$

The mathematical analysis showed that the oscillations as shown in Fig. 7.6d were produced through a Hopf bifurcation. Here, V_i represents the expression of rate of reaction of Glycolytic flux (V_{gly}), Acs flux (V_{Acs}), Pta flux (V_{Pta}), flux to TCA (V_{TCA}), flux for the AcP and OAc^- reaction (V_{Ack}), equilibrium of acid base for acetic acid (V_{AcE}), HOAc transport rate (V_{out}), and R_{LacI}, R_{Pta}, R_{Acs}, represent the synthesis rate of LacI, Pta, and Acs, respectively. $R_{d,}$= (X = LacI, Pta, and Acs) represents the degradation rate of LacI, Pta, and Acs. It was observed that for the presence of oscillations, a high inflow rate is required which can be abolished by the high concentration of M2. The number of relative gene copies has also affected the

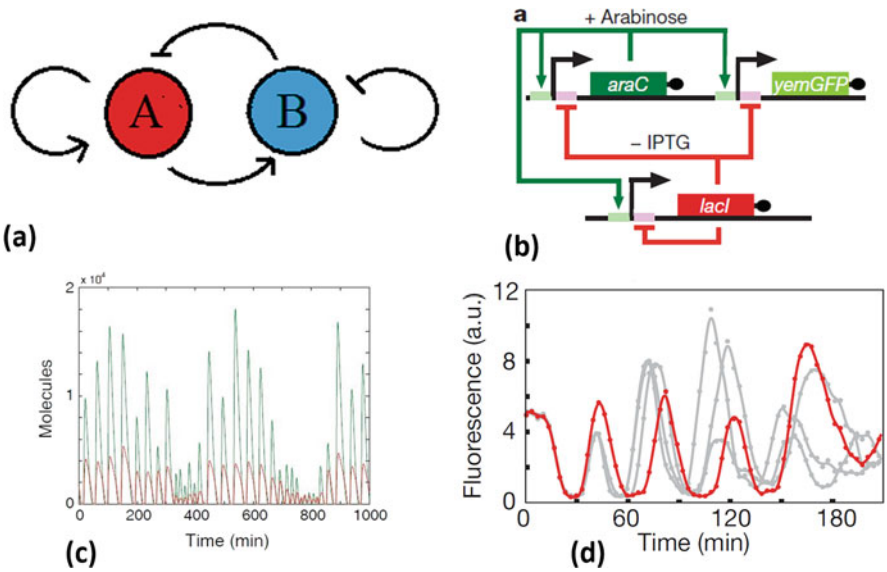


Fig. 7.7 (**a**) Topology of Hasty's two-gene oscillator. (**b**) Schematic working of Hasty's two-gene oscillator. (**c**) Bistable oscillations generated for Hasty's two-gene oscillator. (**d**) Experimental fluorescence trajectories of single cell. Figures (**b**), (**c**), (**d**) are reproduced with permission from Stricker et al. (2008) © (2008) Macmillan Publishers Limited

production of oscillations, while the amplitude variation created by the addition of noise is found by the CLE simulations.

7.6.7 Two Genes Based Oscillator

This oscillator (Stricker et al. 2008) contains two genes as shown in Fig. 7.7a. Gene A, which is the first gene, promotes itself and also of the second gene. Gene B, the second gene, represses itself and also the transcription of gene A. This tendency of self-repression in gene B is the major factor which makes the difference between the topology of Hasty's two-gene oscillator and amplified negative feedback oscillator.

E. coli components were used for the development of this oscillator consisting of a hybrid promoter named as Plac/ara-1, which constitutes the araBAD promoter's activation operator site which is placed in its start site, and lacZYA promoter's repression operator site and placed to the upstream end and downstream end of the start site. AraC protein helps in activation when arabinose is present, and LacI protein acts as a repressor when isopropyl b-D-1-thiogalactopyranoside (IPTG) is absent.

Three co-regulated transcription modules were formed by placing the araC, lacI, and yemGFP (monomeric yeast-enhanced green fluorescent protein) genes that are controlled by three identical copies of the promoter Plac/ara-1. When arabinose and IPTG are added, it activates the promoter which causes the transcription of each component and the presence of arabinose results in the increase of AraC, which corresponds to a positive feedback loop due to which the activity of the promoter gets increased. At the same time, promoter activity gets decreased due to the increased production of LacI, which results in the formation of linked negative feedback loop as shown in Fig. 7.7b. The oscillatory behavior can be driven by the differential activity of these two feedback loops (Stricker et al. 2008).

7.6.7.1 Modeling of Two-Gene Oscillator Using the Stochastic Approach

The dynamics of the promoter is based on the following equations.

$$P_{0,j}^{a/r} + a_2 \overset{k_a}{\rightarrow} P_{1,j}^{a/r} \text{ and}$$

$$P_{1,j}^{a/r} \overset{k_{-a}}{\rightarrow} P_{0,j}^{a/r} + a_2 \; j \in \{0, 1, 2\}$$

$$P_{i,0}^{a/r} + r_4 \overset{2k_r}{\rightarrow} P_{i,1}^{a/r} \text{ and}$$

$$P_{i,1}^{a/r} \overset{k_{-r}}{\rightarrow} P_{i,0}^{a/r} + r_4 \; i \in \{0, 1\}$$

$$P_{i,1}^{a/r} + r_4 \overset{k_r}{\rightarrow} P_{i,2}^{a/r} \text{ and}$$

$$P_{i,2}^{a/r} \overset{2k_{-r}}{\rightarrow} P_{i,1}^{a/r} + r_4 \; i \in \{0, 1\}$$

$$P_{1,2}^{a/r} \overset{k_l}{\rightarrow} P_{L,2}^{a/r} + a_2$$

$$P_{0,2}^{a/r} \overset{k_l}{\rightarrow} P_{L,2}^{a/r}$$

$$P_{L,0}^{a/r} \overset{k_{ul}}{\rightarrow} P_{0,0}^{a/r}$$

Here, $P_{0,j}^{a/r}$ shows the status of the promoter on the plasmid of activator (a) and repressor (r) with AraC dimer (a_2) and LacI tetramer (r_4) bound. Where $i \in \{0, 1\}$ and $j \in \{0, 1, 2\}$. In the next step, the set of reaction is for the transcription, translation, folding of protein, and multimerization of each gene and protein. Due

to the space constraints, all the reactions are not given here. Interested readers may find the detailed reactions in the supplementary file of original article (Stricker et al. 2008).

7.6.8 Mammalian Oscillators

This is an oscillator (Tigges et al. 2009) that is executed in the eukaryotic system to understand their regulatory mechanism, which consists of an expression unit of sense–antisense that encodes for the tetracycline-dependent transactivator (tTA) that is triggered by pristinamycin-dependent transactivator (PIT) as shown in Fig. 7.8c. It comprises two genes: gene A and gene B from which both sense and antisense transcriptions occur from one of the gene. The protein is formed due to the translation of sense transcript which results in the feedback to itself along with the activation of the second gene (B). The translation of the antisense transcript from the first gene (A) is activated by the second gene (B). The transcript is not translated to protein, but it represses the production of sense protein at the translational level,

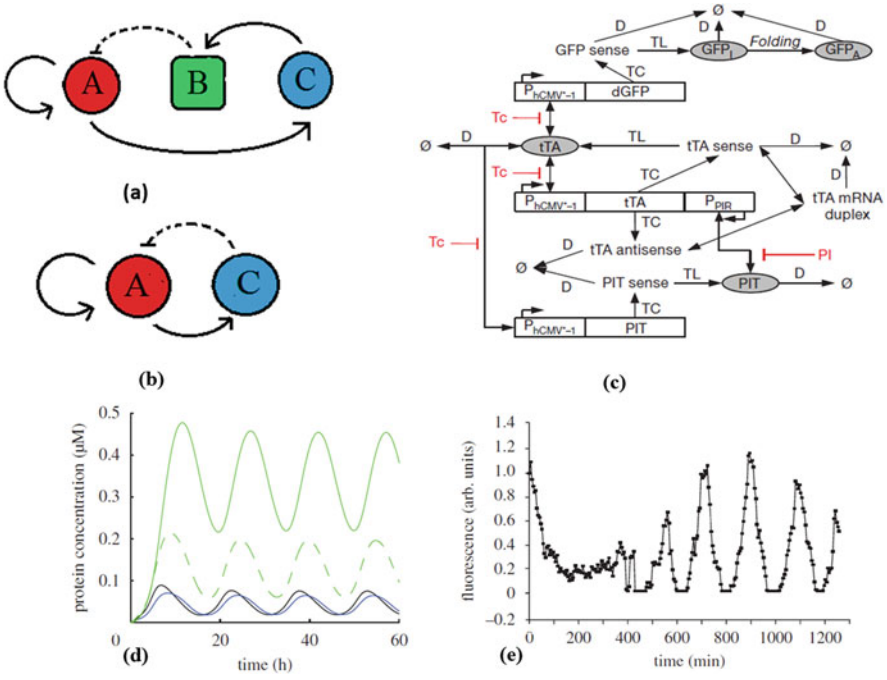


Fig. 7.8 (**a**) Topology of oscillator and (**b**) oscillator as an example of amplified negative feedback topology. (**c**) A cellular process considered for the mathematical model. (**d**) Simulation results of the ODEs of Fussenegger oscillators. (**e**) Oscillations of the oscillator as observed by the time lapse fluorescence analysis. Figures (**c**), (**d**), (**e**) are reproduced with permission from Tigges et al. (2009) © (2009) Macmillan Publishers Limited

due to which negative feedback loop is completed as shown in Fig. 7.8a. Therefore, this Fussenegger oscillator is also a type of amplified negative feedback oscillator, shown in Fig. 7.8b. Moreover, there are delays in the repressive effect which are the extra step involved in a negative feedback loop. ODEs and Gillespie simulations were used for the implementation of this oscillator. These models examined the interaction of RNA polymerases that transcribe the sense and antisense transcripts, which may act as an important repercussion to the repression of sense–antisense strands. After many improvement rounds, a final model is generated by the extensive parameter estimation of data from in vivo analysis.

7.6.8.1 ODEs for the In Silico Implementation of Oscillator

The model is generated by considering the production and degradation of tTA and PIT. The ODEs for the system are written as follows:

$$\frac{dT_m}{dt} = G_1.r_T - k_1 A_m T_m + AT_m.(k_2 k_{DAm}) - k_{DTm} T_m$$

$$\frac{dT_p}{dt} = k_{TL}.T_m - k_{DTp}.T_p$$

$$\frac{dP_m}{dt} = G_2.r_T - k_{DPm}.P_m$$

$$\frac{dP_p}{dt} = k_{TL}.P_m - k_{DPp}.P_p$$

$$\frac{dA_m}{dt} = G_1.r_P - k_1 A_m T_m + AT_m.(k_2 k_{DTm}) - k_{DAm} A_m$$

$$\frac{dAT_p}{dt} = k_1.A_m.T_m - AT_m.(k_2 + k_{DTm} + k_{DAm})$$

$$\frac{dG_m}{dt} = G_3.r_T - k_{DGm}.G_m$$

$$\frac{dG_p}{dt} = k_{TL}.G_m - \left(k_3 + k_{DGp}\right).G_p$$

$$\frac{dG_a}{dt} = k_3.G_p - k_{DGp}.G_a$$

In the given ODEs, T_m is the concentration of mRNA of tTA, T_p is the concentration of protein of tTA, P_m is the mRNA concentration of PIT, P_p is the protein concentration of PIT, A_m is the antisense mRNA of tTA, TA_m is the complex of sense–antisense mRNA, G_m is the GFP in mRNA, G_p and G_a represent the inactive and active GFP, respectively.

7.6.9 Two-Switch Negative Feedback Oscillator

This oscillator (Kim and Winfree 2011) consists of two switches both consisting of a synthetic DNA (each having a regulatory domain, a promoter region, and an output domain), inhibitory molecule, and an activator. An input signal controls these synthetic switches. These switches operate at a particular threshold in response to the input signal of the respective switch. These responses derive from the different hybridization of DNA and RNA molecules, i.e. activation, inhibition, annihilation, and release. The OFF switch consists of a double-stranded DNA having an incomplete promoter region for RNA polymerase as shown in Fig. 7.9b. When a

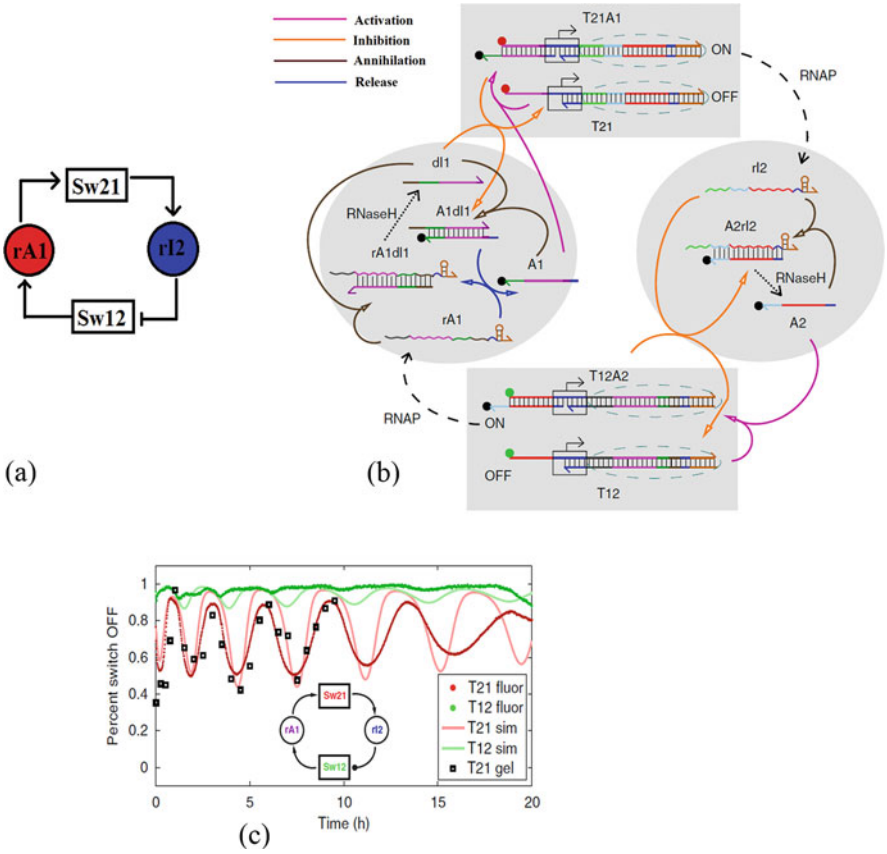


Fig. 7.9 (**a**) Topology of two-switch negative feedback oscillator. (**b**) Schematic circuit diagrams of two-switch negative feedback oscillator. (**c**) Oscillation observed by fluorescence analysis. Figures (**b**), (**c**) are adopted from Kim and Winfree (2011) © (2011) EMBO and Macmillan Publishers Limited

single-stranded DNA activator (A) binds to the DNA molecule, it completes the promoter region due to which this switch gets ON (*activation*). An inhibitor strand of either single-stranded DNA (dI) or single-stranded RNA (rI) molecule can bind to the free floating activators of either single-stranded DNA (A) or single-stranded RNA (rA) that are complementary to the inhibitors resulting in the formation of the activator-inhibitor complex which is functionally inactive (*annihilation*). The free activator strands can reduce the inhibitory strands concentration resulting in the availability of only that strand which had high initial concentration. On the addition of inhibitory strands, if the amount of inhibitory strands rises above the amount of free activators, the switch gets OFF (*inhibition*). There is a toehold region of DNA inhibitor (dI) on A.dI complex that allows the binding of RNA activator (rA) releasing the DNA activator (A) to activate the target switch (*release*) due to the toehold-mediated strand displacement reaction. An inhibitable switch or an activatable switch is formed by assembly of the four hybridization reactions on the basis of the concentration of DNA inhibitor or activator strand.

7.6.9.1 ODEs for the In Silico Implementation

$$\frac{d\,[rA1]}{dt} = k_p.\,[T12A2] - k_d.\,[rA1]$$

$$\frac{d\,[rI1]}{dt} = k_p.\,[T21A1] - k_d.\,[rI2]$$

Here, k_p and k_d are the rate constants of the RNAP and RNaseH, respectively. A change in state of switches is governed by the hybridization reactions depending on the amount of presence of the amount of X. The steady-state response of a switch to the RNA input can be approximated by Hill functions.

$$\tau \frac{d\,[T12A2]}{dt} = \left[T12^{\text{tot}}\right]\left(\frac{1}{1+\left(\frac{[rI1]}{K_I}\right)^n}\right) - [T12A2]$$

$$\tau \frac{d\,[T21A1]}{dt} = \left[T21^{\text{tot}}\right]\left(\frac{1}{1+\left(\frac{[rA1]}{K_A}\right)^n}\right) - [T21A1]$$

Where, n and m represent the Hill exponents, τ shows the relaxation time for the hybridization reaction, and K_A and K_I represent the threshold set by the DNA activator and DNA inhibitor, respectively. Tij represents the sum of the concentrations of ON-state switch [TijAj] and OFF-state switch [TjiAi], and $T21^{\text{tot}}$ gives the sum of all the concentrations of species of [TijAj] and [TjiAi]. We

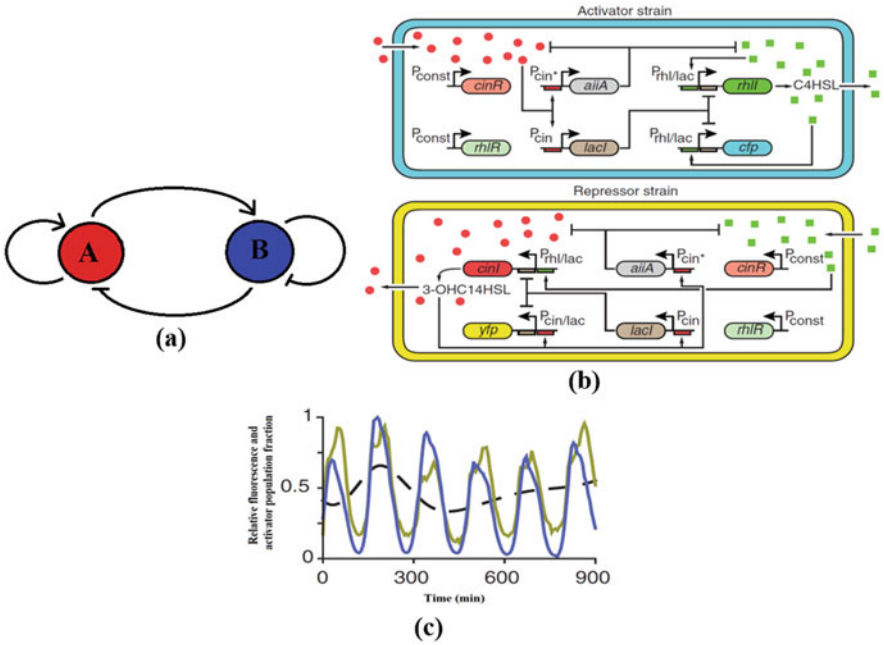


Fig. 7.10 (**a**) Topology of the dual-feedback consortium oscillator. (**b**) Circuit diagrams of both the activator and repressor strains of the dual-feedback consortium oscillator. (**c**) Oscillations observed in fluorescence analysis. Figures (**b**), (**c**) are adopted from Chen et al. (2015) © (2015) AAAS

are omitting the details of the other equations because of the space constraints. Furthermore, positive-feedback loop was also added to study the modularity, and a ring oscillator having three switches was developed and analyzed (Kim and Winfree 2011).

7.6.10 Dual-Feedback Consortium Oscillator

This oscillator was designed primarily to address the challenge of developing a microbial system which shows the population-level behaviors in synthetic biology using the mechanisms of cell signaling for regulation of gene expression. This consortium consists of two different types of cells: activator strain (A) and repressor strain (B) (Chen et al. 2015). Both of these cell types produce two distinct cell-signaling molecules as shown in Fig. 7.10b. These molecules regulate the gene expressions in the cells of both types of strains. The population-level oscillations were produced by these two strains when they were cultured together.

In this model, a signaling molecule C4–homoserine lactone (C4-HSL) is produced by the activator strain which enhances the rate of transcription of *rhII* and *cfp* genes, regulated by the different copies of hybrid promoter $P_{\text{rhI/lac}}$, and also enhances the rate of transcription of target gene in repressor strain. Whereas in the repressor strain, 3-OHC14-HSL is produced as a signaling molecule which inhibits the transcription of *cinI* (regulated by $P_{\text{rhI/lac}}$) and *rhIR* regulated by $P_{\text{cin/lac}}$ as well as represses the target genes in repressor strains by production of repressor LacI as shown in Fig. 7.10b. A coupled negative–positive feedback loop is formed by the two signaling molecules formation mechanisms when these two strains were grown with each other. Furthermore, an enzyme AiiA is produced when both strains were active, which inhibits the formation of both of the signaling molecules.

7.6.10.1 ODEs for the In Silico Implementation of Dual-Feedback Consortium Oscillator

Here, F_a and Y_r represent the productions of reporter proteins CFP and YFP, respectively. The F_a production and Y_r productions are directly proportional to the $P_{\text{rhI/lac}}$ and $P_{\text{cin/lac}}$ promoter's activity, respectively.

$$\frac{dF_a}{dt} = \frac{\eta_{F0}+\eta_{F1}\left(H_a^{\tau}/K_H\right)^{nH}}{1+\left(H_a^{\tau}/K_H\right)^{nH}+\left(L_a^{\tau}/K_L\right)^{nL}} + \frac{d_C F_a}{K_C+R_a+A_a+L_a+F_a+M_a} - dF_a - mF_a$$

$$\frac{dM_a}{dt} = mF_a - \frac{d_C F_a}{K_C + R_a + A_a + L_a + F_a + M_a} - dM_a$$

$$\frac{dY_r}{dt} = \frac{\eta_{Y0}+\eta_{Y1}\left(I_r^{\tau}/K_I\right)^{nI}}{1+\left(I_r^{\tau}/K_I\right)^{nI}+\left(L_r^{\tau}/K_L\right)^{nL}} + \frac{d_C Y_r}{K_C+R_r+A_r+L_r+Y_r+M_r} - dY_r - mY_r$$

$$\frac{dM_r}{dt} = mY_r - \frac{d_C M_r}{K_C + R_r + A_r + L_r + Y_r + M_r} - dM_r$$

M_a represents the mature CFP, and M_r represents the mature YFP, which is described by first-order reaction. η_{F0} and η_{F1} show the basal and final rates of production of CFP, where η_{Y0} and η_{Y1} show the basal and final rates of production of YFP. The rate of production of C4-HSL and 3-OHC14-HSL is given by H_a and I_r, respectively. τ represents the time delay. K_H and K_L represent the EC50 of C4 and IC50 for LacI, respectively. n_H is the Hill-coefficient of C4, and n_L is the Hill-coefficient of LacI. Due to space constraints, we are not detailing here the complete set of equations. Interested researchers may find the details in the original research article describing the construction and realization of this consortium (Chen et al. 2015).

7.7 Hands-On Exercises

In order to practice the simulation of synthetic oscillators, two exercises (using XPPAUT software in a windows machine) are given below in a step-by-step manner.

For that, we need to first download and install XPPAUT following instructions given in http://www.math.pitt.edu/~bard/xpp/xpp.html. Downloaded folder is to be unzipped and to be copied into C:/ drive. One needs to install an X-Window System Server also, if not preinstalled. Xming is one such server that can be installed from https://sourceforge.net/projects/xming/. Input file for the XPPAUT is to be developed as given in the following format and to be saved with an .ode extension.

7.7.1 Sample .ode File for Simulating the Dynamics of Goodwin's Oscillator

$$\frac{dx}{dt} = \left(\frac{a}{(S + k * y)}\right) - b$$

$$\frac{dy}{dt} = (a1 * x) - b1$$

$$\text{param } a = 72, S = 36, k = 1, b = 2, a1 = 1, b1 = 0$$

$$\text{init } x = 7, y = -10$$

done

7.7.2 Sample .ode File for Simulating the Dynamics of Repressilator

$$\frac{dm_1}{dt} = -m_1 + \frac{a}{\left(1 + p_3\hat{}n\right)} + a_1$$

$$\frac{dp_1}{dt} = -b * (p_1 - m_1)$$

$$\frac{dm_2}{dt} = -m_2 + \frac{a}{\left(1 + p_1\hat{}n\right)} + a_1$$

$$\frac{dp_2}{dt} = -b * (p_2 - m_2)$$

$$\frac{dm_3}{dt} = -m_3 + \frac{a}{\left(1 + p_2\hat{}n\right)} + a_1$$

$$\frac{dp_3}{dt} = -b * (p_3 - m_3)$$

#parameters

$$\text{param } a = 10, a_1 = 0.01, b = 0.2, n = 2$$

#initial conditions

$$\text{init } m_1 = 0.2, m_2 = 0.3, m_3 = 0.4, p_1 = 0.1, p_2 = 0.1, p_3 = 0.5$$

$$@\text{bound} = 10,000$$

done

The above given file contains all the differential equations of Goodwin's oscillator and repressilator, initial conditions (init), parameters (param), time step (dt), total run time (total), xplot (x-axis), and yplot (y-axis).

7.7.2.1 Steps for Running the Input Files Using XPPAUT

- Copy the shortcut of XPPAUT to the desktop.
- Run the XMING.
- Drag your file "$_*$.ode" to the shortcut. This will open the file in XPPAUT.
- Set your x-axis from xi vs t option.
- Set your initial run conditions from InitialConds > Range option.
- Go to Window/zoom option and choose fit to see the dynamics.

Once run, one should get the dynamics of Goodwin's oscillator and repressilator, as shown in the Figs. 7.11a, b, respectively.

Acknowledgement AP thanks University Grants Commission, India, for providing PhD fellowship.

Conflict of Interest The authors declare that they have no conflict of interest.

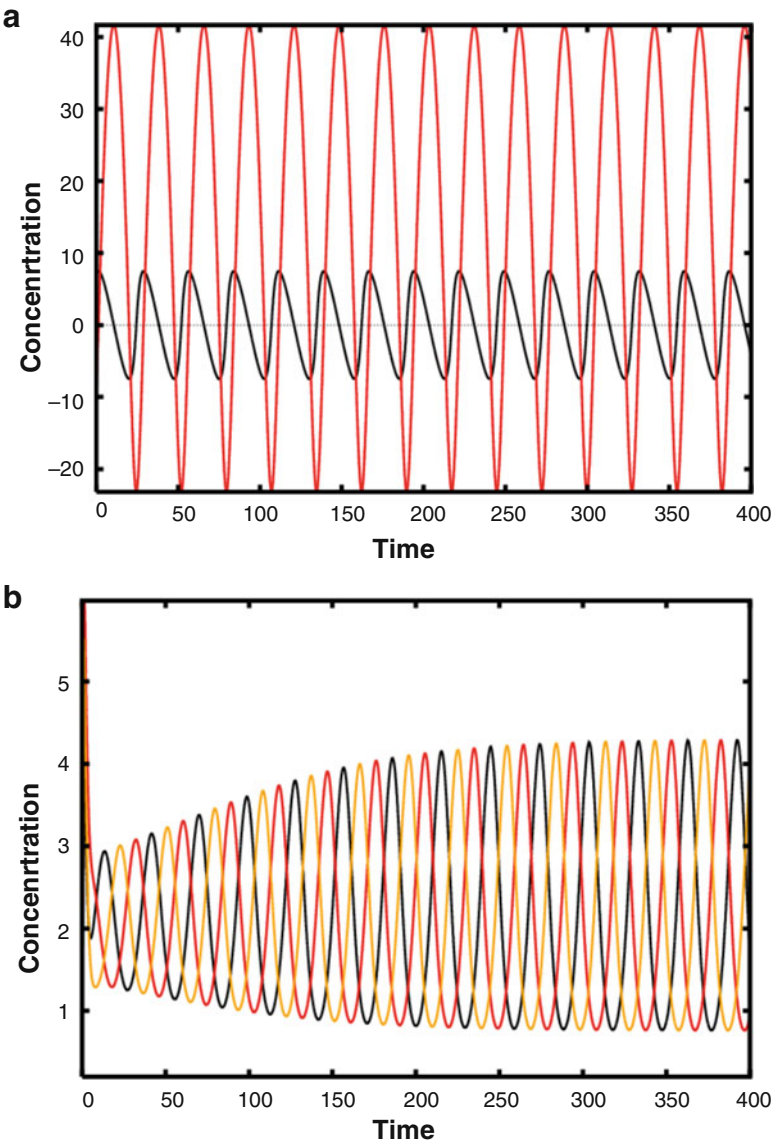


Fig. 7.11 (**a**) Dynamics of Goodwin's oscillator and (**b**) dynamics of repressilator

References

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell 113(5):597–607
Bratsun D, Volfson D, Tsimring LS, Hasty J (2005) Delay-induced stochastic oscillations in gene regulation. Proc Natl Acad Sci USA 102(41):14593–14598
Cao Y, Lopatkin A, You L (2016) Elements of biological oscillations in time and space. Nat Struct Mol Biol 23:1030–1034
Cardon SZ, Iberall AS (1970) Oscillations in biological systems. Biosystems 3(3):237–249
Chandran D, Copeland WB, Sleight SC, Sauro HM (2008) Mathematical modeling and synthetic biology. Drug Discov Today Dis Model 5(4):299–309
Chen Y, Kim JK, Hirning AJ, Josić K, Bennett MR (2015) Emergent genetic oscillations in a synthetic microbial consortium. Science 349(6251):986–989
Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338
Fettiplace R (2017) Hair cell transduction, tuning, and synaptic transmission in the mammalian cochlea. Compr Physiol 7(4):1197–1227
Fraser A, Tiwari J (1974) Genetical feedback-repression II. Cyclic genetic systems. J Theor Biol 47(2):397–412
Fung E, Wong WW, Suen JK, Bulter T, Lee S, Liao JC (2005) A synthetic gene-metabolic oscillator. Nature 435:118–122
Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342
Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361
Glass L, Beuter A, Larocque D (1988) Time delays, oscillations, and chaos in physiological control systems. Math Biosci 90:111–125
Goldbeter A (1997) Modelling biochemical oscillations and cellular rhythms. Curr Sci 73(11):933–939
Goodwin BC (1965) Oscillatory behavior in enzymatic control processes. Adv Enzym Regul 3:425–438
Hahl SK, Kremling A (2016) A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: on fixed points, means and modes. Front Genet 7(46):157
Hasty J, Isaacs F, Dolnik M, McMillen D, Collins JJ (2001) Designer gene networks: towards fundamental cellular control. Chaos (Woodbury, NY) 11(1):207–220
Janson NB (2012) Non-linear dynamics of biological systems. Contemp Phys 53(2):137–168
Karpman VL, Sadovskaya GV (1964) Oscillatory properties of the human body. Bull Exp Biol Med 55(6):649–652
Kim J, Winfree E (2011) Synthetic in vitro transcriptional oscillators. Mol Syst Biol 7:465
Kruse K, Jülicher F (2005) Oscillations in cell biology. Curr Opin Cell Biol 17(1):20–26
Maini P (1996) Oscillations in biology. Trends Biochem Sci 21(10):403–403
Maltoni M, Schwetz T, Tórtola M, Valle JWF (2004) Status of global fits to neutrino oscillations. New J Phys 6:1–37
Purcell O, Savery NJ, Grierson CS, di Bernardo M (2010) A comparative analysis of synthetic genetic oscillators. J R Soc Interface 7(52):1503–1524
Shaviv NJ, Prokoph A, Veizer J (2014) Is the solar system's galactic motion imprinted in the phanerozoic climate? Sci Rep 4:6150
Simakov DSA, Pérez-Mercader J (2013) Noise induced oscillations and coherence resonance in a generic model of the nonisothermal chemical oscillator. Sci Rep 3:2404
Singh V (2015) Modelling methodologies for systems biology. In: Singh V, Dhar PK (eds) Systems and synthetic biology. Springer, Dordrecht, pp 43–62. ISBN: 978-94-017-9513-5. https://doi.org/10.1007/978-94-017-9514-2

Smolen P, Baxter DA, Byrne JH (1998) Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. Am J Phys 274(2):531–542

Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456:516–519

Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. Nature 457:309–312

Tsimring LS (2014) Noise in biology. Rep Prog Phys 77(2):026601

Winfree AT (2001) Populations of attracting-cycle oscillators. In: Winfree AT (ed) The geometry of biological time. Springer, New York, pp 229–257. ISBN: 978-0-387-98992-1. https://doi.org/10.1007/978-1-4757-3484-3

Zeigler BP, Muzy A, Kofman E (2019) Introduction to systems modeling concepts. In: Zeigler BP, Kim TG, Praehofer H (eds) Theory of modeling and simulation. Academic Press, Orlando, pp 3–25. ISBN:978-0-12-778455-7. https://doi.org/10.1016/C2016-0-03987-6

Software-Aided Design of Idealised Programmable Nucleic Acid Circuits

8

Iuliia Zarubiieva and Vishwesh Kulkarni

Abstract

The idea to use nucleic acid as a substrate for design of programmable biomolecular circuits was first introduced almost four decades ago; however, up till now, the field of DNA computing holds many challenges and uncertainties to be discovered. This chapter describes the historical evolution of DNA programming along with its most noticeable breakthroughs till the current days, describes the basics of such important theoretical concepts as DNA strand displacement and Abstract Chemical Reaction Networks, and finally, familiarises the reader with various platforms for in silico synthesis and simulation of genetic circuits.

Keywords

DNA programming · Abstract chemical reaction networks (ACRN) · DNA strand displacement (DSD)

8.1 Introduction

Synthetic biology is an interdisciplinary field laying on the intersection of biology and engineering. It applies the principles from both disciplines in order to design, programme and control the behaviour of biological systems. The origins of synthetic biology can be traced back to1960s when F. Jacob and J. Monod published their work on 'Cellular regulation by molecular networks' (Monod et al. 1961). However, the modern era of this discipline started in early 2000s with the publication of two notable works in *Nature* journal: one on synthetic oscillatory network of transcriptional regulators by M. Elowitz and S. Leibler (Elowitz et al. 2000) and

I. Zarubiieva · V. Kulkarni (✉)
School of Engineering, University of Warwick, Coventry, UK
e-mail: V.Kulkarni@warwick.ac.uk

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_8

another on genetic toggle switch by T. Gardner et al. (2000). In the following years, however, the complexity of programmed circuits increased gradually, so new solutions were required to keep up with the progress.

One of such solutions was the use of nucleic acids to perform computation in biological circuits (Padirac et al. 2013). Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are one of the essential components for all known forms of life. They are responsible for storing genetic information and enabling the production of proteins. DNA can take form of a single strand as well as of multi-stranded complex, where the strands are connected to one another by hydrogen bonds. RNA can only be single-stranded and has quite different functions compared to DNA. Interested readers are directed to learn more about the functionality of nucleic acids in other sources which are widely available.

In the proposed approach, the interaction rules between different nucleic acid strands are defined through well-understood mechanisms of the Watson–Crick base pairing. This way, the biomolecular interactions inside a cell can be precisely programmed simply by choosing the complementary sequences of nucleotides: C (cytosine) G (guanine), A (adenine) or T (thymine), which is replaced by U (uracil) in RNA. Recent advances in synthetic biology approaches paired with the high and predictable DNA programmability made DNA and RNA a highly promising base for the implementation of biochemical circuits.

There are a number of approaches for the design of nucleic acid circuits. One example is the programmed RNA self-assembly for catalysis of hydrogen biosynthesis (Delebecque et al. 2013), and another, DNA structures operating on basic logic were used to release the molecular payload only in case the two antigens, each of them indicating a specific type of cancer, were present simultaneously on the target cell surface (Douglas et al. 2012). But with the recent advances in synthetic biology, more complex dynamics was required for the regulation of biochemical processes in biomolecular circuits (Daniel et al. 2013). A few strategies that can implement such dynamics have recently been proposed, based on principles of DNA strand displacement (Zhang et al. 2011), DNA (Montagne et al. 2011) and RNA enzymes (Kim et al. 2011). They were used for various applications, namely, implementing feedback control mechanisms (Chen et al. 2013), predator–prey dynamics (Fujii et al. 2013), transcriptional oscillators (Weitz et al. 2014) and control of the in vitro protein production (Franco 2012). Each of the various methodologies for implementing nucleic acid circuits can have their own benefits and drawbacks. That is why it is highly useful to apply them to the same system design in order to compare their performance and help identify design strategies that combine the strengths of these different approaches (Fig. 8.1).

A coherent theoretical approach of implementing such circuits by the means of idealised DNA strand displacement reactions has recently been illustrated in Lakin et al. (2016), Song et al. (2016), Yordanov et al. (2014) and Zhou et al. (2017). In the ideal world, a re-use of well-characterised modules, which is called modularity, simplifies the design of circuits. However, it requires modifications for such a scale up of biomolecular circuits since the building blocks that were characterised separately may change their behaviour once connected to other components. In

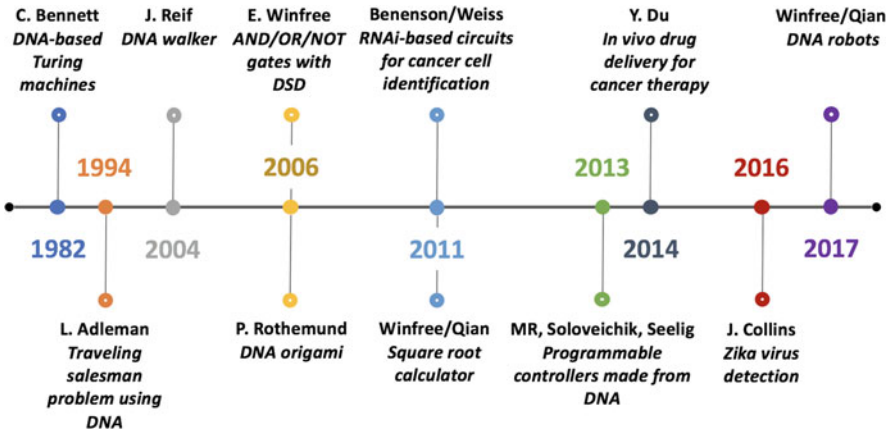


Fig. 8.1 Some of the most notable results in the history of DNA programming. The authors specified are either main authors or group leaders

the context of biochemical networks, Hartwell et al. (1999), Lauffenburger (2000), Andrianantoandro et al. (2006) and Purnick et al. (2009) made the first reliable presentation of this concept.

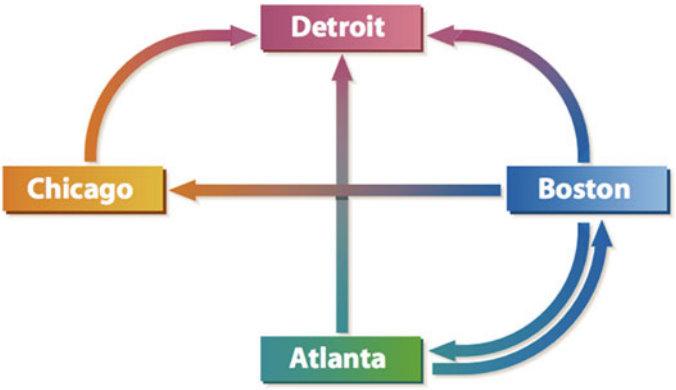
8.2 Evolution of DNA Programming

This section gives an overview of some of the most noticeable results in DNA programming.

DNA-based Turing machine. The history of DNA computing started in 1982, when Charles Bennett designed Turing machine using DNA. In this work, he proposed a conceptual design of Brownian Turing machine and the transition rules for changing the state. However, this approach involved hypothetical enzymes needed for performing the transition. In 1996, Paul Wilhelm and Karl Rothemund extended this work and presented restriction enzymes and ligases that were commercially available (Bennett 1982).

Travelling salesman problem using DNA. Next, in 1994, Leonard Adleman solved a travelling salesman problem using DNA. Travelling salesman is a classic problem in computer science where a salesman must travel between N cities and visit each of them only once. The order in which he visits them is irrelevant, but each connection between two cities has its weight, or cost, which determines the 'difficulty level' of the trip. The salesman wants to keep the cost and the travel distance of the overall trip as low as possible (Fig. 8.2).

Adleman presented a map, where each of seven cities and each of 14 connecting flights between them were assigned a DNA sequence. Each city name has its complement, and each connecting flight consists of half of each corresponding city name—for example, if Atlanta is ACTT**GCAG**, Boston is **TCGG**ACTG, then the



CITY	DNA NAME	COMPLEMENT
ATLANTA	ACTTGCAG	TGAACGTC
BOSTON	TCGGACTG	AGCCTGAC
CHICAGO	GGCTATGT	CCGATACA
DETROIT	CCGAGCAA	GGCTCGTT

FLIGHT	DNA FLIGHT NUMBER
ATLANTA - BOSTON	GCAGTCGG
ATLANTA - DETROIT	GCAGCCGA
BOSTON - CHICAGO	ACTGGGCT
BOSTON - DETROIT	ACTGCCGA
BOSTON - ATLANTA	ACTGACTT
CHICAGO - DETROIT	ATGTCCGA

Fig. 8.2 Adapted from Adleman (1998). A map of cities for travelling salesman problem represented with DNA sequences.

flight Atlanta–Boston becomes **GCAGTCGG**. Because of the Watson–Crick base pairing, the complementary sequences will bind together. Now every flight could pair up into a double-stranded complex with two cities, and every city could be connected by two flights. This way, DNA complex will grow in length, with each DNA flight number connected to complementary DNA city names. It took Adleman seven consecutive days to successfully complete the DNA computation (Adleman 1998).

DNA walker. The next significant discovery was a creation of a DNA walker—a nanodevice that is able to 'walk' autonomously along a DNA 'track'. First presented by John Reif in 2003, he and his group were able to demonstrate it experimentally the following year. The walker operates based on the phenomena of DNA strand displacement, which we will focus on more closely in Sect. 8.4.

The process is as follows: one leg of the DNA walker is *strand A*, which is bound to complementary *strand B*, through normal base pairing. *Strand A* contains an additional unbound sequence on one of its ends, named a toehold. Next, the A-B

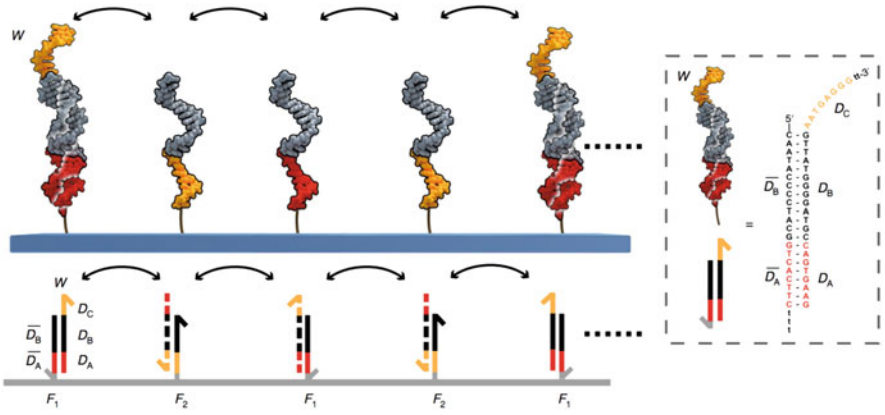


Fig. 8.3 Reproduced with permission from Li et al. (2018). Schematic showing the intended mechanism of DNA walker movement

complex encounters *strand C*. This strand is complementary to *strand A*, including the toehold part. Once the toehold of *strand C* binds with the toehold of *strand A*, it begins to displace each nucleotide of *strand B*, base by base, until *strand B* has been completely replaced by *strand C*. *Strand B* then dissociates from *strand A* and the process can begin again (Yin et al. 2004).

In 2018, the team led by Nils Walter presented a DNA walker able to 'walk' as fast as 300 nm/min. The applications of DNA walkers include nanomedicine, nanorobotics, biosensing and much more (Li et al. 2018) (Fig. 8.3).

DNA origami. A real breakthrough in nanoengineering happened in 2006 when Paul Rothemund presented his work on DNA origami. The concept includes folding of a long single DNA strand supported by multiple small strands, acting like staples. The staple strands hold the main strand in various places, helping to form the required two- or three-dimensional shape. Some of the examples of DNA origami are a smiley face, a star, a triangle (two-dimensional) and various cubic and tube-like shapes (three-dimensional). Results can be directly observed via several imaging techniques (Rothemund 2006) (Fig. 8.4).

Initially created as a form of art, in the coming years, DNA origami was developed into various applications, including drug delivery, nanoelectronic circuitry, enzyme immobilisation, etc. As an example, a group from the Harvard University Wyss Institute reported the vessels for drug delivery that were capable of self-assembly and self-destruction. Such vessels consist of open DNA-origami tube that has a 'door' which can be opened or closed. The tube containing the drug is closed by a DNA aptamer which is designed to recognise certain proteins associated with a disease. Once the container reaches cell recognised as infected, the aptamer breaks apart and the tube releases the drug. The researchers used a disease model for leukaemia and lymphoma. The results of the research were highly promising; however, 1 month after the announcement, the publication was retracted.

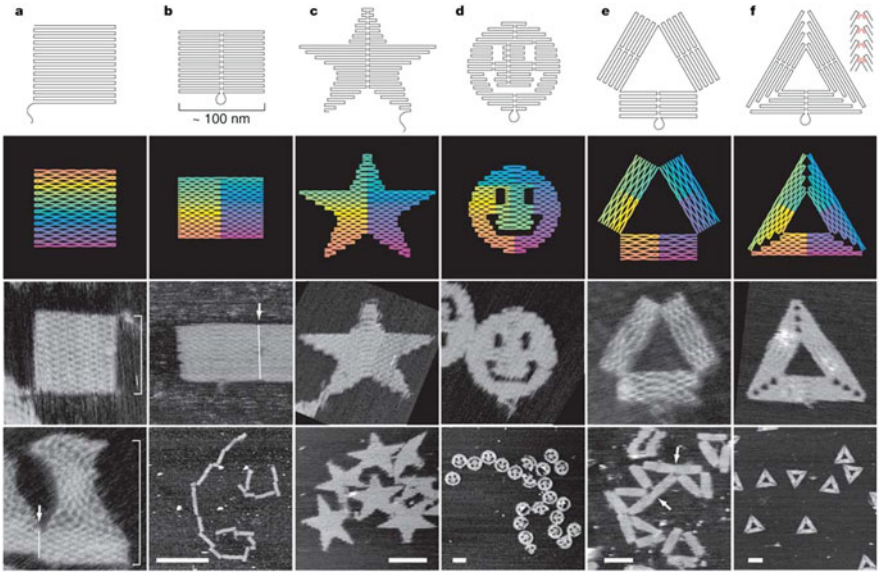


Fig. 8.4 Reproduced with permission from Rothemund (2006). Examples of DNA origami: (**a**) square; (**b**) rectangle; (**c**) star; (**d**) disk with three holes; (**e**) triangle with rectangular domains; (**f**) sharp triangle with trapezoidal domains and bridges between them. All images and panels without scale

AND/OR/NOT gates with DSD. Same year, Eric Winfree's team published the results on building basic logic elements such as AND/OR/NOT gates. The researchers stated that the specified set is enough for computation of any Boolean logic function. The gates treat single DNA strands as inputs and outputs, and the approach works entirely on the DNA strand displacement principle.

An AND gate produces an output signal only when two input signals are both present. The gate design consists of two single strands and a double-stranded complex. In the first reaction, a first oligonucleotide input (G_{in}) reacts with a gate through toehold-mediated strand displacement. This creates a waste and an intermediate product with an open toehold which then participates in the next reaction together with the second input F_{in}. One strand of the original gate complex is fluorescent, and the other one is fluorescence quenching. The second input separates these, eventually causing the fluorescence that can be recorded as the circuit's readout (Seelig et al. 2006).

The designed gates were a subject to errors, specifically, a gate could produce not enough output, or it could 'leak' by releasing the output strand without any trigger. Both types of errors could be corrected using signal restoration.

Square root calculator. Together with Lulu Qian, Eric Winfree has published another prominent work, presenting a circuit that computed a square root of four-bit binary numbers for inputs from 0000 to 1111. The circuit involved 74 initial DNA species, excluding inputs, and in total, it contains 130 different strands, each

of them being 15–33 nucleotides long. The design uses dual-rail logic and consists of a combination of AND and OR logic gates. The results are read off in binary from four different fluorescent outputs (Qian et al. 2011).

RNAi-based circuits for cancer cell identification. In 2011, a group led by Yaakov Benenson and Ron Weiss presented intriguing results for anticancer therapy. They designed a device that detects expression levels of a customisable set of microRNAs and selectively initiates apoptosis for the specified cell type without affecting other cells. They demonstrated successful results for HeLa cells; however, they also observed certain degree of false-positive cell detection and cell death as well as false-negative cell survival. Despite the observed error rate, this work was an important step to demonstrate how synthetic biological networks can trigger pre-programmed biological behaviour upon detection of specific conditions. The work also defined that in order to use the designed circuit in cancer therapeutics, the challenge of efficient in vivo DNA delivery should be overcome. The research opened the possibility of in vitro drug screening and monitoring of the developmental process (Xie et al. 2011).

Programmable controllers made from DNA. After a success of programming language Visual DSD that was developed by Microsoft Research, they teamed up with David Soloveichik and Georg Seelig to design programmable DNA controllers that could work with DNA sensors and motors. This work illustrated the diversity of computing using DNA strand displacement. Unlike Boolean logic circuits which were a more popular choice among researchers, analogue devices that were presented in this work allow complex signal processing of various biological and chemical inputs. Components of the controller can be obtained from biologically synthesised DNA, which is more beneficial compared to its chemically synthesised counterpart due to lower synthesis error. The algorithm was tested on three fundamental reaction types: non-catalytic ($A + B \rightarrow C$), catalytic ($A + B \rightarrow C + B$) and autocatalytic ($A + B \rightarrow C + 2B$) and identified correct stoichiometry and kinetics (Chen et al. 2013).

The consensus algorithm that is used for the controller compares two populations of input signals, and the signal that has a lower concentration is fully removed and substituted by the higher value signal. This approach agrees with the one used in computing across networked systems and provides a proof that CRN algorithms can be directly translated into the DNA controllers. The downside of the design is the computation speed. Andrew Phillips from Microsoft Research notes that the algorithm is still quite slow. 'This algorithm takes 15 hours. You wouldn't have a phone that runs on these algorithms. Their advantage is that they can potentially run inside cells'.

In vivo drug delivery for cancer therapy. In 2014, a group of Chinese researchers presented the results where self-assembled DNA origami nanostructures were used as vehicles for delivering anticancer drug (doxorubicin) into cancer tumours. They showed that triangle-shaped origami structures fitted particularly well for tumour accumulation. The vehicles were prepared through the self-assembly using large number of staple strands, after what the doxorubicin was placed inside of the structure. The result analysis showed that the doxorubicin-containing DNA origami

demonstrated outstanding antitumour efficiency without noticeable toxicity in nude mice with breast tumours. The approach proved to be prominently efficient for in vivo therapeutics and demonstrated the potential of DNA origami nanostructures for the successful drug delivery (Zhang et al. 2014) (Fig. 8.5).

DNA robots. In 2017, the collaboration of groups of Eric Winfree and Lulu Qian gave another outstanding result which is a DNA robot capable to autonomously sort cargo. Most of the robots developed earlier could perform only one simple operation—walking in a direction defined by user. So, the researchers were devoted to create a DNA nanomachine that could perform complex nanomechanical tasks. They aimed to tackle two major difficulties that are met when designing DNA robots—namely, modularity and algorithm simplicity. The robot moves along a DNA origami surface and collects two types of cargo, followed by the delivery of this cargo to a specified final location. The robot does not require any energy supply for its operation. It is also possible for several DNA robots to work simultaneously on the same surface and to perform the task collectively (Thubagere et al. 2017) (Fig. 8.6).

The designed robots were reported to perform in average 300 steps needed for fulfilling the task. The design could be further improved to include the possibility of sorting multiple types of cargo and running multiple tasks separately, where for each task a specific number of robots is allocated depending on the task difficulty level. The developed methodology could also be used for various purposes other than sorting cargo. The authors envision the possibility applying the principles of macroscopic robots to design biomolecular robots that will be able to work in microscopic environment.

8.3 DNA Strand Displacement (DSD)

Various studies have proven that it is possible to describe any abstract chemical reaction network (CRN) with appropriately designed DNA strand displacement reactions. Biomolecular computation, that is, computation performed by biomolecules, has a great potential for use in supercomputing, healthcare, smart materials, nanotechnology and soon. Nucleic acids are particularly suited to serve as such biomolecules since these form stable structures that can be inserted into cells, and, furthermore, since the interactions between these species can be controlled to a good extent by modifying their nucleotide sequences: this essentially exploits the well-known Watson–Crick base pairing. Current section will give an overview of DNA strand displacement process.

DNA strand displacement (DSD) is a competitive hybridisation reaction. It is conceptually subdivided into three stages: toehold binding, branch migration and strand dissociation. The strand displacement is mediated by short domains called *toehold*, and long domains participate in *branch migration*. Although branch

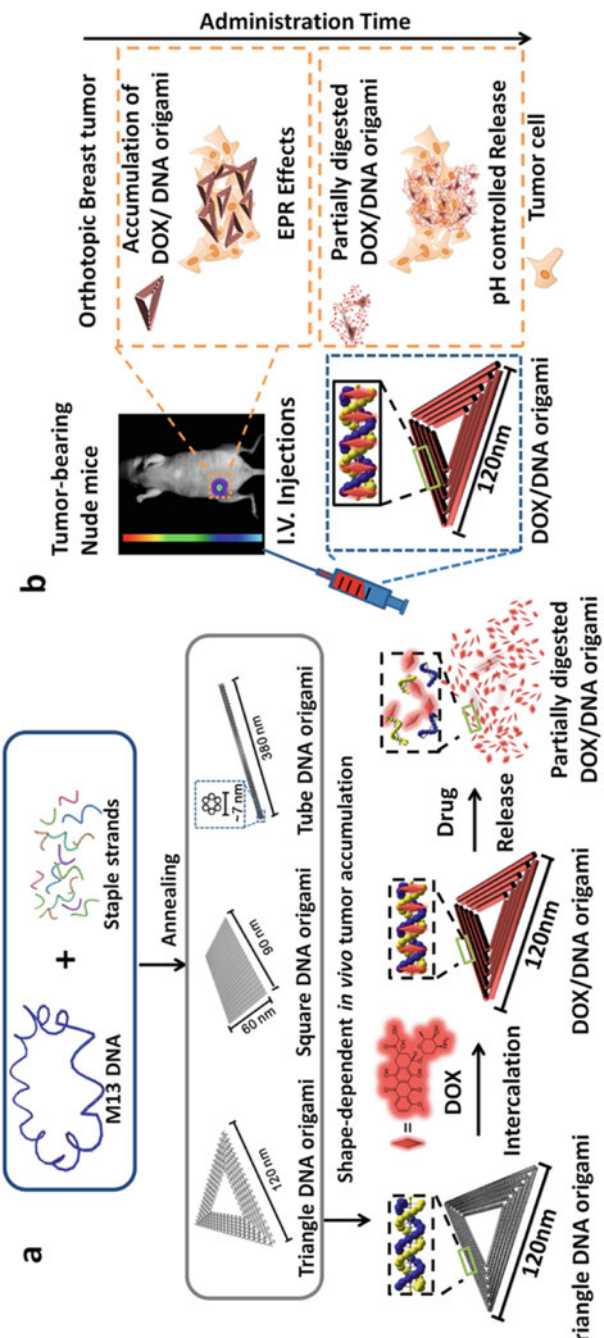


Fig. 8.5 Reproduced with permission from Zhang et al. (2014). Schematic design of DNA-carrier complex. (**a**) Formation of triangular, square and tube origami. The triangle-shaped DNA origami showed the best results for doxorubicin administration. (**b**) Tail-injected DOX/DNA origami complexes accumulated in the breast tumour of nude mice because of EPR effects

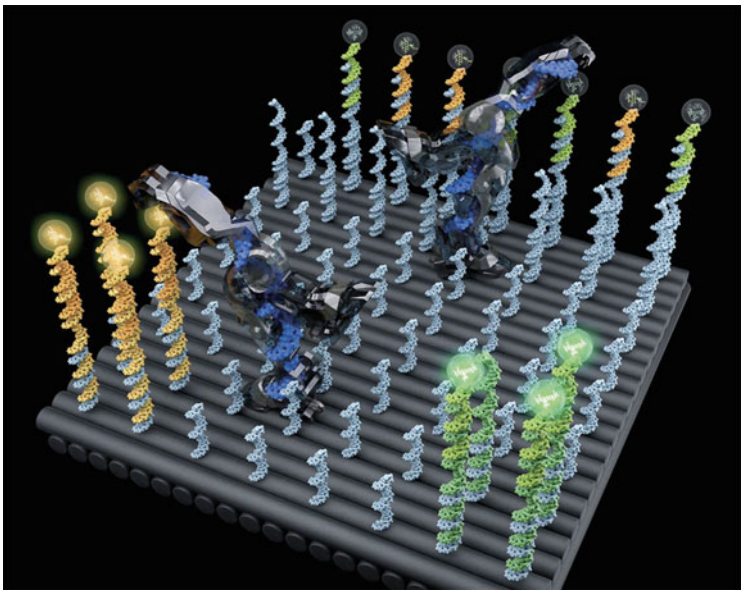


Fig. 8.6 Adapted from Thubagere et al. (2017). Conceptual illustration of two DNA robots that collectively perform cargo-sorting task on a DNA origami surface

migration process was familiar to researchers since 1970s, toehold-mediated strand displacement (TMSD) has been properly introduced only in 2000 by using TMSD for designing DNA tweezers (Yurke et al. 2000).

The process starts with a double-stranded DNA complex. One of the strands in the complex is called *original strand* and another—*protector strand*. Original strand has an unbound toehold domain, which is complimentary to the toehold of the third strand—an *invading strand*, which is single-stranded DNA with nucleotide sequence complementary to the original strand. The toehold regions of original and invading strands bind together with hydrogen bonds and create a DNA complex consisting of three strands. The toehold binding step effects the reaction rate, which can be tuned by varying the length and sequence composition of the toehold region. After the binding of original and invading strands, the invading strand starts replacing the protector strand out of the complex through branch migration process. Eventually, the protector strand is being released from the complex and can further participate in other reactions, for which it can become an invading strand. The forward process is more energetically favoured and proceeds with reaction rate up to six orders of magnitude faster than the reverse reaction. This is a fundamental mechanism for genetic material exchange.

8.4 Chemical Reaction Networks (CRN)

To be able to discuss DNA strand displacement in more details, we need to first mention the theory that gave a basis for describing system performance with DSD. It has been shown that chemical reaction networks can be used to describe the dynamic behaviour of biological systems. Living systems rely on complex networks of chemical reactions to control the concentration of molecules in space and time. Dynamical properties of such networks were widely studied by chemists and physicists after the introduction of the law of mass action, which states that a chemical reaction rate is directly proportional to the product of the activities or concentrations of the reactants.

A chemical reaction network comprises a set of reactants, a set of products and a set of reactions. For example, let us take the following set of chemical reactions:

$$A + B \rightarrow C + D$$
$$D \rightarrow E + F$$
$$F \rightarrow \varnothing$$

We can describe any abstract chemical reaction network (CRN) with linear ordinary differential equations (ODEs). The main focus of mathematical modelling of chemical reaction networks is on the change over time of the concentrations of the various chemical species present in the solution. Following the above set of reactions, let us assume a is the concentration of species A, b—concentration of species B and so on. As all of these concentrations will change over time, they can be written as $a(t)$, $b(t)$, etc.

Then, we can combine them into a vector:

$$x(t) = \begin{pmatrix} a(t) \\ b(t) \\ \vdots \end{pmatrix}$$

and their evolution with time can be written as follows:

$$\dot{x} \equiv \frac{dx}{dt} = \begin{pmatrix} \frac{da}{dt} \\ \frac{db}{dt} \\ \vdots \end{pmatrix}$$

By solving the designed ODEs, it is possible to find the species concentration at the particular time point.

Example. Here, we provide an example of multi-step Michaelis–Menten kinetics model and how it is described with corresponding ODE model:

An enzyme E interacts with substrate S, and their interaction produces an output product P:

$$E + S \underset{k^-}{\overset{k^+}{\rightleftharpoons}} E \cdot S \overset{k_{\text{cat}}}{\rightarrow} E + S + P$$

We are interested in the dynamics of the output P:

$$\frac{dp}{dt} = k_{\text{cat}} e \cdot s$$

Now, we need to express the complex $E \cdot S$. Assume it reaches steady state quickly, then:

$$\frac{de \cdot s}{dt} = k^+ es - k^- e \cdot s - k_{\text{cat}} e \cdot s = 0$$

Equilibrium concentration of complex $E \cdot S$ at steady state is:

$$(e \cdot s)_{ss} = \frac{k^+}{k^- + k_{\text{cat}}} e \cdot s = \frac{1}{KM} e \cdot s$$

where KM is known as Michaelis constant.

If we further assume that the amount of substrate S is much larger compared to E, then $e^{\text{tot}} = e + e \cdot s$. After substituting this into equation of $E \cdot S$ at the steady state, we obtain:

$$e^{\text{tot}} = e + \frac{e \cdot s}{KM} \Rightarrow e = \frac{e^{\text{tot}}}{1 + \frac{s}{KM}}$$

Then, we can obtain the production rate for product P as follows:

$$\frac{dp}{dt} = k_{\text{cat}} e^{\text{tot}} \frac{s}{KM + s}$$

This can be rewritten into more familiar form of:

$$\frac{dp}{dt} = \frac{v_{\max} s}{KM + s}$$

At very high concentrations of substrate S, we obtain the maximum production rate, which is equal to $v_{\max}$. When s is equal to KM, we are at the half-max production rate $v_{\max}/2$.

8.4.1 Representing Signals as Chemical Species

To ensure consistency, the notation used in Yordanov et al. (2014) and Oishi et al. (2011) is used throughout in this chapter. For example, a bidirectional (reversible) is represented as follows:

$$X_1 + X_2 \underset{\delta_2}{\overset{\delta_1}{\rightleftharpoons}} X_3 + X_4$$

where X_i are chemical species with X_1 and X_2 being the reactants and X_3 and X_4 being the products. Here, δ_1 and δ_2 denote the forward and backward reaction rates, respectively. Unlike a unimolecular reaction that has only one reactant, a multimolecular reaction has two or more reactants. Degradation of a chemical species X at rate K (or conversion of X into an inert form at a rate K) is denoted by $X \xrightarrow{K} \varnothing$.

Whereas signals in systems theory can take both positive and negative values, biomolecular concentrations (with Molar (M) as unit) can only take non-negative values. Thus, following the same approach suggested in Yordanov et al. (2014) and Oishi et al. (2011), we represent a signal, x as the difference in concentration of two chemical species, x^+ and x^-. Here, x^+ and x^- are, respectively, the positive and negative components of x such that $x = x^+ - x^-$. The consequence of adopting this scheme is that there is no unique representation for a particular signal. As an example, $x = 20\text{M}$ can be represented by both $x^+ = 50\text{M}$ and $x^- = 30\text{M}$ or equivalently, $x^+ = 20\text{M}$ and $x^- = 0\text{M}$. In practice, x^+ and x^- can be realised as single-strand DNA molecules, as illustrated in Yordanov et al. (2014), where these complementary positive and negative components would annihilate each other at reaction rate η (i.e. $x^+ + x^- \xrightarrow{\eta} \varnothing$). A key advantage of using this scheme is that it allows the realisation of the 'subtraction' operation, as discussed further below.

Oishi and Klavins were first to show how to use idealised abstract chemical reactions to describe theoretic operators of any linear system, namely, integration, summation and gain blocks (Oishi et al. 2011). It was shown that it only requires the following three types of chemical reactions: catalysis, annihilation and degradation. In Yordanov et al. (2014), this number is further reduced to only two.

Throughout the rest of the chapter, equations with superscript $\pm$ and $\mp$ are used as shorthand notations that represent the '+' and '–' individual reactions—for example, $x_i^{\pm} \xrightarrow{K} x_i^{\pm} + x_o^{\pm}$ are implied as the set of two reactions: $x_i^{+} \xrightarrow{K} x_i^{+} + x_o^{+}$ and $x_i^{-} \xrightarrow{K} x_i^{-} + x_o^{-}$. Likewise, the notation $x_i^{\pm} \xrightarrow{K} x_i^{\pm} + x_o^{\mp}$ is used to represent the set of two reactions: $x_i^{+} \xrightarrow{K} x_i^{+} + x_o^{-}$ and $x_i^{-} \xrightarrow{K} x_i^{-} + x_o^{+}$. To be in line with Oishi et al. (2011), we will represent such a set of reactions as $x_i^{\pm} \xrightarrow{K} x_i^{\pm} + x_o^{\pm}$ and $x_i^{\pm} \xrightarrow{K} x_i^{\pm} + x_o^{\mp}$.

As noted in Oishi et al. (2011), one limitation of representing signals as the difference of concentrations is that the requirement of having the same reaction rate, K, for both positive and negative components may not be easy to implement experimentally. However, as shown in Oishi et al. (2011), this requirement can be relaxed if the annihilation rate, η in the annihilation reaction, $x_o^+ + x_o^- \overset{\eta}{\to} \varnothing$ is chosen to be sufficiently large. Hence, we assume this condition of $\eta \gg K$ throughout the rest of this chapter.

8.4.2 Representing Basic Building Blocks as a Set of ACRNs

Below subsections illustrate how to represent basic operators such as scalar gain, integrator, summation/subtraction (linear), as well as multiplication and power of N block (nonlinear) in terms of three basic reaction types.

8.4.2.1 Scalar Gain

Let $x_o = Kx_i$ where x_i is the input, x_o is the output and K is the gain. This operation is implemented using the following set of abstract chemical reactions:

$$x_i^{\pm} \overset{\gamma K}{\to} x_i^{\pm} + x_o^{\pm},$$
$$x_o^{\pm} \overset{\gamma}{\to} \varnothing,$$
$$x_o^{+} + x_o^{-} \overset{\eta}{\to} \varnothing,$$

where γK, γ and η are the kinetic rates associated with catalysis, degradation and annihilation, respectively.

Following generalised mass-action kinetics, it shows that the gain block can be represented with the following set of ODEs:

$$\frac{dx_o^+}{dt} = \gamma\left(Kx_i^+ - x_o^+\right) - \eta x_o^+ x_o^-$$
$$\frac{dx_o^-}{dt} = \gamma\left(Kx_i^- - x_o^-\right) - \eta x_o^+ x_o^-$$
$$\frac{dx_0}{dt} = \frac{dx_o^+}{dt} - \frac{dx_o^-}{dt} = \gamma\left(Kx_i - x_0\right)$$

According to the final value theorem, the steady-state value of x_o for constant input x_i is given by $\lim_{t\to\infty} x_o(t) = Kx_i(t)$.

8.4.2.2 Summation

Consider the summation operation $x_o = x_i + x_d$, where x_i and x_d are the inputs and x_o is the output. The summation is implemented as follows:

$$x_i^{\pm} \xrightarrow{\gamma} x_i^{\pm} + x_o^{\pm},$$
$$x_d^{\pm} \xrightarrow{\gamma} x_d^{\pm} + x_o^{\pm},$$
$$x_o^{\pm} \xrightarrow{\gamma} \varnothing,$$
$$x_o^{+} + x_o^{-} \xrightarrow{\eta} \varnothing.$$

Using the following set of abstract chemical reactions, the subtraction $x_o = x_i - x_d$ is implemented:

$$x_i^{\pm} \xrightarrow{\gamma} x_i^{\pm} + x_o^{\pm},$$
$$x_d^{\mp} \xrightarrow{\gamma} x_d^{\mp} + x_o^{\pm},$$
$$x_o^{\pm} \xrightarrow{\gamma} \varnothing,$$
$$x_o^{+} + x_o^{-} \xrightarrow{\eta} \varnothing.$$

Scaled summation $x_o = K(x_i + x_d)$, and scaled subtraction $x_o = K(x_i - x_d)$, can be implemented by choosing the catalysis rates in the construct of summation block to be γK.

8.4.2.3 Integration

Consider the integrator $x_o = \int x_i dt$, where x_i is the input and x_o is the output. Using the following set of abstract chemical reactions:

$$x_i^{\pm} \xrightarrow{\gamma} x_i^{\pm} + x_o^{\pm},$$
$$x_o^{+} + x_o^{-} \xrightarrow{\eta} \varnothing,$$

such an integrator is implemented.

The ODE representing the operation is $\frac{dx_o}{dt} = K x_i$.

8.4.2.4 Multiplication

Consider the multiplication operation $x_o = x_i \times x_d$, where x_i and x_d are the inputs and x_o is the output. The multiplication is implemented as follows:

$$x_i^{\pm} + x_d^{\pm} \xrightarrow{\gamma} x_o^{\pm},$$

$$x_o^{\pm} \xrightarrow{\gamma} \varnothing,$$

$$x_o^{+} + x_o^{-} \xrightarrow{\eta} \varnothing.$$

The ODEs representing the operation are as follows:

$$\frac{dx_o^{+}}{dt} = \gamma\left(x_i^{+} \times x_d^{+} - x_o^{+}\right) - \eta x_o^{+} x_{o,}^{-}$$

$$\frac{dx_o^{-}}{dt} = \gamma\left(x_i^{-} \times x_d^{-} - x_o^{-}\right) - \eta x_o^{+} x_{o,}^{-}$$

$$\frac{dx_o}{dt} = \frac{dx_o^{+}}{dt} - \frac{dx_o^{-}}{dt} = \gamma\left(x_i \times x_d - x_o\right).$$

8.4.2.5 Power of N

Consider the operation $x_o = x_i^N$, where x_i is the input, x_o is the output and N is the constant specifying power. Similar to multiplication operation, it is implemented using the following set of abstract chemical reactions:

$$N \times x_i^{\pm} \xrightarrow{\gamma} x_o^{\pm},$$

$$x_o^{\pm} \xrightarrow{\gamma} \varnothing,$$

$$x_o^{+} + x_o^{-} \xrightarrow{\eta} \varnothing.$$

8.5 From CRN to DSD—Transition Rules

The transition between chemical reaction networks and its physical representation by DNA molecules was greatly facilitated in 2010 by David Soloveichik, Georg Seelig and Eric Winfree. In their paper, they explored the ways to use DNA strand displacement reactions as a primitive in order to approximate the dynamic behaviour of arbitrary chemical reactions systems (Soloveichik et al. 2010). The authors saw three reasons for the need of such approximation: using only CRNs was limiting the design and functionality of synthetic circuits as it could not fully capture its possibilities and limitations; there was more than extensive literature on CRNs, as well as widely used approach for approximation of CRNs with a set of suitably chosen ODEs; and finally, the fundamental model of chemical reaction system could act as a useful programming model for biomolecular system design.

In the proposed approach, each chemical species is represented as a single-stranded DNA molecule which reacts with double-stranded auxiliary species, which are present in much higher concentration than the signal species. At each step, a signal species can only react with an auxiliary species, eliminating the direct interaction. This way the auxiliary species that are present in significantly larger

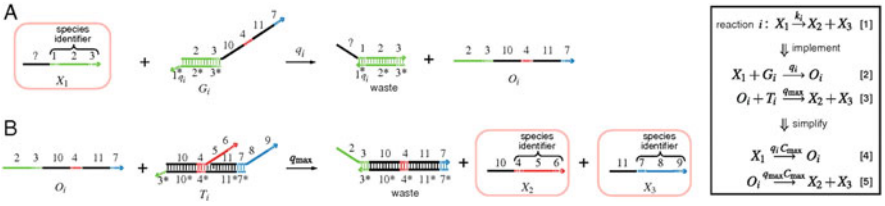


Fig. 8.7 Adapted from Soloveichik et al. (2010). DNA implementation of unimolecular reaction $X_1 \rightarrow X_2 + X_3$. The overall DSD process consists of two reactions (A and B) where product of reaction A serves as reactant of reaction B

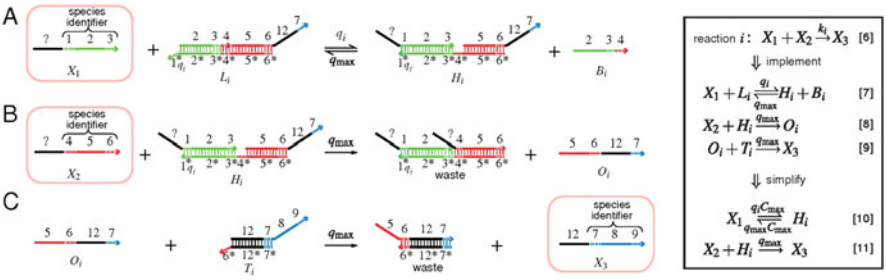


Fig. 8.8 Adapted from Soloveichik et al. (2010). DNA implementation of bimolecular reaction $X_1 + X_2 \rightarrow X_3$. The overall DSD process consists of three reactions (A, B and C) where product of reaction A serves as reactant of reaction B and product of B serves as reactant of reaction C

concentrations can mediate the correct reaction and ensure that all the required signal species participate in it.

Figures 8.7 and 8.8 provide an illustration for a unimolecular ($X_1 \rightarrow X_2 + X_3$) and bimolecular reaction ($X_1 + X_2 \rightarrow X_3$), respectively. Signal species (X_1, X_2, X_3) are represented as four-domain single strands, starting domain being a history domain and the following three being a species identifier. History domain depends on the previous reaction that produced the species, and the species identifier consists of one recognition domain surrounded by two toehold domains. Two molecules with different history domains but same species identifier will be recognised as the same species. The signal strand is considered active when it is fully unbound and inactive otherwise.

The authors also came up with a list of system requirements that are essential for correct operation. First, for all the double-stranded complexes, only toehold domains can be unbound and capable to hybridise with incoming single strands. Second, the toeholds should be short enough so that the toehold binding is reversible and not permanent. And lastly, only the desired target should have the correct combination of domains, which would prevent undesired reactions.

In such a way, any non-catalytic, catalytic, autocatalytic, annihilation or degradation reactions could be represented with DNA. Two-domain and three-domain representations have also probed themselves useful in a number of other studies.

8.6 In Silico and In Vitro Synthesis and Simulation of Genetic Circuits

1.1. Cello

In the last several years, many studies have focused on design analysis of biomolecular circuits. Probably, one of the most noticeable recent researches is a programming language presented by MIT in 2016, which allows users to design DNA circuits that give new functions to living cells. Using this language, a user can create any function of interest, such as detecting and responding to certain environmental conditions, and the software will design the DNA sequences that can achieve it. 'Cello' (a computational language) allows automatically generating circuits for highly specified physical systems and operating conditions. It constructs the desired genetic Boolean logic circuits by choosing from a library of simpler repressor-based NOT/NOR logic gates and connecting them to one other (Nielsen et al. 2016).

Cello makes use of Verilog, which is a commonly used hardware description language for programming computer chips. A user specifies the desired circuit function in Verilog code that captures the desired computational operation. Further, the functional details are transformed into a ready DNA sequence in three steps:

1. Verilog text code is converted into a truth table and further into a circuit diagram.
2. Given the circuit diagram, specific regulators are assigned to each gate in the diagram. Since gates based on different regulators may have different response functions, not all repressors can be paired together. The challenge here is that they may interfere with each other once placed in the complex environment of a living cell. To resolve this issue and help identify correct connections, Cello uses simulated annealing principle.
3. From a circuit diagram and regulators assignment, Cello can now create a DNA sequence using combinatorial design.

Cello language is easy to use and does not require much of a prior knowledge of programming, which makes it easily accessible. This development allowed to save huge amount of time for building complex circuits. Having a ready DNA sequence that you can test greatly reduces the time for obtaining successful results. During the experimentation, 40 of 60 designed circuits with different functions operated correctly from the first time (Fig. 8.9).

Future applications for this kind of programming include designing bacterial cells that can release a cancer drug when they detect a tumour or creating yeast cells that can pause their own fermentation process if too many toxic by-products build up.

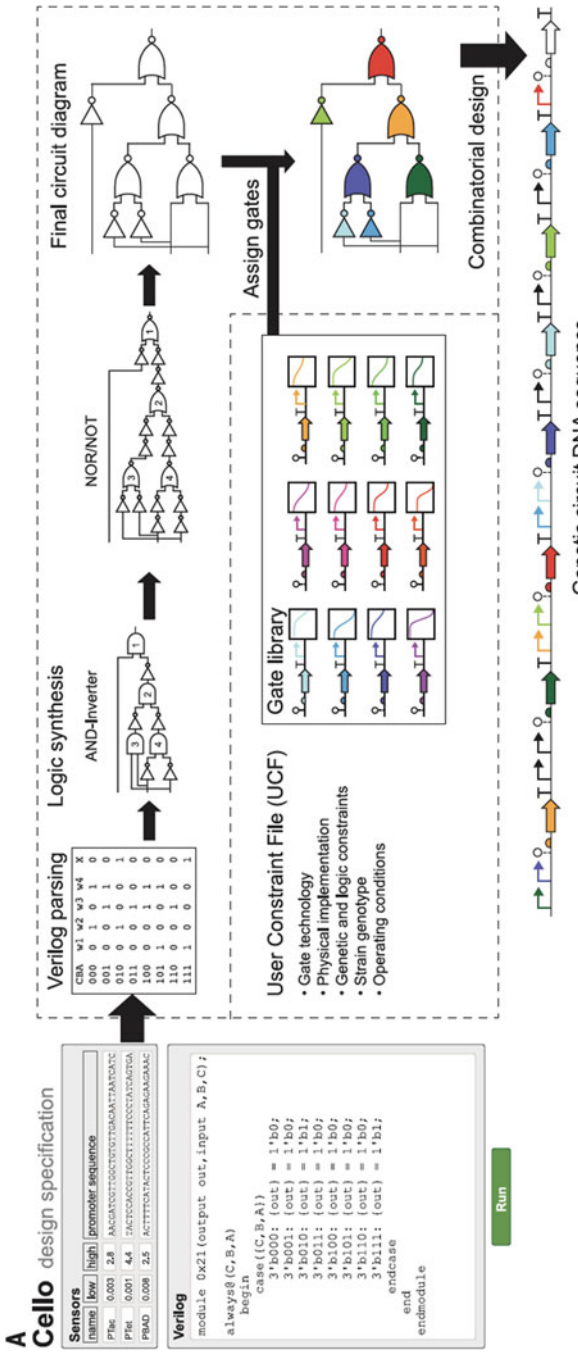


Fig. 8.9 Adapted from Nielsen et al. (2016). Overview of Cello computation process. First step is user input, then Verilog code conversion to circuit diagram, followed by assigning of specific regulators to each gate, and, finally, creation of DNA sequence of genetic circuit

8.6.1 Station B

Another promising platform for programming biology announced in 2019 is Station B by Microsoft Research. The platform is a result of over 20 years of cooperation between Microsoft and their technology, academic and commercial partners which allowed to develop new methods and technology for programming biology.

The goal of the platform is to improve each step of the Design-Build-Test-Learn workflow in the process of programming biological systems. On the first stage, the biological programming languages will be collected into a hierarchy of biological abstractions, each with their associated methods of analysis. Next, high-level programmes will be translated to DNA code, accounting for the biological experiments that are intended. Further, biological experiments will be performed using lab robots, and lastly, the received experimental data will be used as a training data by a wide range of learning methods. As new experiments are performed, the knowledge base will be updated via automated learning.

Station B is a promising tool that uses computational models to describe interactions within a cell that are too complex for humans to parse using notebooks and spreadsheets alone. It will allow scientist to make use of underlying knowledge that has been collected for decades and perform their work more efficiently, reliably and cost-effectively.

8.6.2 Visual DSD

Today, a wide range of tools is available to model the molecular conformations of nucleic acid strands, which are useful in the design of DNA-based biomolecular circuits. The software Visual DSD (Phillips et al. 2009; Lakin et al. 2011, 2012) is a good choice for modelling a broad class of strand displacement systems, including all that can be expressed in terms of chemical reactions.

Visual DSD is a programming language for designing and simulating computational devices made of DNA. Visual DSD uses DNA strand displacement as the main computational mechanism, which allows devices to be designed solely in terms of nucleic acids. It accounts for system complexity and the potential unwanted interference between molecules in the system. After inputting a collection of DNA strands of interest, the software will show how these DNA strands interact with each other. It can also be used to predict their behaviour over time. Visual DSD can be used to detect and fix bugs in silico before the circuit is attempted to be built in vitro or in vivo (Fig. 8.10).

Visual DSD compiles a collection of DNA molecules into a chemical reaction network (CRN). It also includes a stochastic simulator, which computes a possible trajectory of the system and plots the populations of species over time, together with a deterministic simulator, which forms and solves an ordinary differential equation

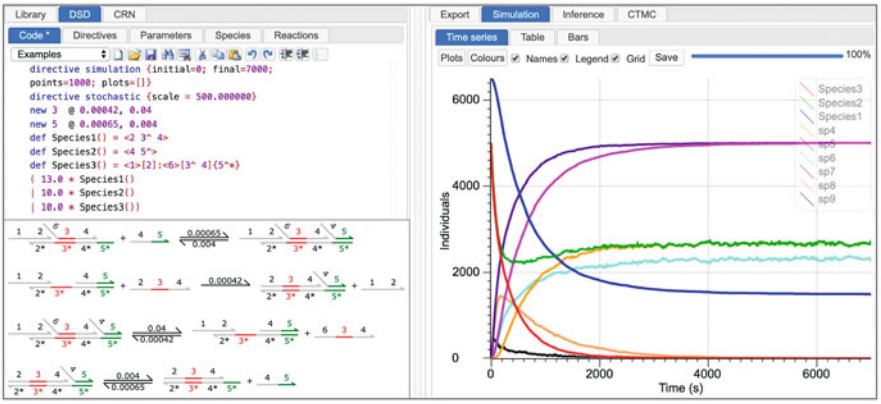


Fig. 8.10 Illustration of Visual DSD interface and functionality. Left top side contains molecular program coded in Visual DSD language, left down side contains the reactions generated by the software, and right side contains simulation results that show the change of species concentration over time

(ODE) representation of the dynamics of the system. The reachable state space of the system can also be constructed as a continuous time Markov chain (CTMC).

8.7 Conclusions and Future Remarks

For the last two decades, the field of synthetic biology experienced a substantial growth and has produced many notable achievements. Moreover, the speed of the technical progress increases gradually, as more and more discoveries are made. As synthetic circuits became more complex, there was a need for moving away from traditional design techniques, and the usage of nucleic acids for circuit design has successfully filled that niche. Programming with DNA has a number of advantages, including its high programmability due to Watson–Crick base pairing, structure stability which allows the structures to be inserted into cells, no requirement for any additional components, precise control of organisation and dynamics at the molecular level and, finally, ability to interact organically with other biological entities. The range of applications for using DNA as a substrate for circuit design varies greatly, from healthcare to smart materials and to nanotechnology.

The field of DNA programming evolved greatly till present day. Started with using DNA for Turing machine in C. Bennet's theoretical work in 1982, it came a long way till the complex DNA robots, DNA origami and programmable DNA controllers. However, many problems are still open for solving, including 'leaky' reactions, fail in modularity and methods for control of kinetic parameters. One of the possible solutions to overcome one or more of these problems could be the use of xeno nucleic acids (XNA) as information carriers instead of DNA. XNA is a synthetic analogue of DNA that has a different sugar backbone structure. Currently,

there are at least six types of successfully synthesised XNAs. The discovery of XNA gave a start to the field of xenobiology, which describes the use of novel biological systems that may not yet be familiar to the research community (Anosova et al. 2016).

Another promising novel approach, currently very popular among synthetic biologists, is the use of CRISPR/Cas9 for targeted genome editing. It is an enzyme that allows to identify and cut desired DNA strands and is target specific (Jinek et al. 2012). This technology gives high versatility for producing user-guided mutations within the DNA and alteration of the genome.

All this suggests that there will be less and less restrictions in technical and, possibly, cost aspects of the circuit synthesis in the near future, but the field is still behind on the full understanding of all the biological and technical aspects that are required to design flawless biomolecular systems.

Acknowledgements This research is supported, in parts, by the EPSRC INDUSTRIAL CASE AWARD (CASE Voucher 16000070), Microsoft Research, and the EPSRC/BBSRC grant BB/M017982/1 to the Warwick Integrative Synthetic Biology Centre.

References

Adleman LM (1998) Computing with DNA. Sci Am, 54–61

Andrianantoandro E et al (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2(1):2006.0028

Anosova I et al (2016) The structural diversity of artificial genetic polymers. Nucleic Acids Res 44(3):1007–1021

Bennett CH (1982) The thermodynamics of computation – a review. Int J Theor Phys 21(12):905–940

Chen Y-J et al (2013) Programmable chemical controllers made from DNA. Nat Nanotechnol 8:755–762

Daniel R et al (2013) Synthetic analog computation in living cells. Nature 497:619–623

Delebecque CJ et al (2013) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333:470–474

Douglas SM et al (2012) A logic gated nanorobot for targeted transport of molecular payloads. Science 335:831–834

Elowitz MB et al (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338

Franco E (2012) Analysis, design, and in vitro implementation of robust biochemical networks. PhD dissertation, California Institute of Technology

Fujii T et al (2013) Predator–prey molecular ecosystems. ACS Nano 7:27–34

Gardner TS et al (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342

Hartwell L et al (1999) From molecular to modular cell biology. Nature 402:C47–C52

Kim J et al (2011) Synthetic in vitro transcriptional oscillators. Mol Syst Biol 7:465

Jinek M et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

Lakin MR et al (2011) Visual DSD: a design and analysis tool for DNA strand displacement systems. Bioinformatics 27(22):3211–3213

Lakin MR et al (2012) Abstractions for DNA circuit design. J R Soc Interface 9(68):470–486

Lakin M et al (2016) Supervised learning in adaptive DNA strand displacement networks. ACS Synth Biol 5(8):885–897
Lauffenburger D (2000) Cell signaling pathways as control modules: complexity for simplicity? Proc Natl Acad Sci U S A 97(10):5031–5033
Li J et al (2018) Exploring the speed limit of toehold exchange with a cartwheeling DNA acrobat. Nat Nanotechnol 13:723–729
Monod J et al (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harbor Sympos Quantit biol 26:389–401
Montagne K et al (2011) Programming an in vitro DNA oscillator using a molecular networking strategy. Mol Syst Biol 7:466
Nielsen AK et al (2016) Genetic circuit design automation. Science 112(6281):aac7341
Oishi K et al (2011) Biomolecular implementation of linear I/O systems. IET Syst Biol 5(4):252–260
Padirac A et al (2013) Nucleic acids for the rational design of reaction circuits. Curr Opin Biotechnol 24:575–580
Phillips A et al (2009) A programming language for composable DNA circuits. J R Soc Interface 6(Suppl 4):S419–S436
Purnick P et al (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10(6):410–422
Qian L et al (2011) Scaling up digital circuit computation with DNA strand displacement cascades. Science 332(6034):1196–1201
Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440:297–302
Seelig G et al (2006) Enzyme-free nucleic acid logic circuits. Science 314(5805):1585–1588
Soloveichik D et al (2010) DNA as a universal substrate for chemical kinetics. PNAS 107(12):5393–5398
Song T et al (2016) Analog computation by DNA strand displacement circuits. ACS Synth Biol 5(8):898–912
Thubagere J et al (2017) A cargo-sorting DNA robot. Science 357:1112
Weitz M et al (2014) Diversity in the dynamical behavior of a compartmentalized programmable biochemical oscillator. Nat Chem 6:295–302
Xie Z et al (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells. Science 333(6047):1307–1311
Yin P et al (2004) A unidirectional DNA walker that moves autonomously along a track. Angew Chem Int Ed 43:4906–4911
Yordanov B et al (2014) Computational design of nucleic acid feedback control circuits. ACS Synth Biol 3:600–616
Yurke B et al (2000) A DNA-fuelled molecular machine made of DNA. Nature 406(6796):605–608
Zhang DY et al (2011) Dynamic DNA nanotechnology using strand-displacement reactions. Nat Chem 3:103–113
Zhang Q et al (2014) DNA origami as an in vivo drug delivery vehicle for cancer therapy. ACS Nano 8(7):6633–6643
Zhou C et al (2017) Four-analog computation based on DNA strand displacement. ACS Omega 2(8):4143–4160

Digital Circuit Design for Biological and Silicon Computers

9

Matthias Függer, Manish Kushwaha, and Thomas Nowak

Abstract

Modern electronic computing devices profit from two far-reaching design choices: the digital abstraction of a continuous signal domain into a binary domain to tolerate noise, and the use of the complementary metal-oxide-semiconductor (CMOS) technique in gates to reduce noise and power consumption. Biological circuit design in the last two decades has been greatly inspired by the digital abstraction of electronic circuits, while still using a relatively analog machinery. With the constantly growing techniques and components available in synthetic biology, the question arises if chip design choices and other lessons learned from silicon computers can be transferred to biological designs. In this context, we discuss here the similarities and differences between the two fields concerning representation of circuits, solutions to computational problems, their implementations, and challenges.

All authors contributed equally.

M. Függer
CNRS, LSV, ENS Paris-Saclay, Université Paris-Saclay, Inria, Saint-Aubin, France
e-mail: mfuegger@lsv.fr

M. Kushwaha (✉)
Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France
e-mail: manish.kushwaha@inra.fr

T. Nowak
Université Paris-Sud, CNRS, Orsay, France
e-mail: thomas.nowak@lri.fr

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_9

Keywords
Circuit design · Logic gates · Biological circuits · Digital circuits · Synthetic biology · CMOS

9.1 Introduction

Evolving for the past four billion years (Dodd et al. 2017), life on Earth today is highly diverse with an estimated 8.7 million species (Mora et al. 2011). Despite this enormous diversity, one key characteristic that differentiates all living cells from non-living matter is their property of response to stimuli (Macklem and Seely 2010). Living cells receive information from their environment, process that information, and then effect a response. Therefore, in the simplest "information processing" sense of computing (Smith 2002), cells can be seen as tiny computing machines. This notion has sustained itself through Monod to today as the successes of molecular and cell biology have continued to reveal insights into the mechanistic bases of cellular functioning (Monod 1972; Nicholson 2019). Conversely, the similarities in information flow between living systems and computing devices have also led to the design and construction of synthetic biological circuits inspired from electronic circuitry (Selberg et al. 2018).

In this chapter, we explore the similarities and differences between computation by biological and silicon computers. We first describe the nature of information in the two system types. Next, we outline how the information flow can be represented as chemical reaction networks or communicating hardware processes. Finally, we define the kind of computational problems and the circuits that can be used to implement their solutions.

9.1.1 Analog Versus Digital Computation

Most information in the physical world is analog in nature, varying between low and high values on a continuous scale (Sauro and Kim 2013). This is true not just of biologically relevant signals like metabolite concentration, heat, light, pH, and osmotic pressure, but also of electrical signals that run through human-made electronic circuits. Current and voltage in electronic circuits can vary across their different parts depending on other circuit properties like material (doping profiles) and geometry (Agarwal and Lang 2005). For both biological and electronic circuits special processing of analog signals can be used to generate their digital abstractions (Teo et al. 2015), using analog components that favor a binary all-or-nothing behavior. Many such circuits have been implemented over the past two decades inside living cells (Moon et al. 2012).

Before being applied to biological circuitry, however, this special treatment of analog signals was exploited for building digital computer architectures (Agarwal and Lang 2005). The main reason for the success of the digital abstraction in computer circuit designs is its tolerance to noise. By introducing a threshold voltage

that separates two voltage zones corresponding to the binary digits 0 and 1, some noise in the signal can be tolerated as long as it does not force the signal outside of the correct zone. In contrast, it is impossible to build an analog adder circuit that outputs the exact sum of two voltages and that tolerates *any* substantial amount of noise. The digital abstraction also significantly simplifies the specifications of circuit interfaces and thus the re-usability of components. Additionally, it enables the use of techniques and results from Boolean logic to further speed up and simplify the design process and the development of design automation tools.

9.1.2 Natural Biological Computation

Natural biological systems process several continuous signals in their extra- and intracellular environments to make decisions (Teo et al. 2015). Bacterial cells sense gradients of chemical signals to modulate their movements to swim towards chemoattractants and away from chemorepellants (Sourjik and Wingreen 2012). Similarly, they make auxotrophic decisions on upregulating and downregulating different parts of their metabolism depending on the available nutrients (Monod and Jacob 1961). Some bacteria use environmental cues to switch between different life cycle stages (Oliva et al. 2018). In developmental pathways, cells of multicellular organisms make differentiation decisions based on gradients of mRNA molecules across their body length (Spirov et al. 2009). In summary, information processing in natural biological systems seems to use hybrid approaches that combine aspects of analog and digital circuitry. While the analog parts of the circuitry process and output continuous signals, digital parts are used for decision-making between discrete options (Erez et al. 2017), often for steps that require substantial downstream commitment. This hybrid architecture appears to be much more powerful and has been applied in bio-inspired design of computational methods, for example, in neural computing where perceptrons use analog logic to generate weighted sums and subsequently a digital activation function to reach a classification decision (Rosenblatt 1958).

9.1.3 Synthetic Biological Computation

In contrast to natural systems that have optimized analog-digital hybrid architectures over evolutionary time scales, rational design of synthetic biological circuits in the past two decades has focused mainly on building digital logic circuitry due to its relative ease of implementation (Teo et al. 2015). This has involved building one or more levels of NOT, AND, OR, NAND, and NOR gates in single cells or across multicellular consortia. Implementation has been done at different layers of biological information processing: the gene expression layer of transcription and translation, the metabolic layer of enzymatic reactions, and the signal transduction layer of small molecules and their sensors (Nielsen et al. 2016; Katz 2017; Kiel et al. 2010; Pandi et al. 2019). Applications of these circuits have ranged

from detection of pollutants and diagnostic biomarkers, to smart therapeutics and metabolic engineering (Khalil and Collins 2010). Despite the extensive work in biological circuit design, the process remains quite ad hoc with many manual steps needed for prototyping and testing. However, new tools for computer-aided design are beginning to show promise for the future of design automation (Nielsen et al. 2016).

9.2 Representation

This section discusses two different symbolic representations of circuits, one from microbiology (chemical reaction networks) and one from electrical circuits (CHP). Symbolic representations, even when similar at a superficial level, can have a profound impact not only for the computer tools that manipulate these representations but also for the ease of understanding and expression for the end user. The persistent community discussions to identify the "best" programming language bears witness to the immense value of symbolic representation. The importance of precise and workable representations is well-understood in the synthetic biology community. Consequently, a number of (visual) representational standards are already being tested (Hucka et al. 2003; McLaughlin et al. 2016).

9.2.1 Chemical Reaction Networks: A Language for Biochemical Reactions

A chemical reaction network specifies a set of reactions involving a set of molecules. Each reaction defines which molecules, and in which quantities, it takes as its input and which molecules, and in which quantities, it produces as its output. Additionally, each reaction has a rate, which determines how often the reaction occurs in a given time interval.

For instance, the reaction $A + B \xrightarrow{\alpha} C + D + E$ takes one unit of A, one unit of B, and produces one unit each of C, D, and E at the rate α. The frequency at which the reaction happens is determined not only by the rate α, but also by the concentrations of A and B. Denoting the concentrations of A and B by $[A]$ and $[B]$, respectively, the propensity of the above reaction is equal to $\alpha \cdot [A] \cdot [B]$.

Different operational interpretations of chemical reaction networks exist. The two most common ones are the ordinary differential equation (ODE) model and the stochastic model. In the ordinary differential equation model, each molecule has a real-valued continuous concentration whose derivative is equal to the sum of the propensities of reactions that produce the molecule minus the sum of the propensities of reactions that consume it. In the stochastic model, concentrations of molecules are measured as integer-valued counts (normalized to volume), and a reaction's next occurrence time is randomly chosen according to an exponential random variable with parameter equal to the reaction's propensity (Gillespie 1977). The deterministic ODE model can be interpreted as the mean field solution of

the stochastic model, that is, it approximates the behavior of the expected value. Figure 9.3 (right) shows ODE and stochastic traces with two different molecule numbers: the higher the number of involved molecules the less likely it becomes for the stochastic traces to differ greatly from the expected ODE trace.

9.2.2 CHP: A Language for Digital Circuit Design

Hardware description languages like Verilog and VHDL provide means to specify the behavior of a circuit not only as a structural composition of basic gates, but also by abstract semantics as they are known from programming languages for software (comprising constructs like loops, conditional execution, etc.). As such, Verilog and VHDL have evolved as standards for design entry and exchange format for circuits in silico, but recently also for microbiological combinational circuits (Nielsen et al. 2016).

Throughout this chapter, we will make use of a semantically simpler hardware description language that has recently gained attention, the CHP (communicating hardware processes) language. Its use in a comparative article like this is motivated by two facts: (a) the language emphasizes a circuit's nature as a collection of simple but highly concurrently operating components and (b) we will argue in Sect. 9.4 that it is closely related to, although being an abstraction of, the reaction based notation used for the design and modeling of microbiological pathways.

CHP Programs Introduced by Martin (1989), CHP is a language that adapts the concept of communicating sequential processes (Hoare 1978), originally developed for sequential software modules that are executed in parallel and communicate with each other, for hardware processes. In this notation a circuit component is one of the following terms that indicates (1) a *variable assignment* $x := x'$ with x being a variable and x' being its assigned value (0 or 1, or any Boolean formula which may include variables), (2) a *guard* $[G]$, where G is a Boolean expression that evaluates to 0 or 1, (3) a *sequential composition* $a; b$ of terms a and b such that b starts only after a is complete, (4) a *parallel composition* $a \parallel b$ of terms a and b such that a and b run concurrently, or (5) a *repeat execution* $*[[a]]$ of term a.

Other brackets are used in straightforward way to group terms.[1] Besides its CHP term, a circuit has an initial value (0 or 1) for all its variables.

Behavior of CHP Programs We will describe the behavior of such a circuit with the use of a simple example with initial values $A = B = 0$ and CHP term

$$*[[[A = 0]; (A := 1 \parallel B := 1)]]$$

[1] In favor of a more concise presentation we present a simplified form of CHP in this chapter. For a full description we refer the reader to Martin's original paper (Martin 1989).

It specifies a circuit that waits until guard $A = 0$ becomes true and then, in parallel, sets the value of variables A and B to 1 (typically with a certain delay). Immediately after that, it repeats and waits until $A = 0$. Since initially $A = 0$, both A and B will indeed be set to 1.

Translation to Gates While the above CHP term might look sufficiently simple to directly derive a gate-level implementation from it, more complex terms may not be that easily derivable. CHP terms are thus iteratively translated into terms of specific structure: a list of production rules. A *production rule* is a term $*[[[G]; x := x']]$, also denoted as $G \to x := x'$ for simplicity, that can be read as: when G becomes true assign x' to x, where x' is either the constant 0 or the constant 1. All production rules are executed in parallel. The behavior, in accordance with the semantics of CHP, is as follows: whenever a guard is true its assignment is executed (with a certain delay). CHP notation can be distinguished from chemical reaction network notation by a rate that is indicated above the arrow when describing a chemical reaction.

For example, the circuit with initial values $A = B = 1$ and $C = D = E = 0$ with production rules

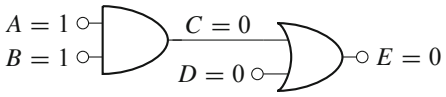
$$A \wedge B \to C := 1 \tag{9.1}$$

$$\neg(A \wedge B) \to C := 0 \tag{9.2}$$

$$C \vee D \to E := 1 \tag{9.3}$$

$$\neg(C \vee D) \to E := 0 \tag{9.4}$$

is now easily seen to be equivalent to the following gate-level implementation (initial values are indicated)



where the first two production rules are the AND gate and the latter two the OR gate. Note that initially rule (9.1) is executed since $A \wedge B = 1 \wedge 1 = 1$, resulting in an update of C to 1 after some delay. Then rule (9.3) is executed, resulting in an update of E to 1.

While in circuits for in silico implementation it is forbidden to have two production rules whose guards are both true at the same time and that set a variable to different values, such rules are plausible in microbiological settings (see Sect. 9.4.2). Depending on the intended semantics such conflicts can be resolved by defining a hierarchy among rules or by resolving conflicts by taking into consideration the time since the guard is activated (or deactivated) and a strength of the rule.

9.3 Computational Problems

From a formal point of view, computational problems can be distinguished into single-shot problems and those problems that require reactive behavior.

Single-shot problems have classically been the main focus of research in computer science. A problem of this class is stated as a function that maps elements of an input space to elements of an output space according to some specification. Classically, elements of the input and output spaces are finite words of the Boolean alphabet $\{0, 1\}$. For example, the problem to compute the logical OR satisfies the specification $\mathrm{OR}(x_1, \ldots, x_n) = x_1 \vee \cdots \vee x_n$. In computer science, solutions to such problems are classically searched for in terms of Turing machines and Boolean circuits, briefly described below.

Turing machines capture an abstract notion of computing architectures where computation is driven by a single processing unit that manipulates strings, one symbol at a time, starting from the input string and finishing at the desired output string. It therein assumes that functions like reliable non-volatile storage of (arbitrarily long) strings and reliable reading and manipulation of a string at a certain position are provided by the concrete architecture, an assumption that is valid for most in silico architectures but not yet for synthetic biological ones.

In *Boolean circuits* computation is driven by Boolean gates, which all operate in parallel, simply transforming binary input signals to output signals and which a priori lack any storage elements. These solutions are closer to the microbiological systems that we are discussing here. While searches for solutions in classical computer science are restricted to the so-called combinational Boolean circuits (cf. Sect. 9.3.1.1) that are stateless, we will consider both stateful and stateless Boolean circuits for their practical relevance in real world implementations.

Importantly, for single-shot problems, it is implicitly assumed that a Boolean circuit that represents a history-less solution can be repeatedly evaluated with changing inputs and its result does not depend on that of a former evaluation.

Reactive systems, by contrast, are systems whose behavior can depend on past inputs. A reactive system continuously receives inputs and produces corresponding outputs that may depend on its history. Given a behavior of the environment which provides the controller's inputs and reacts to its outputs, reactive systems' goal is to provide appropriate outputs such that the combined system, together with the environment, satisfies a given specification.

Analogously to the example of the single-shot OR problem, one may consider the problem of designing a controller that operates in discrete time steps $1, 2, 3, \ldots$: In each time step, it receives an input bit from the environment and has to produce an output that is the OR operation over all input bits received thus far. Such a device could be useful in settings where a controller has to raise an alarm signal once it senses a trigger condition as its input.

9.3.1 Implementations

Reactive problems inherently require implementations that store an internal state while single-shot problems do not. There may, however, be good reasons for using solutions that use internal state for single-shot problems, e.g., smaller or more efficient implementations: think of the single-shot OR problem with thousands of inputs, or at design time a priori unknown number of inputs. One would probably favor a small stateful solution that iteratively solves the problem, remembering intermediate results, over a stateless solution that has to grow with the number of inputs as it cannot store intermediate results from a computation.

9.3.1.1 Implementations Without Internal State

A *combinational Boolean circuit* is a circuit that consists only of stateless gates, like AND, OR, and NOT, and that has no feedback loops: signals strictly propagate from the input to the output. Thus, if a Boolean input is applied to a combinational circuit, its outputs will settle at binary values solely determined by the input values and not by the order in which they are applied, or previously applied input values. It is thus considered stateless.

For the previously discussed single-shot problem to compute the OR of n bit strings, a natural solution as a combinational Boolean circuit is a tree of OR gates; see Fig. 9.1.

9.3.1.2 Implementations with Internal State

A deterministic state machine is a set of states S, one of which is the initial state s_0, an input alphabet I, an output alphabet O, a transition function $\delta : S \times I \to S$, and an output function $o : S \to O$. The behavior of this abstract machine in the presence of an input stream $i_1, i_2, i_3, \ldots$ is as follows: starting from the initial state s_0 where it produces output $o(s_0)$, the machine transitions to state $s_1 = \delta(s_0, i_1)$ producing output $o(s_1)$. Next, it transitions to state $s_2 = \delta(s_1, i_2)$ producing output $o(s_2)$, and so on.

Figure 9.2 depicts a state machine that solves both the single-shot OR problem and the reactive OR problem. Its input and output alphabet is $I = O = \{0, 1\}$, its set of states is $S = \{S_0, S_1\}$ with initial state S_0, its transition function is

$$\delta(S_0, i) = \begin{cases} S_0 & \text{if } i = 0 \\ S_1 & \text{if } i = 1 \end{cases} \quad \text{and} \quad \delta(S_1, i) = \begin{cases} S_1 & \text{if } i = 0 \\ S_1 & \text{if } i = 1 \end{cases}$$

and its output function is $o(S_0) = 0$ and $o(S_1) = 1$.

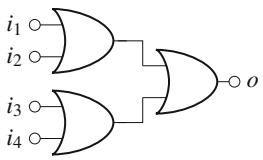


Fig. 9.1 Combinational circuit that solves the OR problem for 4 input bits i_1, i_2, i_3, i_4

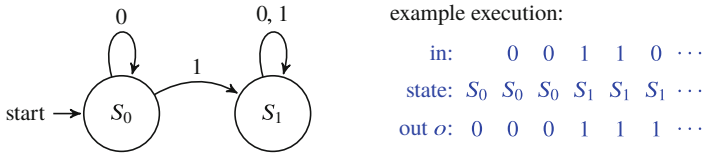


Fig. 9.2 State machine that solves the OR problem. Transitions are depicted as arrows labeled with inputs. Outputs of the states are $o(S_0) = 0$ and $o(S_1) = 1$. An example execution is shown in blue

9.4 Circuits

The combinational circuits and state machines discussed in Sect. 9.3.1 are agnostic to specific target technologies or target machines. In the current section we will discuss more concrete implementations targeted for low-level implementation in silico, that is, in VLSI circuits, and in synthetic biological circuits.

9.4.1 Wires

Conceptually, wires or channels provide means to geometrically route information, to provide separation between different signals, and to translate an input species into an output species. Media and transported species vary greatly: while wires (interconnect) in circuits are made out of metal or polysilicon, insulated from each other, and transport charge (in some cases on-chip fibers transport photons), the situation is far more diverse in microbiological circuits where signals can be metabolites, small molecules, peptides, phages, DNA, or RNA. For all such wires orthogonality of signals, error rates, transport delay, etc., are important questions that need to be addressed.

From a computational point of view, a wire is the simplest computation: a gate that computes the identity, typically with a certain delay. Therefore, in terms of production rules, the specification of a wire can be expressed as:

$$I \to O := 1 \quad \text{and} \quad \neg I \to O := 0 \tag{9.5}$$

When the input is 1, the output becomes 1, and when the input is 0, the output becomes 0.

9.4.1.1 Electrical Wires

The simplest non-trivial model for an electrical wire is an *RC element*. It consists of a resistor and a capacitor in series. This allows the modeling of not only the non-zero resistance of real wires, but also the various loading effects that any real circuit element invariably displays.

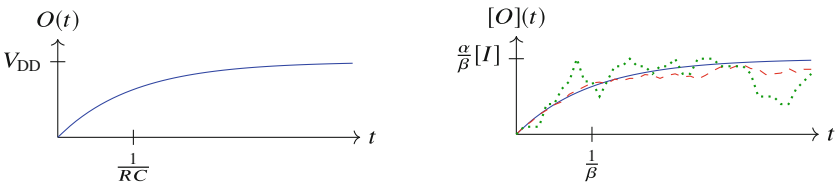


Fig. 9.3 Behavior of electrical (left) and microbiological (right) wires reacting to a constant input signal. The microbiological signal is shown for three settings: stochastic models with $I = 10$ (dotted green) and $I = 100$ (dashed red), and the ordinary differential equation model (solid blue)

Denoting by R the resistance of the resistor and by C the capacitance of the capacitor, the output voltage O of the RC element reacting to an input voltage I is given by the ordinary differential equation

$$\frac{\mathrm{d}O}{\mathrm{d}t} = \frac{I}{RC} - \frac{O}{RC}$$

For example, when starting from an initial voltage of 0 V (= ground, logical 0), the time behavior of an RC wire reacting to a positive input pulse of voltage V_{DD} (= power supply, logical 1) can be calculated as:

$$O(t) = V_{\mathrm{DD}} \cdot \left(1 - e^{-t/\tau}\right)$$

where $\tau = R \cdot C$ is the RC constant. Figure 9.3 (left) depicts this function.

When combining a wire with a Boolean gate, the RC constant of the physical wire is dominated by loading effects inside the gate. In fact, on the scale of gates the RC constant of a physical wire is negligible. In this setting, the derivations from this section are more appropriately applied to an *identity gate*, i.e., a single-input gate whose output is equal to its input.

9.4.1.2 Microbiological Wires

A very simple implementation of a microbiological wire, or identity gate, can be constructed by having the input molecule I be an activator for the gene encoding output molecule O. Likewise considerations, however, hold for other microbiological wires. After adding a decay reaction, which is necessary for the stability of the model, the simple microbiological wire can be described by the two reactions

$$I + S \xrightarrow{\gamma} I + O \quad \text{and} \quad O \xrightarrow{\beta} \emptyset \tag{9.6}$$

where β and γ are some rate constants and S is a substrate assumed to be present in abundance (i.e., its concentration remains constant). By applying mass-action

kinetics to the system of reactions (9.6), we get the following ordinary differential equation for the time behavior of the concentration of the output molecule O:

$$\frac{\mathrm{d}[O]}{\mathrm{d}t} = \alpha[I] - \beta[O]$$

where $\alpha = \gamma \cdot [S]$. An immediate consequence of this equation is that the output is in a steady state only if the input is constant and $[O] = \frac{\alpha}{\beta} \cdot [I]$.

The concentration of output molecules O over time, when given a constant concentration of input molecules I, is equal to

$$[O](t) = \frac{\alpha}{\beta} \cdot [I] \cdot \left(1 - e^{-\beta t}\right) \tag{9.7}$$

when starting at an initial output concentration of zero. Figure 9.3 (right) depicts this function.

Expressed in the CHP language, the reactions (9.6) can be interpreted as the following two production rules:

$$I \to O := 1 \quad \text{and} \quad O \to O := 0$$

Note that this CHP formulation is different from (9.5). In fact, modern electrical implementations directly[2] implement (9.5) by using complementary stacks (explained in Sect. 9.4.2.1), while classical microbiological implementations rely on decay reactions which can lead to a simpler set-up. However, this distinction is not a peculiarity of microbiological settings. In principle, one could also use single-stack implementations in electrical circuits and complementary-stack implementations in microbiological circuits.

9.4.2 Gates

A single output gate is described by two production rules, one that specifies when its output is set to 1, and another that specifies when it is set to 0:

$$P_{\text{up}} \to O := 1 \quad \text{and} \quad P_{\text{down}} \to O := 0 \tag{9.8}$$

Typically P_{up} and P_{down} are required to be mutually exclusive, so that the gate never forces its output to 0 and 1 at the same time. If further P_{up} and P_{down} are negations of each other, that is exactly one of them holds, then the gate is *stateless*. Otherwise, it is *stateful*. An example stateless gate is the OR gate with $P_{\text{up}} = A \vee B$ and $P_{\text{down}} = \neg(A \vee B)$. An example of a stateful gate is the *RS-latch*, a storage element that can

[2] To be precise, an identity gate is usually implemented by two successive inverters.

be set with $S = 1$ and reset with $R = 1$, with $P_{\text{up}} = S \wedge \neg R$ and $P_{\text{down}} = \neg S \wedge R$. Note that, if $R = S = 0$, then the RS-latch holds its previous output; hence, it is stateful.

9.4.2.1 CMOS Gate Implementations

Once the circuit has been rewritten in the form of production rules, a standard complementary metal-oxide-semiconductor (CMOS) implementation using transistors is readily obtainable. Consider the two rules in (9.8) and assume that they represent a stateless gate. For an efficient CMOS implementation, we also require that P_{up} consists only of negated variables and their combination via AND and OR, e.g., $\neg A \wedge \neg B$. Likewise, P_{down} must consist only of positive variables and their combination via AND and OR, e.g., $A \vee B$. Then, rule $P_{\text{up}} \to O := 1$ is implemented by a *stack of p-type transistors* (one transistor per negated variable) that are responsible to pull the gate's output to V_{DD} (= logical 1) in case P_{up} is true. Rule $P_{\text{down}} \to O := 0$ is implemented by a *stack of n-type transistors* (one transistor per positive variable) that pull the gate's output towards the ground (= logical 0) if P_{down} is true. Figure 9.4 shows a CMOS implementation for an inverter, i.e., a gate with $P_{\text{up}} = \neg A$ and $P_{\text{down}} = A$. The p-stack (green) is the above transistor that connects V_{DD} with O if $A = 0$. The n-stack (magenta) is the bottom transistor that connects ground with O if $A = 1$.

Figure 9.5 shows the two cases for input A: If $A = 0$, the p-stack connects and the n-stack is open. Thus the load at O is charged and pulled up to V_{DD} (= logical 1) via the connecting p-stack. Conversely, if $A = 1$, the n-stack connects and the p-stack is open, thus discharging the load at O through the n-stack to ground (= logical 0).

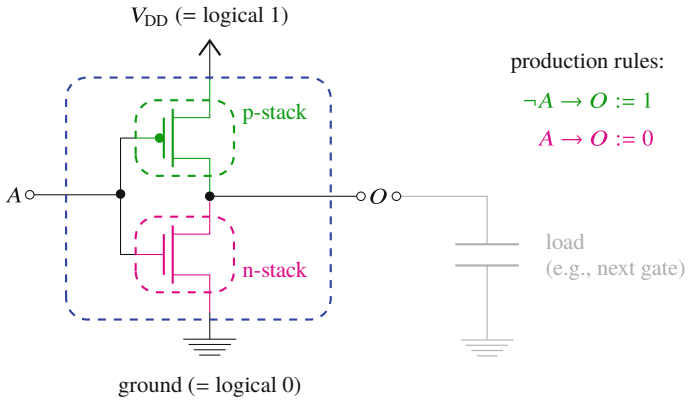


Fig. 9.4 CMOS implementation of an inverter (left) driving the next gate (right) represented by a capacitive load. Note that the capacitive load is a model for the inertia of the next circuit and is not deliberately built into the circuit

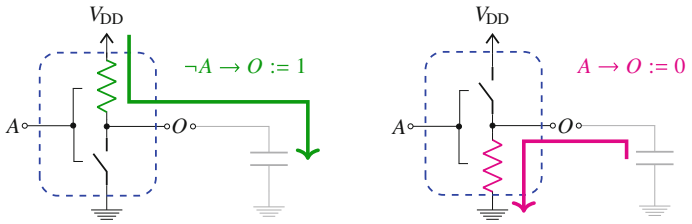


Fig. 9.5 Case $A = 0$: the load at output O is charged (left). Case $A = 1$: the load at output O is discharged (right)

9.4.2.2 Microbiological Gate Implementations

Figure 9.5 shows that the inverter gate does not use charge from input A to charge the output O (so as not to reduce the input charge). Instead, the charge comes from an independent and undepletable source, the power supply V_{DD}.

This concept of decoupling input and output is also observed in naturally occurring reactions and synthetic biological designs. While in the enzymatic reaction

$$A + E \rightleftharpoons AE \to E + O$$

the input molecule A is consumed to create output molecule O via enzyme E, i.e., there is no decoupling, in a reaction like

$$A + E \rightleftharpoons AE \overset{S}{\hookrightarrow} AE + O$$

with a sufficiently available source molecule S, playing the role of V_{DD}, provides decoupling analogous to the CMOS inverter. It resembles the charging via the p-stack in Fig. 9.5 (left) in that it is enabled if A is present and disabled if A is absent, analogously but negated to its silicon counterpart. In terms of production rules, we may express its behavior as $A \to O := 1$. Typically, the resetting of O is not actively driven, but effected implicitly by the decay of O. Generalizations of this scheme to two input species have, e.g., been used to build AND gates (Schaerli et al. 2014, Fig. 1D–F). Inputs are two complementary parts of the phage T7 RNA polymerase, E is a promoter, and S is the pool of nucleotides to transcribe into O.

By contrast, the scheme

$$\begin{array}{ccc} & \overset{S}{} & \\ E & \hookrightarrow & E + O \\ A & \leftrightarrows & \\ & EA & \end{array}$$

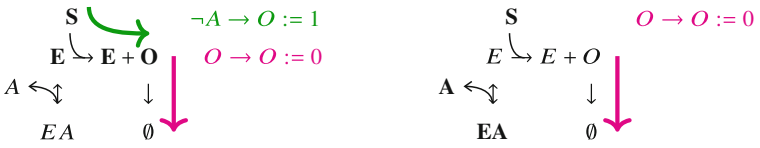


Fig. 9.6 Case $A = 0$: output molecule O is produced (left). Case $A = 1$: output molecule O is decayed if present (right). Molecules with high concentrations are printed in bold

is decoupled and negating, analogous to the p-stack. In terms of production rules, we may write $\neg A \to O := 1$. Again, resetting of O typically is by decay of O. Figure 9.6 shows the reactions of an inverter with "charging" (in green) and "discharging" (in magenta) reactions corresponding to Fig. 9.5. Generalizations of this to two input species have been used to implement NOR gates (Brophy and Voigt 2014, Fig. 2c). It works by repressing expression of O via the CRISPR/dCas9 machinery. Inputs are guide RNAs, E is a promoter, and S is the pool of nucleotides and dCas9.

9.4.3 Stateful Circuits

In this section, we turn our attention to non-combinational circuits, i.e., circuits that contain some internal state. We do this by discussing implementations of the RS-latch with CMOS circuits and microbiological circuits.

The capacity to store bits, as is possible with the RS-latch, is prototypical for stateful circuits. In fact, the latch is used as a component in the so-called flip-flop circuits that are widely used to store states in CMOS circuits. For example, the reactive OR problem from Sect. 9.3.1.2 can be implemented with an RS-latch.

9.4.3.1 CMOS Circuits

Most CMOS circuits possess some internal state. The RS-latch in CMOS circuits is most commonly implemented with two NOR gates, as shown in Fig. 9.7.

The implementation's main idea is to translate the RS-latch representation in the CHP language

$$*[[[S = 1 \wedge R = 0]; Q := 1]] \,||\, *[[[S = 0 \wedge R = 1]; Q := 0]]$$

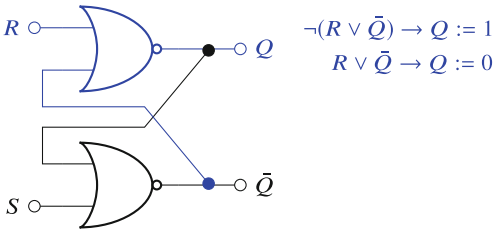


Fig. 9.7 Implementation of RS-latch with two NOR gates. The gate driving Q is shown in blue and its production rules are stated

syntactically into the production rules

$$(S \wedge \neg R) \rightarrow Q := 1 \quad \text{and} \quad (\neg S \wedge R) \rightarrow Q := 0$$

and then to introduce the variable $\bar{Q} = \neg Q$ that allows rewriting the production rules into:

$$\neg(S \vee Q) \rightarrow \bar{Q} := 1 \quad \text{and} \quad (S \vee Q) \rightarrow \bar{Q} := 0$$
$$\neg(R \vee \bar{Q}) \rightarrow Q := 1 \quad \text{and} \quad (R \vee \bar{Q}) \rightarrow Q := 0$$

Given this CHP representation, the possible implementation with two NOR gates is clearly visible.

9.4.3.2 Microbiological Circuits

In synthetic biology, circuits with internal state are less common than in CMOS circuits. While there is a considerable number of microbiological circuits without an internal state (Nielsen et al. 2016), one of the first microbiological circuits with an internal state was the bacterial toggle switch (Gardner et al. 2000). Recently, theoretical microbiological asynchronous circuit designs with internal states have appeared (Nguyen et al. 2019). Figure 9.8 (left) shows the implementation of the toggle switch as a genetic circuit. Interestingly, due to different implementation constraints and the implementation's single-stacked nature, the implementation of the toggle switch does not decompose into two NOR gates, as is the case for the CMOS implementation. We provided the corresponding CHP production rules in Fig. 9.8 (right).

9.4.4 Challenges

In previous sub-sections we discussed an idealized view on CMOS and microbiological circuits. In the following, we point to the limitations of such a view: circuit

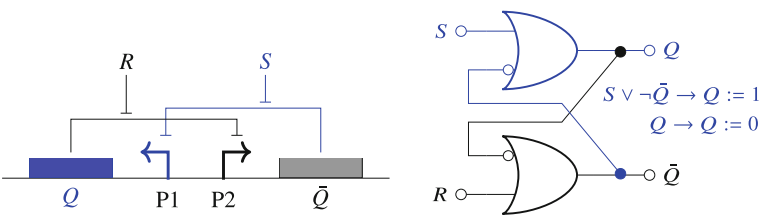


Fig. 9.8 Toggle switch design from Gardner et al. (2000). The components driving Q are shown in blue (left). Corresponding production rules and gate-level description (right). By shifting the input negation to the output observe that signal $\neg\bar{Q}$ in the microbiological circuit corresponds to Q in the circuit in Fig. 9.7 and $\neg Q$ to $\bar{Q}$. Since $Q = \neg\bar{Q}$ during normal operation (R and S are not issued at the same time), the CMOS and the microbiological circuit behave equivalently

components may fail, both at production and mission time, and apparently simple operations like setting an output to 1 are, in fact, continuous processes with a range of implications.

9.4.4.1 CMOS Circuit Implementations

The typical production faults that may cause a permanent unintended behavior of the circuit are open and bridge faults at the transistor level (Ferguson and Shen 1988). In an *open fault* a connection is interrupted, while in a *bridge fault* two points are connected via an unintended (low) resistance. Possible reasons are friction, electrical wear-out, impurities during fabrication, etc., and the effects are typically permanent once they occur. Open and bridge faults within a gate may have different effects on the gate output that depend on the gate, the resistance of the bridge, etc. For simplicity, typically two types of behaviors are assumed, which the circuit is tested for: in a *stuck-at*-0 or *stuck-at*-1 fault the gate output is simply stuck to the logical value of 0 or 1. An example for a stuck-at-1 fault is a low resistance bridge from O to V_{DD} in Fig. 9.4. The second common manifestation at gate-level is the *delay-fault* where logical gate behavior is as expected, but with pronounced delays under certain inputs.

By contrast, a transient fault is a temporary misbehavior of the circuit. In this class falls sudden deposition of charge caused by an ionizing particle hit. Effects are short pulses created within the logic and bit-flips within the memory (Ferlet-Cavrois et al. 2013). Related to transient faults is signal noise, e.g., induced by environmental fluctuations (e.g., temperature) and voltage drops in the power line due to simultaneous switching activity of numerous gates in a circuit. Such noise is an issue in aggressively timed or low-power CMOS circuits.

Unrelated to the above phenomena is metastability that may occur in a circuit in the absence of any external faults. Consider a circuit that has an internal state whose value can be set to at least two different values, say 0 and 1, e.g., an RS-latch. If there are two scenarios, one where the environment can drive the circuit into state 0 and one where it can drive it into state 1, and the scenarios can be continuously transformed into each other, then there also exists a scenario in which the state is in-between where it resides for an arbitrarily long period (Marino 1981), the *metastable state*. This is problematic as a circuit in this state may produce in-between voltage outputs that are neither logical 0 nor logical 1. These can corrupt downstream gates that expect clear voltage inputs.

9.4.4.2 Microbiological Circuit Implementations

Several of the aforementioned faults have corresponding faults in synthetic biological implementations. Brophy and Voigt (2014) provided a survey of common permanent faults that have been observed in synthetic biology designs using the example of an AND gate and an oscillator. A central source of the faults described were mismatched components (e.g., dynamic range and offset mismatches) that essentially lead to a loss of the digital abstraction. In the AND gate the faults led to dynamic range and offset mismatches at its output. While these range mismatches do not play a prominent role in modern CMOS designs, inaccessible ribosome

binding sites (RBS) because of RBS context may be viewed as open faults and lead to stuck-at-0 faults at the gate output. Stuck-at-0 faults were also observed after unintended recombination had taken place in the synthetic circuit. Missing orthogonality between (repressor) signals and corresponding promoters as well as transcriptional read-through because of insufficient insulation by terminators lead to faults that resembled bridge faults in CMOS circuits. Although the gate's behavior was similar to that under bridge fault, we would like to stress an important difference between the biological bridge faults and those in silico: their directionality. While bridge faults in CMOS are modeled as bidirectional, high resistance connections, their microbiological counterparts (missing orthogonality and read-through) a priori are often directional.

In addition to permanent faults, noise plays a pronounced role in both silicon and biological computers. Although signal noise is an issue in low-power CMOS circuits with reduced dynamic range, the situation is intensified in biological circuits; e.g., see the stochastic behavior for $I = 10$ and $I = 100$ input molecules in Fig. 9.3 where unintended short-0 pulses are to be expected. While such pulses may lead to unintended, but time-limited, outputs in stateless circuits, their presence at the input of stateful gates like the toggle switch may lead to a permanent failure of the circuit.

9.5 Conclusions

In this chapter, we presented design techniques for digital circuits, both for silicon and biological computers. We saw that, while many high-level notions, like the abstract concept of Boolean gates and their representation, are identical in both worlds, distinctions have to be made when descending from the abstraction levels towards functional implementations. This distinction is already visible in the Boolean representation domain, as we saw that different implementations require different production rules in the CHP language.

What we have presented in this chapter is only a small part of the vast space of circuit design, both for electrical and biological circuits. But, we have tried to present a relevant selection of similarities and prototypical differences between the two. Clearly, there is a rich interplay possible between the fields of CMOS circuit design and synthetic biology circuit design, though care must be taken to not blindly transfer notions or complete designs from one to the other.

References

Agarwal A, Lang JH (2005) Foundations of analog and digital electronic circuits. Morgan Kaufmann, San Francisco

Brophy JAN, Voigt CA (2014) Principles of genetic circuit design. Nat Methods 11(5):508–520

Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O'Neil J, Little CTS (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–64.

Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S et al (2017) Communication between viruses guides lysis–lysogeny decisions. Nature 541:488–493

Ferguson FJ, Shen JP (1988) A CMOS fault extractor for inductive fault analysis. IEEE Trans Comput Aided Des Integr Circ Syst 7(11):1181–1194

Ferlet-Cavrois V, Massengill LW, Gouker P (2013) Single event transients in digital CMOS: a review. IEEE Trans Nucl Sci 60(3):1767–1790

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403:339–342

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361

Hoare CAR (1978) Communicating sequential processes. In: The origin of concurrent programming. Springer, Heidelberg, pp 413–443

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19(4):524–531

Katz E (2017) Enzyme-based logic gates and networks with output signals analyzed by various methods. ChemPhysChem 18(13):1688–1713

Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11(5):367–379

Kiel C, Yus E, Serrano L (2010) Engineering signal transduction pathways. Cell 140(1):33–47

Macklem PT, Seely A (2010) Towards a definition of life. Perspect Biol Med 53(3):330–340

Marino LR (1981) General theory of metastable operation. IEEE Trans Comput C-30(2):107–115

Martin AJ (1989) Programming in VLSI: from communicating processes to delay-insensitive circuits. Technical Report Caltech-CS-TR-89-1, California Institute of Technology

McLaughlin JA, Pocock M, Mısırlı G, Madsen C, Wipat A (2016) VisBOL: web-based tools for synthetic biology design visualization. ACS Synth Biol 5(8):874–876

Monod J (1972) Chance and necessity: an essay on the nature of philosophy of modern biology. Collins, London

Monod J, Jacob F (1961) General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. In: Cold spring harbor symposia on quantitative biology, vol 26. Cold Spring Harbor Laboratory Press, New York, pp 389–401

Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. Nature 491:249–253

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biol 9(8):e1001127

Nguyen T, Jones TS, Fontanarrosa P, Mante JV, Zundel Z, Densmore D, Myers CJ (2019) Design of asynchronous genetic circuits. Proc IEEE 107(7):1356–1368

Nicholson DJ (2019) Is the cell really a machine? J Theor Biol 477:108–126

Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA (2016) Genetic circuit design automation. Science 352(6281):aac7341

Oliva G, Sahr T, Buchrieser C (2018) The life cycle of L. pneumophila: cellular differentiation is linked to virulence and metabolism. Front Cell Infect Microbiol 8:3

Pandi A, Koch M, Voyvodic PL, Soudier P, Bonnet J, Kushwaha M, Faulon J-L (2019) Metabolic perceptrons for neural computing in biological systems. Nat Commun 10:3880

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386

Sauro HM, Kim K (2013) Synthetic biology: it's an analog world. Nature 497:572

Schaerli Y, Gili M, Isalan M (2014) A split intein T7 RNA polymerase for transcriptional AND-logic. Nucleic Acids Res 42(19):12322–12328

Selberg J, Gomez M, Rolandi M (2018) The potential for convergence between synthetic biology and bioelectronics. Cell Syst 7(3):213–244

Smith BC (2002) The foundations of computing. In: Scheutz M (ed) Computationalism: new directions. MIT Press, Cambridge, pp 23–58

Sourjik V, Wingreen NS (2012) Responding to chemical gradients: bacterial chemotaxis. Curr Opin Cell Biol 24(2):262–268

Spirov A, Fahmy K, Schneider M, Frei E, Noll M, Baumgartner S (2009) Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. Development 136(4):605–614

Teo JJY, Woo SS, Sarpeshkar R (2015) Synthetic biology: a unifying view and review using analog circuits. IEEE Trans Biomed Circ Syst 9(4):453–474

Engineering of Riboregulators for Gene Regulation as a Tool for Synthetic Biology

10

Kinuko Ueno, Kaori Tsukakoshi, and Kazunori Ikebukuro

Abstract

Riboregulators are an RNA tool that can regulate target gene expression by blocking specific sites of the target gene using small RNA fragments. This nature-inspired RNA tool was artificially designed and optimized for specific and highly efficient gene regulation; the modularity and usability of riboregulators have been expanded by application for various cell and bacterial metabolic engineering methods and for the detection of target molecules. Moreover, riboregulators have multiple features such as design feasibility and adjustability of induction level.

In this chapter, we introduce the previous efforts to develop and improve riboregulators. We also focus on several upcoming methods of gene regulation that can benefit from riboregulators.

Keywords

Riboregulator · RNA tool · Gene regulation · Synthetic biology

10.1 Introduction

In the wide field of synthetic biology, target gene regulation is essential for designing and modulating cells. It is important to regulate specific genes because it offers various benefits such as tracking the intercellular phenomena and obtaining a high amount of desired biomass at the optimal time. For decades, a variety of gene regulation methods have been developed in both prokaryotes and eukaryotes. Some of these methods employ various promoters and small RNA fragments for RNA

K. Ueno · K. Tsukakoshi · K. Ikebukuro (✉)
Department of Biotechnology and Life Science, Graduate School of Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan
e-mail: ikebu@cc.tuat.ac.jp

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_10

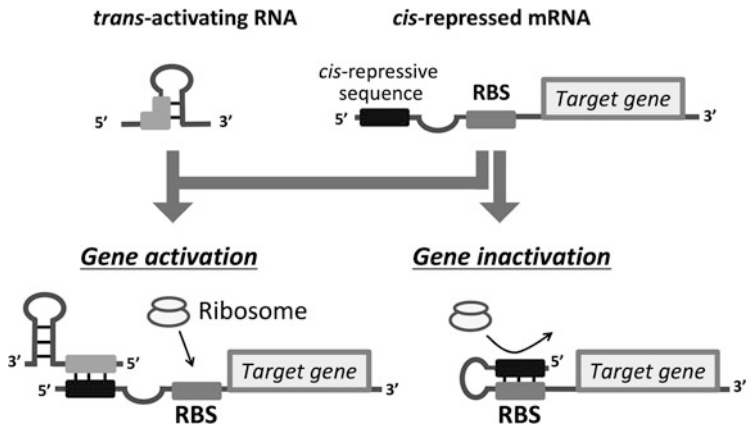


Fig. 10.1 Scheme of riboregulator. Cis-repressed mRNA (crRNA) can form a double-stranded RNA structure to cover the ribosomal binding site (RBS). On the other hand, trans-activating RNA can bind crRNA so that the ribosome can translate the target gene mRNA

interference, or a pair of CRISPR and single-guided RNA to inhibit transcription. Similar to these unique gene regulation tools, riboregulators have been developed and improved as an RNA gene regulation tool.

Riboregulators are an RNA tool, composed of two RNA fragments containing trans-activating RNA (taRNA) and cis-repressed RNA (crRNA) (Isaacs et al. 2004). The sequence of taRNA is complementary to part of the crRNA such that the taRNA and crRNA form an intermolecular hybrid to make the ribosome-binding site (RBS) of the target mRNA available (Fig. 10.1). If taRNA transcription is not induced, the crRNA can form an intramolecular hybrid between the crRNA and ribosome-binding site, which represses the target gene translation. This nature-inspired gene regulation tool was first designed in *Escherichia coli*, and its sequence and application have been subsequently expanded to other cells.

In this chapter, we focus on artificially designed riboregulators and track the development and improvement of this RNA tool from the first report of this RNA tool to recent efforts applying it to biomass production. We would like to elucidate the usability and future possibility of riboregulator usage by presenting an overview of the progress in engineering riboregulators for gene regulation.

10.2 Development of Riboregulator

Before the development of artificial riboregulators, the term riboregulator had another meaning indicating the natural gene regulation system composed of RNA fragments. First, the possibility that the 3′ untranslated region of messenger RNA might have a role in the growth or immortalization of human cells was considered (Janssen and Mier 1997; Jupe et al. 1996a; Jupe et al. 1996b; Gultyaev et al.

1997). Following these reports, the detailed sequence and function of DsrA, the untranslated RNA fragments, were revealed in *E. coli*, indicating that sophisticated gene regulation by untranslated RNA fragments was performed in the cell (Lease et al. 1998). After this discovery, various gene regulation tools based on the untranslated RNA fragments were reported as natural riboregulators in several species of bacteria, archaea, and eukaryotes (Horos et al. 2019; Bohn et al. 2010; Papenfort et al. 2008; Breaker et al. 2002). These natural riboregulators regulate translation by forming a specific structure so that they can inhibit ribosome binding.

Inspired by the riboregulators, an artificial riboregulator was first developed in *E. coli* to regulate a specific target gene (Isaacs et al. 2004). Collins and coworkers designed a taRNA with a 19-nucleotide complementary sequence of the RBS and four candidate sequences of 25-nt crRNA, which can bind to taRNA. In this design, the crRNA sequence was designed to fulfill the following conditions: (1) the crRNA does not alter the coding frame of the target gene, (2) it can form a stem-loop structure, and (3) it has the YUNR motif (pYrimidine-Uracil-Nucleotide-puRine) consensus sequence. These criteria were applied to design the crRNA such that its structure can be flexible enough to bind taRNA in the cell. In these results, one of the candidate crRNA sequence showed a 19-fold increase of GFP expression in *E. coli*. Further, this study successfully showed the specificity of taRNA and crRNA binding by evaluation of several cognate RNA pairs. It was, thus, shown that the artificially designed riboregulator can be functionalized in the cell. This strategy for posttranscriptional regulation has been improved by sequence modification and has been applied to various purposes such as biomass production, gene circuit, and tracking intracellular conditions, and we would like describe further applications of this strategy later.

As another type of riboregulator, Yin and coworkers developed Toehold switches that showed high dynamic range and specificity (Green et al. 2014). The toehold switch consists of switch RNA and trigger RNA as depicted in Fig. 10.2. The strategy for translational regulation was similar to the riboregulator system that is mentioned above, but this toehold switch has two different points from the first reported riboregulator. First, this toehold switch can cover both the RBS and start codon, which enabled us to maximize the repression of translation. Second, one of the RNA fragments of the toehold switch, trigger RNA, does not need to contain the RBS sequence so that the variation in the design of trigger RNA can be expanded. In the design of toehold switch, the sequence of the RBS and start codon was completely unpaired with the trigger RNA such that translation was induced by linear–linear RNA interactions between the trigger RNA and the 5′-end of mRNA. It should be noted that the trigger RNA has to contain the complementary sequence of the start codon, but this design resulted in an increase of the inducing level and resulted in high expression of the target gene under an induced state. A riboregulatory system composed of a trigger RNA without RBS and a start codon

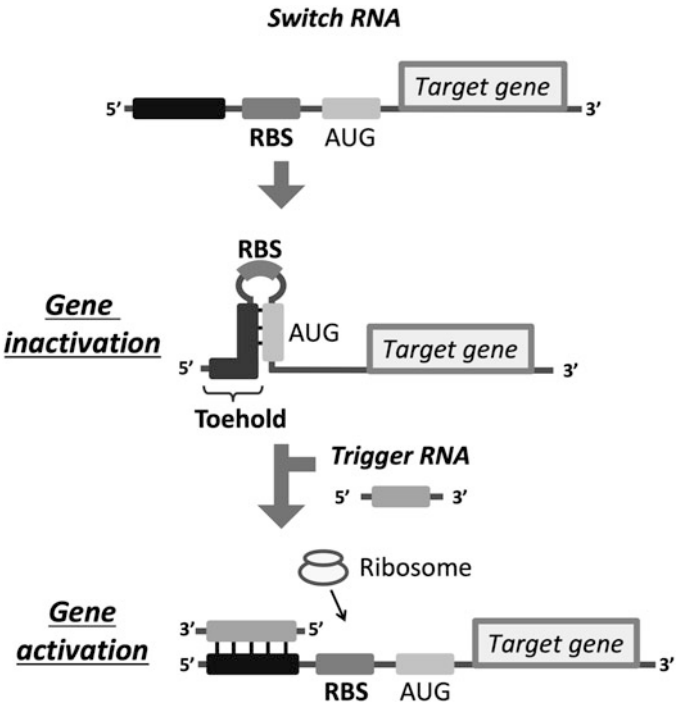


Fig. 10.2 Scheme of toehold switch. Transcribed switch RNA can form a double-stranded RNA structure that can cover both the RBS and start codon. This structure can be distorted by the binding of trigger RNA to the toehold region and can induce target gene expression

sequence has also been applied to various detection methods and cellular component tracking, whose details are mentioned later.

10.3 Improvement of Riboregulator

10.3.1 Stability

One of the limitations of riboregulators is that their induction level is relatively lower than that of protein-based transcriptional regulation (Lutz and Bujard 1997). For example, one of the most widely used promoter, the Tetracycline promoter system, showed ~350-fold ON/OFF ratio with high protein expression level, whereas the original riboregulator showed up to 13-fold.

To overcome this issue, it is an effective strategy to stabilize taRNA fragments in order to maximize its induction level. Previously, Yu and coworkers used the Hfq-binding sequence to improve the function of the RNA-based gene regulation tool (Man et al. 2011). Hfq is one of the RNA chaperone proteins that binds to adenine-

uracil-rich scaffolds so that the complex can modulate the stability of RNA-RNA interaction (Brennan and Link 2007; Sun et al. 2002). Further, it was assumed that Hfq has several roles in inducing intramolecular hybridization between sRNAs and target mRNA (Lease and Woodson 2004). Based on these roles, the Hfq-binding sequence was examined and it was then used to stabilize small RNAs in *E. coli* (Sharma et al. 2012).

To improve the stability of taRNA, our research group focused on the Hfq-binding sequence and applied to stabilize the taRNA fragments in the cell (Sakai et al. 2014). In this report, several small RNA scaffolds from DsrA, GcvB, MicF, and Spot42 were selected and fused with taRNA. The Mif-F scaffold-fused taRNA showed the highest On/OFF ratio, and further mutation of this scaffold-fused taRNA resulted in a high level of target gene induction.

10.3.2 Gene Circuit

As riboregulators are one of the posttranscriptional regulation systems, they can be combined with other transcriptional regulation processes to modulate convoluted regulation such as a gene circuit and counter system. The large benefit of designing a regulation network using multiple response regulators, promoters, and RNA tools is that this regulation system can be applied to complicated digital circuits and design of cell behavior. To obtain a new strategy for cell control, there are several reports that combine the riboregulator with multiple gene regulation tools.

The first effort to apply a riboregulator to a gene circuit was performed by employing a riboregulator for a Riboregulated Transcriptional Cascade by combining multiple RNA polymerases, promoters, and a riboregulator (Cabello et al. 2010). Collins and coworkers designed a riboregulator that is controlled under taRNA-located downstream of the arabinose promoter. As this riboregulator sequence has the T7 or T3 RNA polymerase-coding region, induction of the riboregulator results in GFP gene expression under the T7 or T3 promoter. As a result, the fluorescence intensity increases with a corresponding number of arabinose addition to the cell environment. Other strategy for designing a gene circuit using RNA tools indicates the slow response speed of protein-based gene regulation networks (Lucks et al. 2011); however, this regulation network system showed a rapid response toward the arabinose pulse within 1 h, which implies that a combination of RNA gene regulation tools and a strong protein-based expression system can be applied to the gene network design with a quick response.

10.3.3 Redesigning Riboregulators by Computational Modeling

The sequence of the cognate RNA fragments pair, trRNA and crRNA, is the essential factor of gene regulation by riboregulators. Although there has been an effort to analyze the binding ability of taRNA toward crRNA in vitro (Senoussi et al. 2018), there are technical hurdles to analyze the behavior of two RNA fragments in the

cell. To remove this difficulty and to design an optimized sequence, several research groups have employed computational calculation to predict the structure of RNA fragments to achieve a riboregulator sequence that can show a high ON/OFF ratio, induction level, and specificity. For example, Jaramillo and coworkers first reported the de novo designed riboregulator sequence by screening the best combination of modified natural small RNAs and their pair sequence (Rodrigo et al. 2012). The authors employed two parameters to classify the previously reported small RNA structure: (1) free energy of complex formation and (2) activation energy of complex formation. The authors also hypothesized that the RNA-RNA interaction can occur after an individual folding state, followed by a transition state in which the two RNA fragments bind partially. In this report, the authors imposed the previously reported natural small RNAs and artificial small RNAs, and calculated the parameters to explore the candidates that can modulate stable hybridization of the RNA-RNA interaction without covering the RBS. As a result, the algorithm revealed six candidates from 1040 sequences. It was further clarified that these candidate sequences functioned as riboregulators in vivo. Notably, the designed taRNA and crRNA showed high orthogonality. Though the detailed mechanism and structure of these obtained taRNA and crRNA cannot be revealed in this report, the automated methodology for designing regulatory RNA fragments can be one of the solutions to obtain an optimized riboregulatory sequence, and further efforts to explore the candidates of de novo design are underway (Peters et al. 2018).

10.3.4 Ligand-Inducible Riboregulator

The riboregulator itself can be applied for various gene regulation tools through a simple strategy, as natural occurring or the originally developed riboregulator was not capable of responding to specific ligands, such as small molecules or proteins. These gene regulation methods that can respond to ligands can be represented by a riboswitch or two-component system of ligand-inducible promoters.

To fulfill the needs of ligand-inducible riboregulator, Smolke and Bayer developed antiswitches, which utilize RNA fragments containing an aptamer sequence and a cap-region covering sequence (Bayer and Smolke 2005). As this gene regulation tool was developed in the *Saccharomyces cerevisiae*, the antiswitches can inhibit target gene expression by covering the cap-region of mRNA (Fig. 10.3). Additionally, the authors fused the theophylline or tetracycline aptamer to the antiswitches, so that this RNA fragment can change its structure to bind the cap region in the presence of the target ligand, resulting in target gene repression. As a result, this antiswitch effectively repressed the target protein expression after ligand addition within 3–4 h. The sequence design itself differed from other riboregulators developed in the bacterial cell; however, this report showed the possibility that fusion of an aptamer sequence with regulatory RNA fragments can expand the capability of the riboregulator system.

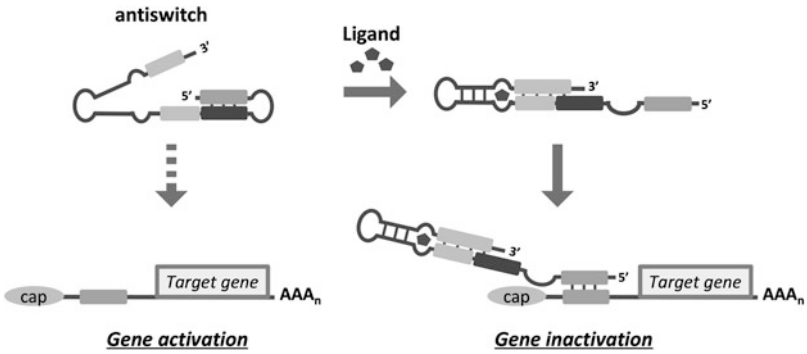


Fig. 10.3 Scheme of ligand-inducible riboregulator in mammalian cell. The antiswitch can change its structure in the presence of a ligand so that this RNA fragment can bind upstream of the target gene to repress gene expression

10.4 Applying Riboregulators to Gene Regulation in Cyanobacteria

Cyanobacteria are one of the photosynthetic organisms that can convert carbon dioxide to several high-value compounds including oil lipids and chemical agents. Cyanobacteria are considered an effective and economical host because they need only sunlight, water, and a few nutrients to grow, and their growth rate is fast enough to be considered for industrial applications. In addition, cyanobacteria are amenable to gene modification, which enables designing a cyanobacterial strain that can produce specific target compounds by introducing gene regulation tools (Mazard et al. 2016; Khan et al. 2018). However, gene regulation tools that can be used in cyanobacteria are still limited because the genetic tools developed in *E. coli* have different ribosomal binding site sequences and promoters in cyanobacteria, which could affect their function in cyanobacteria (Huang et al. 2010; Camsund and Lindblad 2014).

To overcome the limitation of gene regulation tools in cyanobacteria, our research group optimized the riboregulator sequence to the RBS sequence, so it could function in cyanobacteria (Abe et al. 2014). We designed several riboregulator sequences and evaluated their function in cyanobacteria and *E. coli*. We redesigned the riboregulator based on the riboregulator sequences crR12 and taR12, which were reported by Collins and co-workers (Isaacs et al. 2004). In detail, we replaced the RBS and chose a candidate sequence that was predicted as the most stable secondary structure by the mfold web server (Zuker 2003). We also optimized the crRNA sequences by inserting mismatches into the stem region in order to enhance hybridization and minimize the effect of RNase III after hybridization of crRNA and taRNA (Hjalt et al. 1995). After the functionality of the designed riboregulators was first evaluated in *E. coli*, we evaluated the candidate sequences in cyanobacteria.

In a cyanobacteria strain, Synechocystis sp. PCC 6803, transcription of crRNA and taRNA was regulated by the P_{trc} promoter without the lacO sequence and a nickel-inducible promoter. As a result, the redesigned riboregulator successfully controlled the reporter gene and showed an ON/OFF ratio of 13, which is at the same level as the original riboregulator in *E. coli*.

Though optimization of the riboregulator sequence for cyanobacteria did not result in an improvement of the ON-OFF ratio for this RNA tool, we succeeded in improving the fold-induction of the redesigned riboregulator by inserting a spacer region on crRNA to strengthen the taRNA-crRNA intermolecular hybridization and to loosen the crRNA intramolecular hybridization (Sakamoto et al. 2018). This modification resulted in a 78-fold higher induction of the riboregulator. It was assumed that that this sequence optimization resulted in repression of the noninduced condition and resulted in a high ON-OFF ratio because the induced fluorescence level was not enhanced. Though the induction level of the riboregulator for cyanobacteria still needs improvement, our research shows the possibility of using riboregulators for gene regulation in other bacterial species. Later in this chapter, we describe further applications of this optimized riboregulator for biomass production in cyanobacteria.

10.5 Biomass Production Using a Riboregulator

10.5.1 Succinic Acid Production in *E. coli*

Gibson assembly is a DNA assembly method which allows for the joining of mul-As part of the metabolic engineering field, producing valuable compounds using bacteria is attractive for industrial progress. Recent advances in synthetic biology have enhanced the development of biomass production strategies; however, maximizing the yield of desired products from hosts is still a challenge. To optimize the cellular conditions for target compound production, it is important to consider the metabolic flux and control the concerned genes at an optimal time.

To produce succinic acid using bacteria, it is important to control the expression level of several key enzymes. Succinic acid is one of the valuable metabolites that is widely used for plastic products and has received attraction as a bio-based chemical. Using the production pathway of succinic acid in *E. coli*, it can be produced by carboxylation of phosphoenolpyruvate (PEP) to oxaloacetate, followed by conversion toward malate, fumarate, and succinic acid. To boost the productivity of succinic acid, overexpression of *ecaA* and *pepc* genes, which encode carbonic anhydrase and PEP carboxylase, was shown as an effective strategy (Liu et al. 2013). However, it was also reported that excessive expression of these enzymes in the lag phase can lead to lower growth rate (Wang et al. 2015).

On the contrary, expression of the *mgtC* gene from *Salmonella* in *E. coli* can improve cell growth, whereas overgrowth of cells possibly affects the target biomass production. To maximize the total productivity of succinic acid without affecting cell growth, the induction timing of these enzymes is important.

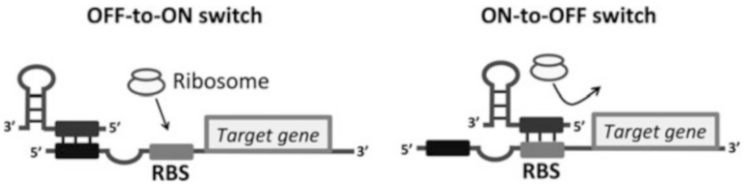


Fig. 10.4 Design of OFF-to-ON switch and ON-to-OFF switch using riboregulators. The complementary sequence of taRNA and crRNA is shown as the black region. The OFF-to-ON switch can induce expression in the presence of taRNA, whereas the ON-to-OFF switch can repress the target gene using taRNA

To develop a novel method for succinic acid production in *E. coli*, Li and coworkers employed a riboregulator as a gene regulation tool in metabolic engineering (Wang et al. 2018). First, the authors developed two types of riboregulatory sequences, an OFF-to-ON switch and an ON-to-OFF switch to control multiple genes (Fig. 10.4). For the development of an OFF-to-ON switch, the authors designed six crRNA candidates that can partially hide RBS sequences up to 6 nucleotides. Further, corresponding taRNA candidates were designed, and both sequences were introduced upstream of the kanamycin resistance gene so that the activation or repression level of the target gene can be determined by the growth of *E. coli* transformants using kanamycin as the sensor. In contrast, the authors designed an ON-to-OFF switch by employing the trRNA sequence that can totally cover the RBS sequence. This repressive taRNA was further fused with an MicF scaffold for stabilization. In the next step to apply the proposed RNA tool for metabolic engineering, the authors used the OFF-to-ON switch to regulate the *ecaA* and *pepc* genes, and the ON-to-OFF switch for regulating the *mgtC* gene, which should be repressed in log phase. The taRNAs for regulating *ecaA* and *pepc* had the same sequence, and both taRNAs were under the control of the IPTG promoter. Thus, after IPTG addition at 4 h of cultivation, succinic acid accumulation was increased without an effect on cell growth. Further, this riboregulator-mediated fermentation production of succinic acid was carried out in a 3-L bioreactor and resulted in 3.25 g L^{-1} h^{-1} productivity. In their previous study, excessive expression of *mgt*B and *mgt*C genes in the one of the *E. coli* strains, CBMG111, showed a productivity of 2.15 g L^{-1} h^{-1}. Although several parameters, such as *E. coli* strain and fermentation media, were not identical to other reports on the improvement of succinic acid production (Okino et al. 2008; Wang et al. 2015; Wang et al. 2016), it was clearly shown that the productivity of target products was significantly increased by multiple gene regulation via the dual-riboregulator system.

10.5.2 Glycogen Production in Cyanobacteria

To expand the availability of RNA-mediated gene regulation tools for metabolic engineering in cyanobacteria, it is necessary to regulate chromosomal genes.

Although homologous gene recombination is widely used for metabolic engineering in cyanobacteria, this classic method cannot be used for essential genes. By applying RNA tools to chromosomal gene regulation, both repression and induction of target genes could be modulated at an optional timing. Metabolic engineering in cyanobacteria can be facilitated using an RNA tool to knock down chromosomal genes, as genes related to metabolic pathways tend to be essential for cell growth.

As a new chromosomal gene regulation method, we applied a riboregulator for knockdown in cyanobacteria (Ueno et al. 2017). One of the remarkable features of a riboregulator is the specificity and strength of target gene repression by crRNA. Because the crRNA sequence is complementary to the RBS sequence, crRNA can immediately hybridize and block the RBS site located adjacent to the crRNA. This scheme enables this RNA tool to realize a high repression level. Further, it is possible to use a riboregulator as a strong repression tool of chromosomal genes. Focusing on this feature, we used the riboregulator to knock down target genes in cyanobacteria.

To apply riboregulators as a knockdown tool in cyanobacteria, we inserted the riboregulator sequence into a chromosomal gene of cyanobacteria (Fig. 10.5). As the target gene, we chose one of the global transcription regulators, *cyabrB2*, which also regulates the accumulation of glycogen granules (Yamauchi et al. 2011; Kaniya et al. 2013). In this report, we controlled the transcription of taRNA and crRNA by a nickel-inducible promoter or constitutive promoter and inserted this riboregulator sequence upstream of the *cyabrB2* gene. Further, it should be mentioned that we cloned the *cyabrB2* gene itself and replaced the endogenous

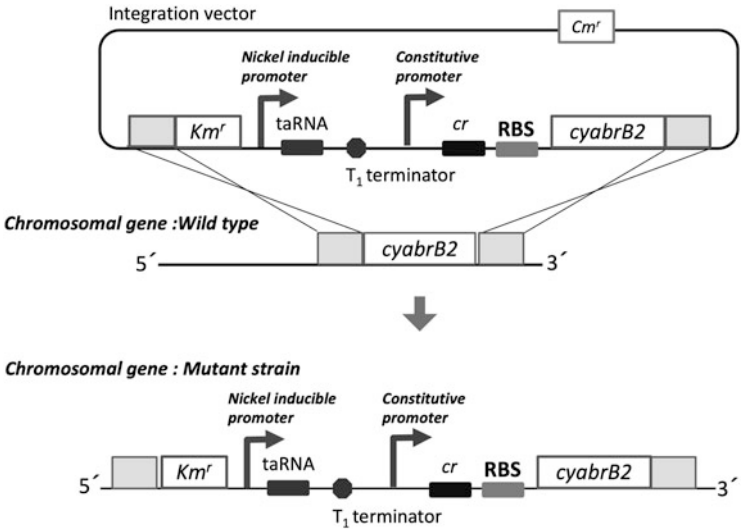


Fig. 10.5 Design of chromosomally integrated riboregulator in cyanobacteria. The endogenous cyabrB2 gene was replaced with the constructed integration vector containing the riboregulator sequence. cr shows a cis-repressive sequence

cyabrB2 gene with a riboregulator-inserted *cyabrB2* gene to remove the effect of the native chromosomal gene. After we constructed the riboregulator-inserted mutant, we evaluated the transcription and expression level of cyAbrB2 and observed high levels of repression for cyAbrB2. Additionally, high glycogen accumulation was observed when *cyabrB2* gene was knocked down by the riboregulator. Overall, we showed that the riboregulator could be applied to as a new tool for knockdown in cyanobacteria

10.6 Riboregulator for Tracking Cell Behavior

Though the riboregulatory system was first developed as a tight gene regulation system, this tool can be also applied as a sensing and tracking tool that can sensitively detect intercellular target molecules. Because repressor and inducer RNA fragments can be transcribed under certain promoters, we can utilize the riboregulator as a sensitive sensor in vivo.

To use the riboregulator as a tracking tool, Collins and coworkers regulated the transcription of the riboregulator using a promoter that can respond to intercellular ions to detect changes in the intercellular ion environment (Callura et al. 2010). In this report, transcription of the crRNA and GFP was controlled by the pTonB promoter, which can be repressed by binding the transcriptional regulator, Fur, in the presence of iron. Moreover, taRNA was controlled by the anhydrotetracycline-inducible promoter to adjust the sensitivity of this sensor. As a result, fluorescence intensity was drastically increased in the presence of an ion chelator, indicating that the riboregulator under the control of a sensitive promoter successfully detected the intercellular ions. This was the first trial showing the modularity of riboregulators as a tracking tool in vivo.

Additionally, several efforts to apply riboregulators for the detection of intercellular RNA fragments, such as small RNA and micro RNA, are underway. The strategy to utilize riboregulators for RNA fragment detection is that reporter gene expression can be induced only if the target RNA fragments bind to the crRNA sequence. This strategy itself was first demonstrated in vivo by fusing the complementary sequence of target RNA fragments, Chemokine Receptor 6 sequence, to the loop region of crRNA, and it was shown that the reporter gene was successfully expressed in the presence of target RNA fragments in the cell-free translation system (Narita et al. 2005). Though the availability of this system was not confirmed in vivo in this report, a recently published report showed that this concept can be applied for intercellular RNA fragment detection (Wang et al. 2019). Liu and coworkers redesigned the toehold switch for detecting miR-155 and miR-21, which are microRNAs in cancer cells. The complementary sequence of miRNA was fused to the upstream region of the toehold switch so that the reporter gene could be expressed in the presence of target miRNA. The vector containing this designed miRNA-targeting toehold switch was transcribed in mammalian cells and successfully detected the miRNA. Additionally, the authors succeeded in multiple detection of two miRNAs by employing two riboregulator sequences

with different reporter genes. It is surprising that the designed riboregulator can respond to small miRNA fragments just by fusing 23 bases of the miRNA-binding sequence. It should be noted that the toehold switch needs to contain only the three consecutive bases of the start codon and does not need to contain the RBS-binding complementary sequence bearing more than 6 consecutive bases on its regulatory region. Because of its high design feasibility, the toehold switch can be applied to the detection of other target RNA fragments.

10.7 Conclusion and Future Perspective

In this chapter, we traced the development and improvement of riboregulators focusing on their usability in synthetic biology. Though this RNA tool was first developed as a tight and strong gene regulation tool, there is still room for its improvement. For example, there are several efforts to increase its stability and induction level by fusing a scaffold to the crRNA sequence or by redesigning the RNA sequence through computational analysis. Addition of functionality has also been attempted by fusing other RNA aptamers to taRNA. Throughout these improvements, it was clearly shown that the feasibility of riboregulator sequence design is one of the large features that also enables its utilization in different bacterial cells, such as cyanobacteria.

Following sequence improvement of the riboregulator, this RNA tool was employed for further applications, such as biomass production or detection of target molecules. For details, tight and rapid regulation of the target gene can be applied to multiple gene regulation for succinic acid production in *E. coli*. Our research group also succeeded in higher glycogen production in cyanobacteria using a chromosomally integrated riboregulator. Moreover, applying a riboregulator for detection of cellular conditions and small RNA fragments was performed by redesigning the crRNA sequence.

Because the reports covered in this chapter did not employ exactly the same riboregulator sequence, we cannot make a sweeping statement about the features of a riboregulator, such as its induction level and ON/OFF ratio. However, it should be noted that this RNA-based gene regulation tool has wide application possibilities because of its design ease, orthogonality, and high repression level. It can be assumed that combination with other gene regulation tools, such as ligand-sensitive promoters, RNA scaffolds, and other protein-based tools, might further expand the applications of riboregulators.

Competing Interests There is no competing interest.

References

Abe K, Sakai Y, Nakashima S et al (2014) Design of riboregulators for control of cyanobacterial (*Synechocystis*) protein expression. Biotechnol Lett 36:287–294

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol 23:337–343

Bohn C, Rigoulay C, Chabelskaya S et al (2010) Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism. Nucleic Acids Res 38:6620–6636

Breaker RR, Winkler W, Nahvi A (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419:952–956

Brennan RG, Link TM (2007) Hfq structure, function and ligand binding. Curr Opin Microbiol 10:125–133

Cabello CM, Bair WB, Lamore SD et al (2010) Science 46:220–231

Callura JM, Dwyer DJ, Isaacs FJ et al (2010) Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. Proc Natl Acad Sci 107:15898–15903

Camsund D, Lindblad P (2014) Engineered transcriptional systems for cyanobacterial biotechnology. Front Bioeng Biotechnol 2:40

Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold switches: De-novo-designed regulators of gene expression. Cell 159:925–939

Gultyaev AP, Franch T, Gerdes K (1997) Programmed cell death by *hok/sok* of plasmid R1:coupled nucleotide covariations reveal a phylogenetically conserved folding pathway in the hok family of mRNAs. J Mol Biol 17:26–37

Hjalt TÅH, Gerhart E, Wagner H (1995) Bulged-out nucleotides protect an antisense RNA from RNase III cleavage. Nucleic Acids Res 23:571–579

Horos R, Büscher M, Kleinendorst R et al (2019) The small non-coding vault RNA1-1 acts as a riboregulator of autophagy. Cell 176:1054–1067

Huang HH, Camsund D, Lindblad P, Heidorn T (2010) Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. Nucleic Acids Res 38:2577–2593

Isaacs FJ, Dwyer DJ, Ding C et al (2004) Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22:841–847

Janssen RA, Mier JW (1997) Tropomyosin-2 cDNA lacking the 3′ untranslated region riboregulator induces growth inhibition of v-Ki-ras-transformed fibroblasts. Mol Biol Cell 8:897

Jupe ER, Liu XT, Kiehlbauch JL et al (1996a) The 3′ untranslated region of prohibitin and cellular immortalization. Exp Cell Res 224:128

Jupe ER, Liu XT, Kiehlbauch JL et al (1996b) Prohibitin in breast cancer cell lines: loss of antiproliferative activity is linked to 3′ untranslated region mutations. Cell Growth Differ 7:871–878

Kaniya Y, Kizawa A, Miyagi A et al (2013) Deletion of the transcriptional regulator cyAbrB2 deregulates primary carbon metabolism in *Synechocystis* sp. PCC 6803. Plant Physiol 162:1153–1163

Khan MI, Shin JH, Kim JD (2018) The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. Microb Cell Factories 17:36

Lease R, Cusick M, Belfort B (1998) Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA: RNA. Proc Natl Acad Sci U S A 95:12456–12461

Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. J Mol Biol 344:1211–1223

Liu R, Liang L, Wu M et al (2013) CO2 fixation for succinic acid production by engineered *Escherichia coli* co-expressing pyruvate carboxylase and nicotinic acid phosphoribosyltransferase. Biochem Eng J 79:77–83

Lucks JB, Qi L, Mutalik VK et al (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci U S A 108:8617–8622

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

Man S, Cheng R, Miao C et al (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. Nucleic Acids Res 39:e50

Mazard S, Penesyan A, Ostrowski M et al (2016) Tiny microbes with a big impact: the role of cyanobacteria and their metabolites in shaping our future. Mar Drugs 14:e97

Narita A, Ogawa K, Sando S, Aoyama Y (2005) Highly sensitive genotyping using artificial riboregulator system. Nucleic Acids Symp Ser 49:271–272

Okino S, Noburyu R, Suda M et al (2008) An efficient succinic acid production process in a metabolically engineered Corynebacterium glutamicum strain. Appl Microbiol Biotechnol 81:459–464

Papenfort K, Pfeiffer V, Lucchini S et al (2008) Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. Mol Microbiol 68:890–906

Peters G, Maertens J, Lammertyn J, De Mey M (2018) Exploring of the feature space of de novo developed post-transcriptional riboregulators. PLoS Comput Biol 14(8):e1006170

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 U S A 109:15271–15276

Sakai Y, Abe K, Nakashima S et al (2014) Improving the gene-regulation ability of small RNAs by scaffold engineering in *Escherichia coli*. ACS Synth Biol 3:152–162

Sakamoto I, Abe K, Kawai S et al (2018) Improving the induction fold of riboregulators for cyanobacteria. RNA Biol 15:353–358

Senoussi A, Lee Tin Wah J, Shimizu Y et al (2018) Quantitative characterization of translational Riboregulators using an in vitro transcription-translation system. ACS Synth Biol 7:1269–1278

Sharma V, Yamamura A, Yokobayashi Y (2012) Engineering artificial small RNAs for conditional gene silencing in *Escherichia coli*. ACS Synth Biol 1:6–13

Sun X, Zhulin I, Wartell RM (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. Nucleic Acids Res 30:3662–3671

Ueno K, Sakai Y, Shono C et al (2017) Applying a riboregulator as a new chromosomal gene regulation tool for higher glycogen production in *Synechocystis* sp. PCC 6803. Appl Microbiol Biotechnol 101:8465–8474

Wang J, Qin D, Zhang B et al (2015) Fine-tuning of *ecaA* and *pepc* gene expression increases succinic acid production in *Escherichia coli*. Appl Microbiol Biotechnol 99:8575–8586

Wang J, Wang H, Yang L et al (2018) A novel riboregulator switch system of gene expression for enhanced microbial production of succinic acid. J Ind Microbiol Biotechnol 45:253–269

Wang J, Yang L, Wang D et al (2016) Enhanced succinic acid productivity by expression of mgtCB gene in *Escherichia coli* mutant. J Ind Microbiol Biotechnol 43:505–516

Wang S, Emery NJ, Liu AP (2019) A novel synthetic toehold switch for MicroRNA detection in Mammalian cells. ACS Synth Biol 8:1079–1088

Yamauchi Y, Kaniya Y, Kaneko Y, Hihara Y (2011) Physiological roles of the cyAbrB transcriptional regulator pair Sll0822 and Sll0359 in *Synechocystis* sp. strain PCC 6803. J Bacteriol 193:3702–3709

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

Recent Advances, Challenges, and Opportunities in Riboswitches

11

Chandrasekaran Ramakrishnan, Manish Kesharwani, and Devadasan Velmurugan

Abstract

Riboswitch is a regulatory element of the mRNA, binds small molecule, and modulates the translation process of synthesizing the corresponding protein. Hence, the RNA is capable with self-regulation. This is an important mechanism of transcription and posttranscription level gene regulations in the central dogma. This regulatory element has a conserved motif and is present in the 5′ UTRs, an upstream region of genes. In this chapter, we have provided a detailed information on the mechanism of riboswitches with an up-to-date scientific findings and the design and applications of synthetic riboswitches in the industry and medicine domains. In this review, riboswitches with unique molecular features such as ligand-binding, termination-regulating, and translational-regulating properties are also discussed. Besides the many experimental techniques, the nonredundant role of computational techniques in the design of riboswitches is emphasized. Additionally, stability and conformational dynamics of riboswitches and their biophysical models explaining ligand-binding and translation initiation rate are also highlighted.

Keywords

Synthetic riboswitch · Noncoding RNA · Gene regulation · Riboswitch classification

C. Ramakrishnan
Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

M. Kesharwani · D. Velmurugan (✉)
Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai, India

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_11

11.1 Introduction

Riboswitches are present in the noncoding segment of the messenger RNA and act as structured element for gene regulation. Many microorganisms, like bacteria and fungi, and plants have riboswitches in their mRNAs. The structural element comprises two distinguished domains called "aptamer domain" and "expression platform" (Fig. 11.1). The aptamer domain binds the small molecule ligand from cellular metabolites. Mostly, an increase concentration of the ligand leads to more stable fold of aptamer domain and that results in further folding of expression platform. Hence, the small molecule ligand binding transduces into a signal that regulates gene expression. The key structural change that happens during folding is the formation of the secondary structure, and it is in response to the allosteric communication between the expression platform and aptamer domain. This secondary structure formation is acting as a switch that is similar to the termination/antitermination stem-loop or ribosome-binding site (RBS), responsible for the transcription and posttranscription level gene regulations, respectively (Breaker 2011; Garst et al. 2011). As of now, the riboswitched are identified with 24 unique features and many of them are characterized with three-dimensional structures. Also, several structures of ligand-bound aptamers are also reported. The important ligands are c-di-AMP, flavin mononucleotide (FMN), glycine, lysine, purines, S-adenosyl methionine (SAM), and thiamine pyrophosphate (TPP) (Breaker 2011). Riboswitches are broadly classified based on the size, shape, type of gene regulation (transcriptional and posttranscriptional), charge of sensor ligands, and structural properties. Particularly, type I riboswitches have the tendency of binding its congate ligand at the helical bundle as well as pseudoknots. Some of the ligands bind type I riboswitches are glucosamine-6-phosphate (GlmS) and purines. Here, the ligand binding causes a conformational change in the ligand-binding site of the aptamer domain. On the other hand, the type II riboswitches bind ligands through their peripheral structural elements and they undergo a global conformational change upon ligand binding (Montange and Batey 2008).

The mechanism of natural riboswitches was exploited in synthetic biology to produce synthetic riboswitches. Particularly, in gene therapy and metabolic engineering, the synthetic riboswitches play a crucial role in regulation of gene expression, externally. RNA aptamers are generated using de novo method where large collections of unique and randomized oligonucleotides ($\sim 10^{14}$) are synthe-

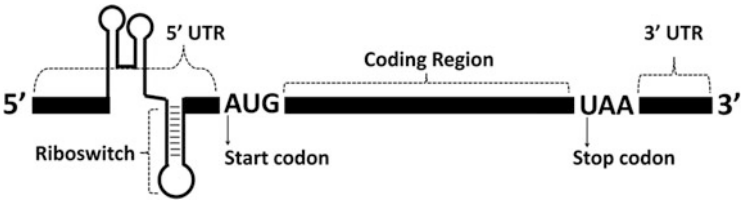


Fig. 11.1 Structure of a riboswitch (Edwards and Batey 2010)

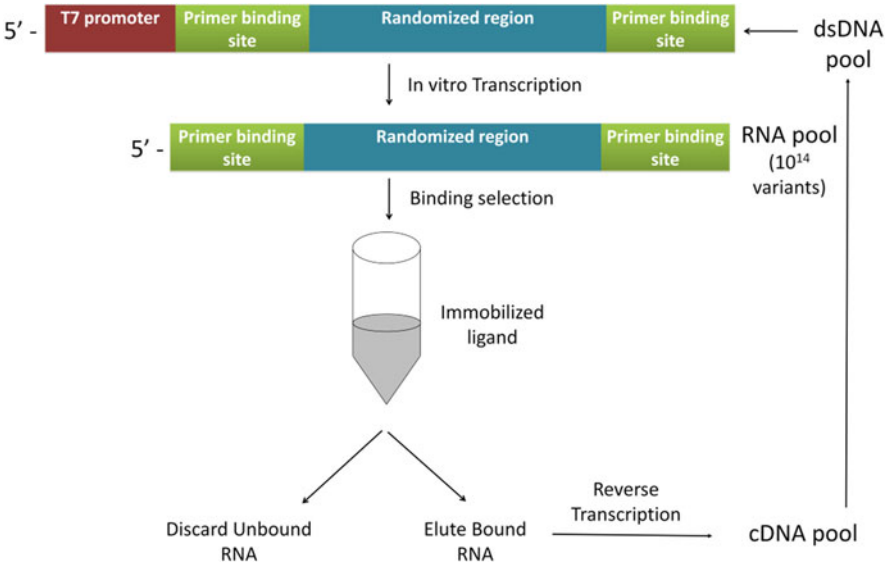


Fig. 11.2 Aptamer selection method. Reconstructed image with reference to Topp and Gallivan (2010)

sized and in vitro methods, say, SELEX (Systematic Evolution of Ligands by EXponential enrichment), are used to choose the candidate (Topp and Gallivan 2010) as shown in Fig. 11.2.

Among the millions of random fragments, only few fragments can bind small molecule ligands, have the characteristics of synthetic riboswitches (Weigand et al. 2011), and have a molecular feature of switching from apo/ground to ligand-bound conformation through a large conformational transition (Fig. 11.1). A study by Suess and coworkers (Weigand et al. 2014) confirmed that the synthetic riboswitches serve as a gene expression control device (Weigand et al. 2008) through mutations introduced in the distal to the ligand-binding site which affect the biological activity.

Synthetic riboswitch design and development are still in growth phase. The aptamers selected using the conventional SELEX method result only high-affinity candidates, and only few of them have conformational switch feature. However, many studies demonstrated the successful development of riboswitch using theophylline aptamer. A recent study by Suess et al. dealt with these limitations through their riboswitch developmental pipeline. Herein, they have reported paromomycin-binding riboswitch with a Kd value of 20 nM, and they were able to characterize the conformational switch between ligand-free and ligand-bound forms (Boussebayle et al. 2019).

11.2 RNA-Mediated Gene Regulation

As this chapter focuses on the scope of riboswitches, the mechanism of noncoding RNA-mediated gene regulation is briefed in this section. Different types of translation control mechanisms have been discovered, and they are mediated by the miRNA or siRNA–protein complex. Eventually, the mRNA is being degraded in the process. In brief, the miRNA and siRNA control translation by nonredundant mechanisms involving the ribonuclear proteins (RNA-induced silencing complex, RISC). In addition to gene expression control, these molecules are found to involve in epigenetic modifications during evolution (Kloc et al. 2008). Moreover, siRNAs are found to be involved in the antiviral defense also. Mainly, in plants, the antiviral defense mechanism is influenced by the conversion of ssRNA to dsRNAs and cleavage of dsRNA, which results in the small siRNAs that interfere with the translation process.

11.2.1 Riboswitch in Prokaryotes

Bacterial infections are a major concern due to the antibiotics resistance acquired by the pathogens. Particularly, treating bacterial infections caused by *E. coli* and methicillin-resistant *S. aureus* (MRSA) bacteria is very challenging to date. Most of the riboswitches were identified in microorganisms and plant sources, and there is no clear evidence for their occurrence in human. These riboswitches endeavor as specific potential drug target to design novel antibiotics. Crystal structures of TPP-bound aptamer domain of thiM riboswitch reveal only static event of ligand binding (Serganov et al. 2006). A typical riboswitch scaffold in bacterial system with different conformational states in response to ligand unbound (−M) and bound (+M) is depicted in Fig. 11.3.

11.2.2 Riboswitch in Eukaryotes

The only riboswitch found in the eukaryotes is TPP riboswitch. It is also present in the bacteria and governs gene regulation at the transcriptional level (Bian et al.

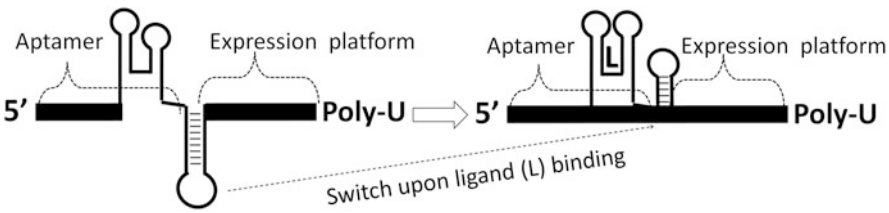


Fig. 11.3 Riboswitch-mediated gene regulation in bacteria

2011). In eukaryotes, it regulates the splicing and stability of mRNA. Importantly, the eukaryotic TTP riboswitch is present in introns or 3′ UTRs, in contrary to those found in prokaryotic sources (Cheah et al. 2007; Wachter et al. 2007). This necessitates the search of riboswitch and its expression based on the introns while regulation of eukaryotic gene expression is of interest. On the other hand, a riboswitch of a bacterial origin, cyclic-di GMP riboswitch is found functioning in the human system. Herein, the riboswitch regulates the expression of RNA-activated protein kinase (RPK) and modulates immuno-signaling (Hull et al. 2016).

11.3 Classification of Riboswitches

A broad classification made based on the ligand type defines four distinct riboswitches. The S-adenosyl methionine (SAM/SAM-I/SAM-II/SAM-III) riboswitches are of ligand type "enzyme cofactors"; guanine, adenine, 2′-deoxy-guanosine are "nucleotide precursors"; lysine is "amino acid"; and magnesium is "metal ion." Table 11.1 describes the classes of riboswitch based on ligand type and their discovery details.

11.4 Structure–Function Relationship of Riboswitches

11.4.1 Structure of Riboswitches

Riboswitch is a structured element of the mRNA that is evolved or designed to recognize a specific low-molecular-weight ligand which results in the modulation of gene expression. A series of conformational changes occur during the recognition process that connects ligand-free and ligand-bound states which are the two functionally relevant states. This also depends on the concentration of the ligand at the intracellular environment. To date, many structures (274 as of May, 2019) of riboswitches are available in the protein data bank (PDB). In these, 264 are determined using X-ray crystallography, 9 using NMR, and 1 using electron microscopy. About 80 different small molecule ligands binding different riboswitches are found in the database. Number of structures available for different organisms and different riboswitch types (bacteria 140, unassigned 107, eukaryota 25, other 10, virus 1) are depicted in Fig. 11.4a, b, respectively.

The 22 structures available for *Homo sapiens* include a wild-type and different mutant GMP riboswitches, whereas 140 structures available for bacteria include SAM-I, PRPP, thiM TPP, fluoride, and FMN riboswitches.

11.4.2 Conformational Transition

The conformational transition in riboswitch is a well-studied structural feature. using NMR and ITC experiments reported that a synthetic neomycin (an amino-

Table 11.1 Classification of riboswitches and structural data available in the protein data bank

Type of ligand	Ligand	Family/class	PDB ID:Chain ID
Enzyme cofactor		SAM-I	2gis:a; 2ydh:a; 2ygh:a; 3gx2:a; 3gx3:a; 3gx5:a; 3gx6:a; 3gx7:a; 3iqn:a; 3iqp:a; 3iqr:a; 3v7e:c; 3v7e:d; 4aob:a; 4b5r:a; 4kqy:a; 5fjc:a; 5fk1:a; 5fk2:a; 5fk3:a; 5fk4:a; 5fk5:a; 5fk6:a; 5fkd:a; 5fke:a; 5fkf:a; 5fkg:a; 5fkh:a
	S-adenosyl-methionine	SAM-II	
		SAM-III	
	Thiamine pyrophosphate	TPP	2cky:a; 2cky:b; 2gdi:x; 2gdi:y; 2hoj:a; 2hok:a; 2hol:a; 2hom:a; 2hoo:a; 2hop:a; 3d2g:a; 3d2g:b; 3d2v:a; 3d2v:b; 3d2x:a; 3d2x:b; 3k0j:e; 3k0j:f; 4nya:a; 4nya:b; 4nyb:a; 4nyc:a; 4nyd:a; 4nyg:a
Nucleotide precursor	Guanine	Purine/G	1y26:x; 1y27:x; 2b57:a; 2ees:a; 2eet:a; 2eeu:a; 2eev:a; 2eew:a; 2g9c:a; 2xnw:a; 2xnz:a; 2xo0:a; 2xo1:a; 3ds7:a; 3ds7:b; 3fo4:a; 3fo6:a; 3g4m:a; 3gao:a; 3ger:a; 3ges:a; 3gog:a; 3got:a; 3la5:a; 3rkf:a; 3rkf:b; 3rkf:c; 3rkf:d; 4fe5:b; 4fej:b; 4fel:b; 4fen:b; 4feo:b; 4fep:b; 4 lx5:a; 4 lx6:a; 4tzx:x; 4tzy:x; 4xnr:x; 5c7u:b
	Adenine	Purine/A	
	2′-Deoxyguanosine	Purine/dG	
Amino acid	Lysine	Lysine	3d0u:a; 3d0x:a; 3dig:x; 3dil:a; 3dim:a; 3dio:x; 3diq:a; 3dir:a; 3 dis:a; 3dix:a; 3diy:a; 3diz:a; 3dj0:a; 3dj2:a; 4erj:a; 4erl:a
Metal ion	Magnesium	Mg2+/ykoK	2qbz:x; 3pdr:x; 3pdr:a

Note: This classification scheme was adopted from the Nature—Scitable

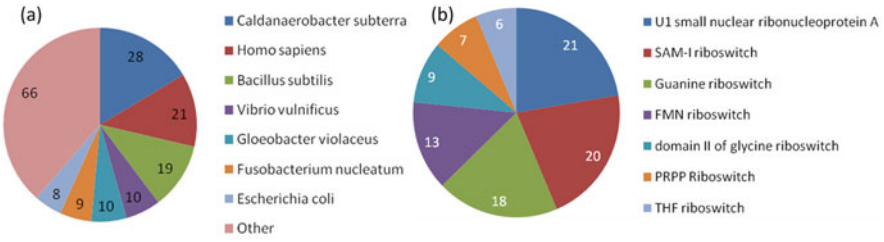


Fig. 11.4 Statistics of the various riboswitch structures deposited in the PDB

glycoside antibiotic) riboswitch undergoes conformational changes (Weigand et al. 2014). An open conformation (no interference with ribosomal scanning) in the presence of ligand was characterized with a stem-loop structure, a canonical U-turn fold of internal loop as an important structural feature for regulation (PDB ID:2KXM).

Harald and co-workers (Furtig et al. 2015) reported that the RNAs have a hierarchical free energy landscape, and they have low-energy barriers between alternative tertiary structures. Hence, the RNAs tend to switch between different conformations, rapidly. From the available crystal structures, the ligand-bound conformations of riboswitches, apo-state, and the unbound conformation, holo-state, are well differentiable. Particularly, the three-dimensional structures available for the SAM-1 riboswitch from *T. tencongensis*, lysine riboswitch from *Thermotoga maritima*, and the glycine riboswitch from *V. cholerae* are helpful to understand the distinct apo conformation. In this, the aptamer segment shares the same secondary structure. The biophysical means of the stabilization of three unique conformational states of riboswitches (between the ligand-free-apo and ligand-bound-holo states) can be understood by the chemical equilibrium constant (K_{pre}) (Furtig et al. 2015) defined as follows:

$$K_{\text{pre}} = \frac{[\text{apo}A]}{[\text{apo}B]} = \frac{\text{K}_{\text{apoB}} \rightarrow \text{K}_{\text{apoA}}}{\text{K}_{\text{apoA}} \rightarrow \text{K}_{\text{apoB}}} = e^{-\frac{\Delta G(\text{apo}A - \text{apo}B)}{R.T}}$$

11.4.3 Ligand Recognition

The structural determinants of ligand recognition by riboswitches vary with the type and source of organisms. Basically, the aptamer domain is folded into a three-way junction and forms a ligand-binding site. This region is being exploited in synthetic chemistry to make random fragments followed by selection of the potent ones. This can be seen naturally by the existence of only four classes of riboswitches for many decades while there are more than 30 different structural architectures for binding different ligands (Weinberg et al. 2017). Here, bioinformatics approach becomes more efficient in the prediction of ligand-binding sequences and calculation of ligand-binding energy. Molecular dynamics simulations have also become a robust approach to assess the stability of the riboswitch-ligand complexes. A study by Micura et al. (Haller et al. 2013) explained the molecular mechanism of TPP riboswitch binding to the ligand through a three-state model which was earlier proposed by SAXS (Ali et al. 2010; Baird and Ferre-D'Amare 2010). Further, Velmurugan and co-workers (Kesherwani et al. 2018) explained the molecular mechanism of the TPP riboswitch and the nucleotides that are crucial for stability and ligand binding. Similarly, the recognition mechanism can be referred in the literature for other riboswitches as well (Gilbert et al. 2006; Kim and Breaker 2008; Mulhbacher and Lafontaine 2007; Serganov 2010; Serganov and Patel 2009).

11.5 Computational Prediction of Riboswitches

Development of riboswitch has the importance in clinical and synthetic biology where its application becomes robust. In order to speed up the process, many algorithms and tools are being developed and improved (Antunes et al. 2017). For the past few decades, algorithms and methods are being developed for accurate prediction of secondary structure, which is a crucial part of RNA folding and even in the prediction of riboswitches (Andronescu et al. 2010; Mathews et al. 2004; Zuker 1989). A recent study based on the machine learning techniques is a successful case in identification of synthetic riboswitches. The hydrogen bond pattern and biophysical parameters are identified as important features influencing tc-riboswitch-mediated gene regulation (Groher et al. 2019). The in vitro approaches based on aptamers are not successful in making functional riboswitch, and hence, the computational methods become a robust to overcome the time and cost involvement in the process. In this way, secondary structure prediction-based development of theophylline riboswitches is noteworthy (Domin et al. 2017). Here, de novo methods followed by optimization using structure-guided experimental methods were employed. Moreover, tetracycline aptamers, Theo tet aptamers, and their variants substantiate the robustness of computational methods to predict and design the apatamers. Computational software modules, such as RNAComposer and Rosetta, and Langevin dynamics simulation based on coarse-grained self-organized polymer (SOP) and BarMap approach are some of the important and popular tools for riboswitch and RNA structure predictions (Gong et al. 2017). Particularly, RNAComposer (Popenda et al. 2012) tools under the Rosetta project are helpful with well-optimized models to predict the structures of the riboswitch and structural changes in the aptamers. Moreover, to control the mammalian gene expression control and controlling the viral replications in the mammalian host, mathematical models to predict the *aptazymes* (a combination of riboswitch and ribozyme) are also applied in the synthetic riboswitch.

A comprehensive list of tools available for riboswitch motif prediction and analyses is as follows:

1. *CARNA*: http://rna.informatik.uni-freiburg.de/CARNA/Input.jsp
2. *DRD:* http://drd.denison.edu/
3. *HMMER*: http://hmmer.org/
4. *RegRNA* 2.0: http://regrna2.mbc.nctu.edu.tw/index.html
5. *RegRNA*: http://regrna.mbc.nctu.edu.tw/html/prediction.html
6. *RNAComposer*: http://rnacomposer.cs.put.poznan.pl/
7. *RiboSW*: http://ribosw.mbc.nctu.edu.tw/
8. *riboswitch Finder*: http://riboswitch.bioapps.biozentrum.uni-wuerzburg.de/
9. *Riboswitch scanner*: http://service.iiserkol.ac.in/~riboscan/application.html
10. *RibEx*: http://www.ibt.unam.mx/biocomputo/ribex.html
11. *RNAConSLOpt*: http://genome.ucf.edu/RNAConSLOpt/
12. *Rosetta*: https://www.rosettacommons.org/
13. *SwiSpot*: http://www.iet.unipi.it/a.bechini/swispot/

11.6 Artificial Riboswitches and Their Applications

Through characterization of natural riboswitches as well as their synthetic variants and understanding their molecular mechanism, scientists enabled to modulate the function of existing riboswitches. A coupled approach with mathematical modeling and experimental methods eases the investigation of structure–function relationship of almost all types of riboswitches existing to date (Beisel and Smolke 2009). However, attaining the three basic features of successful aptamers, such as (1) binding its ligand, (2) conformational switch from free to ligand-bound form, and (3) fast recognition of ligand by the aptamer, is very challenging. As the in vitro selection process addresses only affinity, in vivo method is needed to compensate the other features (Wittmann and Suess 2012; Page et al., 2018). In parallel with the development of riboswitches, artificial or synthetic riboswitches can also be based on the "*aptazymes.*" They are also defined as a ligand-dependent self-cleaving ribozymes. They are very well functional in the mammalian hosts and control the gene expression. In another case, an aptazyme was designed to control the E1A protein expression in adenovirus and has an important role in oncolysis (Ketzer et al. 2014). Further, successful addition of aptazymes into the guide RNAs resulted with an aptazyme-embedded guide RNA and small molecule-controlled genome editing (Tang et al. 2017), which involves a popular CRISPR-Cas9 system. Together, they are all robust tools for ligand-driven gene expression control and are instrumental in synthetic biology and clinical applications (Lee et al. 2016; Stifel et al. 2019).

11.6.1 Applications

Applications of synthetic riboswitches become successful as the basis of RNA-based interactions and folding are common among the bacterial and eukaryotic sources. The following list gives a range of application of synthetic riboswitches that are successful currently:

- Bacterial infections are being treated with novel antibiotic targeting bacterial riboswitches. (Villa et al. 2018; Stav et al. 2019)
- Logic gates for diagnostics and biocontainment detection (Gallagher et al. 2015).
- Paper-based diagnostics to detect viral RNAs (Pardee et al. 2014).
- Conditional gene regulation.
- Biosensor engineering and fluorescent biosensors.
- Engineering metabolic flux.
- Protein detecting reagents.
- Preparation of affinity columns.
- Applications in hematology.
- Design of antiviral agents.
- Antibody development with controlled autoimmune effects.
- Deliver siRNA to the target cells.

11.7 Conclusion

Synthetic riboswitch becomes a unique tool among the many RNA variants in modulating the gene expression and a potential framework for RNA engineering with various applications. It includes feasibility of structural characterization, defined genetic footprint, and is modular in nature. Scope for engineering the riboswitches increased as the aptamer domain is very sensitive toward even the synthetic organic compounds and paves the way to orthogonal gene regulation. There are many strategies, such as development of mathematical and computational models, structure-based drug design, and screenings. The in vitro screening of engineered aptamers has bottleneck with flood of random variants, does not hit low-affinity aptamers, and has less yielding. Overall, a defined platform for developing novel antibiotics is provided by the riboswitches due to their application in modulating gene regulation in a wide spectrum of microorganisms including bacteria and viruses. A continuous effort with novel methods in the field of synthetic riboswitches definitely results in the production of novel therapeutics in human health care.

References

Ali M et al (2010) The ligand-free state of the TPP riboswitch: a partially folded RNA structure. J Mol Biol 396(1):153–165

Andronescu M et al (2010) Computational approaches for RNA energy parameter estimation. RNA 16(12):2304–2318

Antunes D et al (2017) Using RNA sequence and structure for the prediction of riboswitch aptamer: a comprehensive review of available software and tools. Front Genet 8:231

Baird NJ, Ferre-D'Amare AR (2010) Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. RNA 16(3):598–609

Beisel CL, Smolke CD (2009) Design principles for riboswitch function. PLoS Comput Biol 5(4):e1000363

Bian J et al (2011) The riboswitch regulates a thiamine pyrophosphate ABC transporter of the oral spirochete Treponema denticola. J Bacteriol 193(15):3912–3922

Boussebayle A et al (2019) Next-level riboswitch development-implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. Nucleic Acids Res 47(9):4883–4895

Breaker RR (2011) Prospects for riboswitch discovery and analysis. Mol Cell 43(6):867–879

Cheah MT et al (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature 447(7143):497–500

Domin G et al (2017) Applicability of a computational design approach for synthetic riboswitches. Nucleic Acids Res 45(7):4108–4119

Edwards AL, Batey RT (2010) Riboswitches: a common RNA regulatory element. Nat Educ 3(9):9

Furtig B et al (2015) Multiple conformational states of riboswitches fine-tune gene regulation. Curr Opin Struct Biol 30:112–124

Gallagher RR et al (2015) Multilayered genetic safeguards limit growth of microorganisms to defined environments. Nucleic Acids Res 43(3):1945–1954

Garst AD, Edwards AL, Batey RT (2011) Riboswitches: structures and mechanisms. Cold Spring Harb Perspect Biol 3(6):a003533

Gilbert SD et al (2006) Structural studies of the purine and SAM binding riboswitches. Cold Spring Harb Symp Quant Biol 71:259–268
Gong S et al (2017) Computational methods for modeling aptamers and designing riboswitches. Int J Mol Sci 18(11):2442
Groher AC et al (2019) Tuning the performance of synthetic riboswitches using machine learning. ACS Synth Biol 8(1):34–44
Haller A et al (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. Proc Natl Acad Sci U S A 110(11):4188–4193
Hull CM, Anmangandla A, Bevilacqua PC (2016) Bacterial riboswitches and ribozymes potently activate the human innate immune sensor PKR. ACS Chem Biol 11(4):1118–1127
Kesherwani M, N HVK, Velmurugan D (2018) Conformational dynamics of thiM riboswitch to understand the gene regulation mechanism using Markov state modeling and the residual fluctuation network approach. J Chem Inf Model 58(8):1638–1651
Ketzer P et al (2014) Artificial riboswitches for gene expression and replication control of DNA and RNA viruses. Proc Natl Acad Sci U S A 111(5):E554–E562
Kim JN, Breaker RR (2008) Purine sensing by riboswitches. Biol Cell 100(1):1–11
Kloc A et al (2008) RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol 18(7):490–495
Lee CH, Han SR, Lee SW (2016) Therapeutic applications of aptamer-based riboswitches. Nucleic Acid Ther 26(1):44–51
Mathews DH et al (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A 101(19):7287–7292
Montange RK, Batey RT (2008) Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys 37:117–133
Mulhbacher J, Lafontaine DA (2007) Ligand recognition determinants of guanine riboswitches. Nucleic Acids Res 35(16):5568–5580
Page K et al (2018) Engineering riboswitches in vivo using dual genetic selection and fluorescence-activated cell sorting. ACS Synth Biol 7(9):2000–2006
Pardee K et al (2014) Paper-based synthetic gene networks. Cell 159(4):940–954
Popenda M et al (2012) Automated 3D structure composition for large RNAs. Nucleic Acids Res 40(14):e112
Serganov A (2010) Determination of riboswitch structures: light at the end of the tunnel? RNA Biol 7(1):98–103
Serganov A, Patel DJ (2009) Amino acid recognition and gene regulation by riboswitches. Biochim Biophys Acta 1789(9–10):592–611
Serganov A et al. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nat 441(7097):1167–1171
Stav S et al (2019) Genome-wide discovery of structured noncoding RNAs in bacteria. BMC Microbiol 19(1):66
Stifel J, Spöring M, Hartig JS (2019) Expanding the toolbox of synthetic riboswitches with guanine-dependent aptazymes. Synth Biol 4(1):ysy022
Tang W, Hu JH, Liu DR (2017) Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. Nat Commun 8:15939
Topp S, Gallivan JP (2010) Emerging applications of riboswitches in chemical biology. ACS Chem Biol 5(1):139–148
Villa JK et al (2018) Synthetic biology of small RNAs and riboswitches. Microbiol Spectr 6(3). https://doi.org/10.1128/microbiolspec.rwr-0007-2017
Wachter A et al (2007) Riboswitch control of gene expression in plants by splicing and alternative 3′ end processing of mRNAs. Plant Cell 19(11):3437–3450
Weigand JE et al (2008) Screening for engineered neomycin riboswitches that control translation initiation. RNA 14(1):89–97
Weigand JE et al (2011) Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity. Nucleic Acids Res 39(8):3363–3372

Weigand JE et al (2014) Sequence elements distal to the ligand binding pocket modulate the efficiency of a synthetic riboswitch. Chembiochem 15(11):1627–1637
Weinberg Z et al (2017) Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. Proc Natl Acad Sci U S A 114(11):E2077–E2085
Wittmann A, Suess B (2012) Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. FEBS Lett 586(15):2076–2083
Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244(4900):48–52

Recent Advances in Gene and Genome Assembly: Challenges and Implications

12

Tanu Agarwal, Renuka Suravajhala, Mahak Bhushan, Prittam Goswami, Lahiru Iddamalgoda, Babita Malik, Jayaraman Valadi, and Prashanth Suravajhala

T. Agarwal
Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, India

NIIT University, Neemrana, Rajasthan, India

R. Suravajhala · B. Malik
Department of Chemistry, School of Basic Sciences, Manipal University Jaipur, Jaipur, Rajasthan, India

M. Bhushan
Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, India

Indian Institute of Science Education and Research, Kolkata, West Bengal, India

P. Goswami
Bioclues.org, Hyderabad, India

L. Iddamalgoda
Research Centre for Data Analytics and Cognition La Trobe University, Melbourne, VIC, Australia

J. Valadi
Computer Science, Flame University, Pune

P. Suravajhala (✉)
Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, India

Research Centre for Data Analytics and Cognition La Trobe University, Melbourne, VIC, Australia
e-mail: prash@bisr.res.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_12

Abstract

The last decade has seen a surge in genome assembly and sequencing technologies. From the good old DNA sequencing and electrophoresis reads enunciated by Sanger, Maxam, and Gilbert, several tools were embodied through the use of these methods and yet, the challenges these methods imposed have had a paradigm shift. In this focused review, we provide a gist of the tools used by the genome assemblers and discuss the challenges and implications of the next generation sequencing chemistries shaped up by the development of tools.

Keywords

Genome assembly · Genome · Next generation sequencing · Assembly algorithms · Neural network

12.1 Introduction

The next generation sequencing (NGS) technologies allow us to use high-throughput, fast, and correct strategies to determine the precise order of nucleotides (Henson et al. 2012). With the advent of NGS, bioinformatics has taken a big leap for interpretation and application of this biological information (Wheeler et al. 2008). Assembling a genome would produce fragments which are obtained and processed through many sequencing technologies (Pareek et al. 2011; Ekblom and Wolf 2014) that are introduced recently. These high-throughput technologies are characterized by typically shorter browse lengths (as low as 30–50 bp) and have different error characteristics than Sanger data (Nagarajan and Pop 2010; Ma et al. 2012a, b). Existing package tools for assembly and analysis of sequencing information are, therefore, not perfect to handle the new kinds of information (Mardis 2008). On the other hand, machine-driven sequencing technologies have allowed researchers to rewrite the genomes of the many organisms. With short read sequencing (SRS) technologies accelerating at a pace to explore the nature, it has provided challenges to package tools used to reconstruct genetic data (Pop 2004; Pop and Salzberg 2008). One among the foremost encouraging applications of SRS technologies is metagenomics, (Poinar et al. 2006; Richter et al. 2008) wherein deoxyribonucleic acid is extracted from associate degree atmosphere like soil, water, or a part of the organic structure, and also the mixture of species is sequenced employing a random amplicon or shotgun technique (Noguchi et al. 2006). The ensuing short reads originate from a whole bunch or thousands of various species, give comparatively a bigger assembly challenge than one genome sequencing project.

For NGS, annotation is an important key method for deducing the structure and performance of the assembled sequences (Dominguez Del Angel et al. 2018). Beside the protein-coding genes, annotation of different options like non-coding RNAs or presence of repetitive or regulative sequences can be done, and as a result final contiguous sequences could be scaffolded and curated before finally assembling them (Bussotti et al. 2013; Stricklin et al. 2005–2008). If there is a

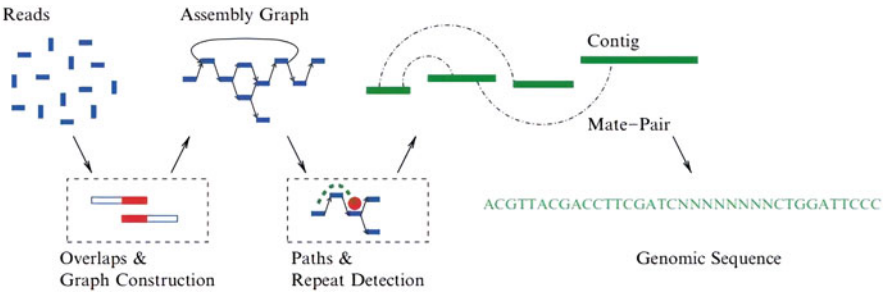


Fig. 12.1 Overview of de novo assembly. De novo assembly programs typically use the overlap between reads to construct a graph structure. After some simplifications of the graph, unambiguous paths in the graph are used to reconstruct contiguous sequences (contigs). Information such as the presence of mate-pair links between contigs may also allow the construction of gapped sequences (scaffolds)

reference available for the genome, one could go for it but the last decade has seen a surge in de novo genome assemblies. In a de novo genome assembly (Chaisson et al. 2015), the genome is first assembled by itself without a template, and then annotated. Two important classes of such assembly algorithms exist, viz. overlap–layout–consensus (OLC) and de Bruijn graphs (Miller et al. 2010). Their comparison is well documented elsewhere in a review by (Li et al. 2010). With high-quality browse assemblies revealing repeat structures and structural variation otherwise incomprehensible by SRS, the de novo genome assembly resulting in overlaps between reads that are long to distinguish (Zhang et al. 2011) (Fig. 12.1). On the other hand, the long sequencing reads has allowed the analysis in extremely repetitive satellite sequence in flies (Khost and Larracuente 2016) and birds (Weissensteiner et al. 2017), paving the approach for purposeful studies in areas of the genome not accessible (Phillippy 2017).

12.2 Next Generation Sequencing Chemistry

In 1977, two main methods for DNA sequencing were developed based on the automation of thermal cycle and nonhazardous chemicals. Sanger's sequencing method in which chain termination is caused by base-specific termination, i.e., (terminator ddNTP) of primed DNA synthesis by dideoxynucleotide analogs with specific fluorescent dyes has been extensively used in research and was referred as "First Generation Sequencing" (FGS) (Sanger et al. 1977). Whereas another approach in which the products were separated by labeled terminal DNA fragments through gel electrophoresis (Maxam and Gilbert 1977) is also relatively used. Later, second generation arose in the form of parallel sequencing that could produce huge data of a sample, cost effectively with a single run. As described earlier, it uses short reads and so known as short read length technology (Metzker 2005, 2009; Scholz et al. 2012). While this has been explored in various advanced scientific studies

related to wide range of human diseases and plant related microbial methods, a large application of it was seen towards ecological and evolutionary studies. With further refinement, third-generation sequencing that could read longer length at low cost was developed (Scholz et al. 2012). In the previous methods, long strands of DNA are cleaved into small segments and through amplification and synthesis, nucleotide sequences were read. We give a gist of details below.

12.3 First-Generation Sequencing

DNA double-helix structure that was published by Watson and Crick in 1953 (Watson and Crick 1953) and sequence of the natural polynucleotide (Bleidorn 2016) made its beginning of the first sequencing methods for tRNA and oligonucleotides. Further, it has led to sequencing of RNA and DNA and took a decade to develop DNA fragmentation with more than 100 different tRNA sequences (Franca et al. 2002). The first genome sequence of MS2 RNA long bacteriophage was with 3569 nucleotides and T4 DNA bacteriophage in lysozyme, 24-bp lac gene sequences (Callanan et al. 2018). Later on Maxam and Gilbert developed method for DNA sequence chemical degradation with specific bases at labeled DNA fragments terminals separated by electrophoresis method (Maxam and Gilbert 1977). The specific primer extension method was used to generate sequencing data of bacteriophage DNA that lead to Sanger's method (Holley et al. 1965). Thus, the Sanger based DNA chain-termination sequencing method effectively overleads Maxam and Gilbert's chemical degradation method (Balasubramanian 2015) due to its simple, accurate, less toxic, lower amounts of radioactivity chemicals. During this era, Applied Biosystems Instruments (ABI) has developed automated first-generation sequencer of DNA which was based on Sanger method, in which fluorescent dye and terminator reagents in a reaction lead to sequencing than that of four separate reactions (Shendure and Ji 2008; Barba et al. 2014; Franca et al. 2002). The sequencers with advance computers were used to validate and store the collected sequenced data (Guzvic 2013). Later, thermal resistant enzyme (Taq polymerase) extracted from *Thermus aquaticus* was used for PCR technology along with random or specific region resequencing (Barba et al. 2014). In early 1940s, discovery of reverse transcriptase developed to RNA sequencing through reverse transcribed cDNA from RNA evolved (Smith et al. 1986; Temin and Mizutani 1970). In 1991, Sanger based systematic sequencing of 373A DNA wherein semi-automated sequencers could generate cDNA long sequences with 397 bases known as expressed sequence tags (ESTs) along with substrates and markers were used for mapping RNA contigs. The updates were submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank) that has generated huge DNA sequences data during 1980–1990. Then came the breakthrough research in sequencing of first human genome sequence (Baltimore 1970) that has enhanced increased number of DNA and RNA sequences in GenBank between 1992 and 2004 (http://www.ncbi.nlm.nih.gov/genbank/statistics) wherein automated sequencers and industrialization have led to identification of 337 new human genes and 48 homologous genes from various

organisms (Venter et al. 2001). About 83 million nucleotides of cDNA sequences, 87,000 human cDNA sequences, and also two bacterial species, *Haemophilus influenzae* (Adams et al. 1991) and *Mycoplasma genitalium* (Fleischmann et al. 1995) whole genome sequences were reported in 1999 by EST data sequence assembler, TIGR (Fraser et al. 1995).

An automated sequencer with automated Prism 3700, 96 capillaries with a capacity to produce sequence data of 1.6×10^5 bases per day produced the whole genomic sequences and were developed by TIGR, Sanger Centre, and RIKEN in 2001 (Sutton et al. 1995). Sanger sequencing was most widely applied although it was expensive and consuming a lot of time. In postgenomic era, DNA sequencing, functional genomics predictions, SNPs, and transcript arrays had rapid applications in biological research studies (Sutton et al. 1995; Stein 2001). Soon after this, a rapid growth was seen towards various applications in prokaryotes and eukaryotic gene expression studies (Peltonen and McKusick 2001; Kiechle and Zhang 2002; Kulski et al. 2005). Later in 2005, Sanger dideoxy sequencing method came up to challenge the cost (Balasubramanian 2015; Shendure and Ji 2008; Barba et al. 2014) which is known as next generation sequencing (NGS), now aptly called as "now generation sequencing." It was designed to produce huge data of sequence from several samples besides shown to have high-throughput, degree of sequence coverage albeit less accuracy for individual reads through parallel strategies. On a lateral front, single-cell transcriptome study has seen towards a change as a big data. The error rate in Illumina-capillary (0.1–1%), Ion Torrent-454 sequencing (1.78%), Oxford nanopore (38.2%), SOLiD (1%), and Pacific Biosciences (15%), respectively (Rieber et al. 2013) and this short time and rapid development has lead to evolution of sequencing technologies.

12.4 Second-Generation Sequencing

Second-generation sequencing involves the use of cDNA which is reverse transcribed from RNA by shotgun sequencing without cloning the foreign host cells. The template libraries are constructed by linkers and adapter sequences and ligated to the ends of DNA. Later, amplification of solid surface or beads libraries is performed, and tiny emulsion droplets were collected and detected through luminescence for monitoring the electric charges of the nucleotides. With the NGS approaches generating enormous number of short reads, the nucleotides are analyzed in parallel in less time when compared to that of Sanger method. The fragment ends were generated either as single or pair end reads and thus, the data generated by NGS is easily accessible for comparing quantitatively. In this method, an important step would be downstream methods such as preparation and amplification of libraries (Kulski et al. 2005; Zhu 2003; Lenoir and Giannella 2006; Head et al. 2014; Hart et al. 2010). Before the library preparation, fragmentation of genomic DNA through acoustic shearing, sonication and later, enzymatic digestion with DNase I or fragmentase are carried. By ligation and PCR methods, the adapters, tags, barcodes, and primers are further labeled. There are other alternatives in the

form of Illumina's fragmentation technology where sample preparation time could be reduced (Head et al. 2014). On the other hand, in targeted sequencing, the regions of interest are captured in the fragmented DNA and supplemented custom-designed primers with probe-hybridization-capture or PCR amplification. RNA sequencing libraries are generated from RNA-binding protein, immunoprecipitation and the non-coding RNAs could be isolated, whereas in methylation sequencing, the genomic DNA library is prepared after reaction with bisulfite. In ChIP-seq and RIPseq, the genomic DNA fragments for sequencing were prepared after specific sequence was subjected with trapped antibody as supplement. Among genomic DNA libraries, the RNA and ChIP libraries with less starting materials produce less amounts of DNA extract from histones immunoprecipitation or DNA-binding proteins with restricted sequence coverage.

To prepare DNA and RNA, library kits with semiautomated or fully automated machines were developed in second- or third-generation sequences such as GemCode from 10× genomics, Raindrop's Thunderstorm for all sequencing, cBot for Illumina (Hart et al. 2010), and Ion Torrent platform (Hart et al. 2010). Thus, the cost and workload were reduced by these sequencing kits and machines. The labels with sample tag barcodes of DNA libraries are then done using multiplex identifier (MID) developed by Roche/454 sequencing. While sequencing libraries, it is better to go for maximum resolution for each sequence. Later, the DNA fragments with primers are attached to the solid surface by amplification with PCR microbead emulsion (bp (Check Hayden 2009; Illumina 2015)) or by solid phase (Check Hayden 2009, Ion ChefTM or the Ion OneTouchTm 2, Rothberg et al. 2011) to generate single-stranded DNA molecules to produce efficient sequencing data to detect (Rothberg et al. 2011). Other methods, viz. SOLiD (Roche 454, Life Technologies) and Ion Torrent platforms and solid-phase PCR considered in HiSeq/MiSeq (Illumina) platforms (Check Hayden 2009) were on a rise in this generation.

12.5 Third-Generation Sequencing

In recent times sequencing technologies were emerged to next level, i.e., third-generation sequencing that can reduce cost and simple preparation and sequencing method (Check Hayden 2009; Liu et al. 2012; Aird et al. 2011). Furthermore, single-molecule real-time (SMRT) sequencing was developed by Pacific Biosciences (Liu et al. 2012; Aird et al. 2011). Later, the sequencing using SMRT was carried out on cells with 150,000 ultra-microwells with a molecule of DNA polymerase immobilized at the well bottom by biotin-streptavidin. In this method, the single-strand DNA template and immobilized DNA polymerase coupled and labeled with fluorescent dNTP analogs and the growing stand with incorporated nucleotide is detected. Thus, single molecular trace is converted as template sequence wherein four nucleotides are added in real time simultaneously and measured with high speed and reported 99.3% accuracy with read lengths of 900 bp (Check Hayden 2009). Later, with SMRT bell templates method, the sequences are repeated several times to read longer length and accuracy is improved to >99.999% (108, 109).

The real time processing signal, base calling, quality assessment are done by the genetic analysis system known as Helicos sequencer (a single-molecule fluorescent sequencer) developed by Helicos Biosciences (Travers et al. 2010; Koren et al. 2013).

12.6 The Now to the Future Generation Sequencing

This accounts for Oxford Nanopore Technologies MinION and PromethION (Thompson and Steinmann 2010; Ashley et al. 2010). The MinION is a portable device for sequencing DNA and RNA by connecting to a laptop/computer via a USB port and PromethION is a small benchtop system. The pores are made from proteins such as hemolysin of *Staphylococcus aureus* and are used by Nanopore sequencer (Hickman et al. 2013). Nanopores with ionic conductivity vary when the nucleic acid strand passes through it and the concept was applied on DNA and RNA sequencing (Thompson and Steinmann 2010). While the ionic flow of the molecule translocates through the pore, the sample preparation is simple and read long lengths with Kbp range are developed. Furthermore, no amplification or ligation is carried out before sequencing. In nanopore, DNA translocation and speed optimization are the major drawbacks for generating accurate ionic fluctuations to reduce the high error rates (Ashley et al. 2010; Hickman et al. 2013; Bayley 2015; Wang et al. 2005)). The DNA molecules are then visualized and identified by electron microscopy wherein they are converted to single strands with the help of enzymes. They are labeled by dNTPs with heavy metal atoms that are separated and identified as large black dots, small black dots, and large gray dots in electron microscope generated images. The labeled DNA reads and analysis are carried out by image analysis software using real time sequence data by standard image-based technologies. The sequence with range of 5–50 kb in length reads providing de novo genome assembly and full haplotypes and copy number variants can be analyzed.

12.7 Overview of Genome Assembly

Conventionally, very little volumes of reagents are used for synthesis of oligonucleotides, taken in small columns. Chemicals and solvents are sent through these columns followed by oligonucleotide chain synthesis (Ma et al. 2012a, b). The reaction cycles proceed in keeping with the standard four-step phosphoramidite chemistry, viz., coupling, capping, and oxidization. This is called the solid-phase phosphoramidite chemistry and during the past decade it has been thought of as the "method of choice" for many industrial deoxyribonucleic acid synthesizers (Tian et al. 2009) because of high productivity and a smaller variety of errors. As a result of limitations in chemical action potency, the length of artificial oligonucleotides usually does not exceed "150–200 bases" (Gibson et al. 2010a, b). Assembly is a straightforward task, wherein the given reads correspond to overlapping positions on the reference genome (Gibson et al. 2010a, b; Gibson 2009). While overlapping

is often the case and if they are from overlapping locations on the genome, the reads that have arisen from two totally different copies of same sequence would be wrong and make no sense (Bao et al. 2011).

Before annotating the genome, it is to be checked whether the assembly is prepared. To verify the completeness, associate degreed adjacency of an assembly, N50 statistics are crucial (Yandell and Ence 2012). It is a measure that describes the quality of assembled genomes fragmented in contigs of different length. In addition, the average variety of gaps per frame are checked for good assembly before meeting the minimum standards for submission to "public databases" (Chain et al. 2009). If the scaffold N50 is round the moderate gene length, then it is assumed that close to five hundredth of the genes are going to be contained on one scaffold, thus complete genes along with fragments from the remainder of the genome can offer a large resource for downstream analyses (Cantarel 2008). Though majority of genome annotation pipelines work differently, they share certain features irrespective of the type of sequences to be annotated. The erstwhile expressed sequence tags (ESTs), restricted fragment length polymorphism (RFLP) sequences are aligned to the genome and evidence-driven or ab initio gene predictions are generated (LeProust et al. 2010). With current pipelines targeted on the annotation of protein-coding genes, databases like Ensembl have some capabilities for annotation of non-coding ribonucleic acids (ncRNAs) especially the long non-coding RNA (lncRNA), that are thought of as major regulatory regions (Tripathi et al. 2017; Xu et al. 2017; Mazzoni et al. 2015). Furthermore, this will allow easy detection of novel proteins from classifying the known unknown regions or pseudogenes in the genome (Shidhi et al. 2014).

12.8 Machine Learning in Genome Assembly

Genome assembly tasks belong to the specific class of problems known as non-deterministic polynomial-time hard (NP-hard). To get the optimal solutions for such class of problems in acceptable time (polynomial time) is not feasible. It is well known due to this limitation that different heuristics are applied and there is no guarantee that these solutions are optimal. Hence the practically important genome assembly task poses a computational challenge to provide solutions with maximum reliability. Recently, machine learning methods have been proposed with a view to producing a better solution, computationally more efficient and faster (Padovani de Souza et al. 2018). Survey of literature indicates three stages of applications of machine learning methods, viz. (1) pre-assembly, (2) post-assembly, and (3) auto-assembly. In pre-assembly, the machine learning methodologies have been essentially employed to assist the process by efficient grouping of reads. Angeleri et al. (1999) built a recurrent neural network (RNN) wherein the trained network was deputed to predict the similarities between test (unseen) reads. Constantinescu (2015) also proposed the construction of an RNN-based binning of reads to assist assembly process, similar to Angeleri et al. (1999); while the earlier method used an OLC-based assembler, the current method was with assemblers based

on Greedy and "de Bruijn" strategies. Though this method did not provide any performance improvement, there was a significant improvement in assembly time. Post-assembly machine learning based methods were mainly used for validating genome assemblies. Choi et al. (2008) described a supervised machine learning method employing random forests and decision trees. Their model was able to identify portions of genomic assemblies by splitting contigs into blocks of 500 bp. In addition, Lanc and Emrich (2013) approach based on unsupervised machine learning to validate genome assemblies works in such a way that segments with assembly errors were placed in the same group. Unsupervised clustering algorithms were used and the results obtained were extended to other assemblies. Bodily et al. (2015) presented a method, called ScaffoldScaffolder, aiming to improve the scaffolding step of genomes of diploid heterozygous using ML. Palmer et al. (2010) demonstrated an application of ML during assembly using an OLC-based assembler called Minimus, specifically for identifying read overlaps as true or false, prior to the construction of contigs. They employed several machine learning approaches within WEKA framework. They were able to double the median contig length. Bocicor et al. (2011) proposed a novel solution to the problem of genome assembly using reinforcement learning (RL) wherein they described an assembler that predicts the optimal order of the reads to later perform the assembly. In a given state space, each state corresponds to an ordered combination of reads and the actions which lead to states that represent distinct and overlapped reads were programmed to receive greater rewards. The problem finally amounted to identifying the correct order of reads by maximizing rewards. Zhu et al. presented an assembler based on supervised learning employing SVM, which combines greedy strategy with auxiliary methods to assemble contigs and scaffolds. On the other hand, the SVM based prediction models were able to make decision dynamically whether to continue extending contigs or not. Recently, Nathan Lapierrre et al. have employed a novel deep learning-based methodology for de novo long read scrubbing using approximate alignment. Furthermore, selecting genes using multi-objective machine learning approaches would be made easier for accurate prediction (Dashtban et al. 2018) even as we hope that de novo scrubbing would improve other downstream analyses such as read correction, structural variation detection, etc.

12.9 Applications of Genome Assembly

With the information from deoxyribonucleic acid sequencing used to find genetic variations related to any given disease/phenotype, the most intriguing part would be the use of reference-based alignment methods to predict the variants/SNPs that account for less than 30–50% of the discovered variations (Potamias et al. 2014). In the recent past, with the addition of several robust tools (See Table 12.1), the applications of genome annotation and assembly have taken a big leap starting from manipulation of genetic circuits and metabolic pathways to development of artificial genomes (Tian et al. 2004). Recently, J. Craig Venter Institute performed de novo synthesis and assembly of 1,000,000 nucleotide microorganism genome

Table 12.1 Overview of genome assembly tools

S. no	Tool	Brief	References	URL
1	ALLPATHS-LG	ALLPATHS-LG is based on the Eulerian de Bruijn graph, and is considered to be the most accurate assembler for short NGS reads	Gnerre et al. (2011), Ribeiro et al. (2012)	https://software.broadinstitute.org/allpaths-lg/blog/
2	SOAPdenovo	SOAPdenovo2 implements an algorithm based on the sparse k-mer and reduces the required memory by up to 35 GB for assembly of a human whole genome	Luo et al. (2012)	http://soap.genomics.org.cn/soapdenovo.html
3	Sparse Assembler	SparseAssembler uses the sparse k-mer and drastically reduces the required memory by an order of magnitude	Ye et al. (2012)	https://sites.google.com/site/sparseassembler/
4	String Graph Assembler (SGA)	SGA implements assembly algorithms using the FM-index, a compressed substring index based on the Burrows–Wheeler transform	Simpson (2014), Simpson and Durbin (2012), Nagarajan and Pop (2013), Chikhi and Rizk (2013), Li and Waterman (2003), Melsted and Pritchard (2011), Pell et al. (2012)	https://github.com/jts/sga
5	Maryland Super Read Cabog Assembler (MaSuRCA)	MaSuRCA (hybrid method of OLC and Eulerian de Bruijn graph) compresses overlapping reads into super-reads	Zimin et al. (2013), Myers et al. (2000), Zimin et al. (2014)	https://github.com/alekseyzimin/masurca
6	Meraculous	Meraculous builds a lightweight hash table, in which only high-quality extensions are stored	Berkeley et al. (2011)	https://sourceforge.net/projects/meraculous20/
7	JR-Assembler	The JR-Assembler extends seed reads by overlapping other reads instead of using the de Bruijn graph, and is similar to greedy assemblers	Chu et al. (2013)	http://jr-assembler.iis.sinica.edu.tw/
8	Velvet	Velvet uses bidirectional de Bruijn graphs	Zerbino and Birney (2008)	https://www.ebi.ac.uk/~zerbino/velvet/
9	SPAdes	SPAdes is one of Eulerian de Bruijn graph assemblers, and was designed for single-cell sequencing	Bankevich et al. (2012), Medvedev et al. (2011)	http://cab.spbu.ru/software/spades/

10	ABySS	ABySS is a de novo parallel assembler for short reads that enhances distributed k-mer hash table	Simpson et al. (2009)	http://www.bcgsc.ca/platform/bioinfo/software/abyss
11	Hinge	Hinge is an Open Source OLC (Overlap–Layout–Consensus) assembler that assembles both third- and fourth-generation long read sequences obtained from Oxford nanopore and PacBio	Kamath et al. (2017)	https://github.com/HingeAssembler/HINGE
12	Hierarchical Genome Assembly Process (HGAP)	HGAP is the first long-read assembler developed by Pacific Biosciences that SMRT sequence data	Chin et al. (2013)	https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-in-SMRT-Analysis
13	Canu	Canu is the successor of Celera assembler and assembles long reads obtained from Oxford nanopore and PacBio through adaptive k-mer weighing	Koren et al. (2017)	https://github.com/marbl/canu
14	LightAssembler	LightAssembler is a lightweight assembly algorithm designed and uses graph traversal and simplification modules	Sara El-Metwally et al. (2016)	https://github.com/SaraEl-Metwally/LightAssembler
15	Falcon	Falcon is an Open Source assembler developed by Pacific Biosciences for large genome assembly of diploid organism using SMRT data obtained from PacBio	Chin et al. (2016)	https://github.com/PacificBiosciences/FALCON/
16	HapSVAssembler	HapSVAssembler first builds a consensus sequence, identifies the heterozygous loci of SNPs/SVs, and reconstructs the paternal and maternal haplotypes	Ting (2016)	https://github.com/ythuang0522/HapSVAssembler
17	Ray	Ray is an Open Source short-read assembler that can assemble genomes in parallel interconnected computers using Multiple Passing Interface (MPI)	Boisvert et al. (2010)	https://github.com/sebhtml/ray
18	Short sequence assembly by k-mer search and 3′ read extension (SSAKE)	SSAKE is a de novo assembler that uses greedy extension (“overlap then extend”) approaches for assembling millions of very short DNA sequence reads	Warren et al. (2006)	https://github.com/warrenlr/SSAKE

(continued)

Table 12.1 (continued)

S. no	Tool	Brief	References	URL
19	Short-read assembler based on robust contig extension for genome sequencing (SHARCGS)	SHARCGS is a fast and highly accurate de novo assembler that uses greedy extension algorithms to assemble short read data (25–40 mers). It is also based on "overlap then extend" idea	Dohm et al. (2007)	http://sharcgs.molgen.mpg.de/software/sharcgs.shtml
20	Verified Consensus Assembly by k-mer extension (VCAKE)	VCAKE is a de novo assembler that can assemble millions of short reads obtained from Illumina's Solexa sequencing data even in the presence of sequencing errors. It uses greedy algorithms based on the idea of "overlap then extend"	Jeck et al. (2007)	https://sourceforge.net/projects/vcake/
21	Newbler	Newbler is a de novo sequence assembler that was specifically created for assembling reads produced by 454 GS-series of pyrosequencing. It implements OLC (Overlap–Layout–Consensus) twice to assemble the sequence reads	Margulies et al. (2005), Miller et al. (2010)	https://ngs.csr.uky.edu/Newbler
22	Celera	The Celera assembler is a de novo assembler that was originally developed by Celera genomics to assemble data obtained through Whole Genome Shotgun Sequencing using sophisticated string and graph algorithms based on OLC approach. Its license was made Open Source in 2004. An extended version of Celera assembler that supports combinations of ABI 3730 and 454 FLX reads was launched under the name CABOG (Celera Assembler with the Best Overlap Graph). Similarly, PCbR is another extended version of Celera that supports PacBio reads. The present version of Celera supports any combination of reads (greater that 64 bps) produced by Sanger, 454, Illumina, Oxford nanopore and PacBio platforms	Myers et al. (2000), Miller et al. (2008), Berlin et al. (2015)	https://sourceforge.net/projects/wgs-assembler/files/wgs-assembler/wgs-7.0/

23	PHRAP	Phrap is a de novo assembler. It was originally developed for the assembly of cosmid sequences obtained from large scale cosmid shotgun sequencing in the human genome project. It uses banded Smith–Waterman–Gotoh algorithm and quality scores generated by PHRED basecaller to assemble the sequencing data obtained from shotgun sequencing	De la Bastide and McCombie (2007)	http://www.phrap.org/phredphrapconsed.html
24	QSRA	QSRA is a de novo assembler that assembles short read sequences based on a quality value. It is based on VCAKE algorithm. It is more efficient than the VCAKE algorithm and provides better error handling	Bryant et al. (2009)	http://mocklerlab.org/tools/2
25	TIGR	TIGR is a de novo assembler generally used for assembling data obtained from shotgun sequencing. It uses OLC-based greedy algorithm for assembly based on Smith–Waterman algorithm. A modified version of TIGR is TIGR2 which uses BLAST-like algorithms to compare the fragments	Sutton et al. (1995)	https://www.jcvi.org/tigr-releases-version-20-tigr-assembler
26	Contig Assembly Program 3(CAP3)	CAP3 is a de novo assembler used for small scale assembly. It clips the end regions of the reads having poor quality. The fragments are then assembled using a OLC-based greedy algorithm which uses Smith–Waterman technique to compare and arrange the fragments. A quality score may or may not be used	Huang and Madan (1999)	http://seq.cs.iastate.edu/cap3.html
27	AV454	It is a de novo assembler based on ARACHNE algorithm, it is used to assemble small non-repetitive reads obtained specifically from a mixed population of viruses	Henn et al. (2012)	https://www.broadinstitute.org/viral-genomics/av454
28	Phusion	It is a de novo assembler that assembles data obtained from NGS using a strategy called read clustering. It has been used for the successful genomic assembly of mouse and C. briggsae. Phusion 2 is its newer version	Mullikin (2002)	https://sourceforge.net/projects/phusion2/

(continued)

Table 12.1 (continued)

S. no	Tool	Brief	References	URL
29	Shorty	Shorty is a de novo assembler that has been developed for the assembly of very short microreads (25–30 bps) that are produced by next generation sequencing technologies	Hossain et al. (2009)	https://www3.cs.stonybrook.edu/~skiena/shorty/
30	Taipan	Taipan is a hybrid de novo assembler. It assembles short read fragments based on both greedy algorithms and graph based methods	Schmidt et al. (2009)	https://sourceforge.net/projects/taipan/
31	Mimicking Intelligent Read Assembly (MIRA)	MIRA is a multipass assembler/mapper that can work on data obtained through Sanger, Illumina, 454 and Ion torrent platform. It can be used for carrying out true hybrid assembly using the data produced by different platforms, for mapping data obtained from one platform into the assembly produced by data from another platform and also to carry out simple mapping of data against the consensus sequence of other assemblies. It is useful for assembling reads having lots of repetitive sequences	MIRA Assembler	https://sourceforge.net/p/mira-assembler
32	Minia	Minia is a de novo assembler that requires very low memory to assemble large genomes. It uses de Bruijn graph coupled with bloom filter to assemble short read sequences generally obtained from Illumina. It was able to assemble the whole human genome in 23 hours just by using 5.7 GB of RAM	Chikhi and Rizk (2013)	http://minia.genouest.org/
33	Bowtie 2	Bowtie 2 is a reference guided genome assembler. It is a fast and efficient tool, ideal for aligning 50–1000 bp sequences against a relatively large genome. It can carry out local, gapped, and paired end alignment	Langmead et al. (2009), Langmead and Salzberg (2012)	http://bowtie-bio.sourceforge.net/bowtie2/

34	Low-coverage short-read assembler (LOCAS)	LOCAS is a de novo assembler that assembles short reads from NGS and is ideal for homology guided assembly of eukaryotic genomes and can be used for the detection of novel sequences. SUPERLOCAS is an extended version of LOCAS that provides additional features for resequencing purposes	Klein et al. (2011)	http://ab.inf.uni-tuebingen.de/software/locas/
35	Mapping and Assembly with Quality (MAQ)	MAQ is designed to map scores of terribly short reads accurately to a reference genome by taking into consideration the standard values related to bases	Li et al. (2008a, b)	https://sourceforge.net/projects/maq/files/
36	SHRiMP	SHRiMP is one among the primary alignment programs specifically targeted at SOLiD sequencing information	Nagarajan and Pop (2010), Earl et al. (2011)	Unavailable

and placed it within a host lacking its own genetic machinery, and was ready to with success generate an artificial living cell (Gibson et al. 2010a, b). Another application in advancement could be attributed to microarrays towards development of microarray-derived oligo pools, paving the way for new developments of the standard, efficiency, and automation of microarray-based oligo synthesis and gene assembly (Levy et al. 2007; Caruthers et al. 1987). In addition, the implications of NGS technology became apparent with many assemblers designed to alter the new issues and methods. While tools used heuristics, de Bruijn/OLC-based approaches, in the light of artificial/synthetic genomes, machine learning based approaches will have a big role to play in the development of assembly tools. One important thing, however, would be to efficiently handle large amounts of SRS to check alignment sensitivity and performance. Performance is usually achieved by constructing economical indexes of either the reference genome or the set of reads, permitting the speedy identification of supposed matches that are then refined through a lot of time-intensive algorithms. For example, some programs permit solely some variations between a browse and also the reference genome and apparently do not permit indels. Moreover, the alternative of alignment program and corresponding parameters depend on the precise application: as an example, in SNP discovery it is vital to permit for variations between the reads expected because of sequencing errors, whereas in CHIP-seq experiments, actual or almost-exact alignments are most likely decent.

12.10 Conclusions

The NGS has led to revolutionary advancement in basic and molecular research studies. It has ventured into personalized genomics for community analysis with cost effect huge data generation. While NGS has reduced the cost, manpower, and time, with high error rate and short read length compared to that of Sanger sequencing, there poses a problem to counter the challenges at the algorithmic scale. In this chapter, we have attempted to discuss the challenges and implications of genome assembly while narrating down the NGS chemistry and tools.

References

Adams MD, Kelley JM, Gocayne JD (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Aird D, Ross MG, Chen WS (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12:18

Angeleri E, Apolloni B, de Falco D (1999) DNA fragment assembly using neural prediction techniques. Int J Neural Syst 9(6):523–544

Ashley EA, Butte AJ, Wheeler MT (2010) Clinical assessment incorporating a personal genome. Lancet 375:1525–1535

Balasubramanian S (2015) Solexa sequencing: decoding genomes on a population scale. Clin Chem 61:21–24

Baltimore D (1970) Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226:1209–1211

Bankevich A, Nurk S, Antipov D (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477
Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ (2011) Evaluation of next-generation sequencing software in mapping and assembly. J Hum Genet 56(6):406–414
Barba M, Czosnek H, Hadidi A (2014) Historic perspective, development and applications of next-generation sequencing in plant virology. Viruses 6:106–136
Bayley H (2015) Nanopore sequencing: from imagination to reality. Clin Chem 61:25–31
Berkeley L, Chapman JA, Ho I (2011) Meraculous: de novo genome assembly with short paired-end reads. PLoS One 6:e23501
Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33(6):623–630
Bleidorn C (2016) Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Syst Biodivers 14(1):1–8
Bocicor MI, Czibula G, Czibula IG (2011) A reinforcement learning approach for solving the fragment assembly problem. In 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, pp 191–198
Bodily PM, Fujimoto MS, Ortega C, Okuda N, Price JC, Clement MJ, Snell Q (2015) Heterozygous genome assembly via binary classification of homologous sequence. BMC Bioinf 16(7):S5
Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17(11):1519–1533
Bowtie 2: fast and sensitive read alignment. bowtie-bio.sourceforge.net/bowtie2/index.shtml
Bryant DW, Wong WK, Mockler TC (2009) QSRA—a quality-value guided de novo short read assembler. BMC Bioinf 10(1):69
Bussotti G, Notredame C, Enright A (2013) Detecting and comparing non-coding RNAs in the high-throughput era. Int J Mol Sci 14(8):15423–15458
Callanan J, Stockdale SR, Shkoporov A, Draper LA, Ross RP, Hill C (2018) RNA phage biology in a metagenomic era. Viruses 10(7):386
Cantarel BL (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196
Caruthers MH, Barone AD, Beaucage SL, Dodds DR, Fisher EF, McBride LJ, Matteucci M, Stabinsky Z, Tang JY (1987) Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. Methods Enzymol 154:287–313
Chain PSG, Grafham DV, Fulton RS, FitzGerald MG, Hostetler J et al (2009) Genome project standards in a new era of sequencing. Science 326(5950):236–237
Chaisson MJP, Wilson RK, Eichler EE (2015) Genetic variation and the de novo assembly of human genomes. Nat Rev Genet 16(11):627–640
Check Hayden E (2009) Genome sequencing: the third generation. Nature 457(7231):768–769
Chikhi R, Rizk G (2013) Space-efficient and exact de Bruijn graph representation based on a Bloom filter. Algorithms Mol Biol 8:22
Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Korlach J (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10(6):563–569
Chin CS, Peluso P, Sedlazeck FJ (2016) Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13(12):1050–1054
Choi JH, Kim S, Tang H (2008) A machine-learning approach to combined evidence validation of genome assemblies. Bioinformatics 24(6):744–750
Chu TC, Lu CH, Liu T (2013) Assembler for de novo assembly of large genomes. Proc Natl Acad Sci 110:E3417–E3424
Constantinescu R-I (2015) A machine learning approach to DNA shotgun sequence assembly. Masters Dissertation, University of the Witwatersrand, p 101
Dashtban M, Balafar M, Suravajhala P (2018) Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics 110(1):10–17

De la Bastide M, McCombie WR (2007) Assembling genomic DNA sequences with PHRAP. Curr Protoc Bioinformatics 17:11

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2007) SHARCGS, a fast and highly accurate short-read assembly algorithm for *de novo* genomic sequencing. Genome Res 17(11):1697–1706

Dominguez Del Angel V, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, Vinnere Pettersson O, Lantz H (2018) Ten steps to get started in genome assembly and annotation. F1000Res 7:148

Earl DA, Bradnam K, St John J, Darling A, Lin D, Faas J, Yu HO, Vince B, Zerbino DR, Diekhans M (2011) Assemblathon 1: a competitive assessment of de novo short read assembly method. Genome Res 21:2224–2241

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7(9):1026–1042

El-Metwally S, Zakaria M, Hamza T (2016) LightAssembler: fast and memory-efficient assembly algorithm for high-throughput sequencing reads. Bioinformatics 32(21):3215–3223

Fleischmann RD, Adams MD, White O (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512

Franca LTC, Carrilho E, Kist TBL (2002) A review of DNA sequencing techniques. Q Rev Biophys 35:169–200

Fraser CM, Gocayne JD, White O (1995) The minimal gene complement of Mycoplasma genitalium. Science 270:397–404

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

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM (2010a) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56

Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C (2010b) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7:901–903

Gnerre S, MacCallum I, Przybylski D (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci 108(4):1513–1518

Guzvic M (2013) The history of DNA sequencing. J Med Biochem 32:301–312

Hart C, Lipson D, Ozsolak F (2010) Single-molecule sequencing: sequence methods to enable accurate quantitation. Methods Enzymol 472:407e430

Head SR, Komori HK, LaMere SA (2014) Library construction for next-generation sequencing: overviews and challenges. Biotech 56:61–77

Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Allen TM (2012) Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog 8(3):e1002529

Henson J, Tischler G, Ning Z (2012) Next-generation sequencing and large genome assemblies. Pharmacogenomics 13(8):901–915

Hickman SE, Kingery ND, Ohsumi T (2013) The microglial sensome revealed by direct RNA sequencing. Nat Neurosci 16:1896–1905

Holley RW, Apgar J, Everett GA (1965) Structure of a ribonucleic acid. Science 147:14625

Hossain M, Azimi N, Skiena S (2009) Crystallizing short-read assemblies around seeds. BMC Bioinf 10(Suppl 1):S16

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Illumina (2015) cBot. http://www.illumina.com/documents/products/datasheets/datasheet_cbot.pdf. Last Accessed 27 May 2019

Jeck WR, Reinhardt JA, Baltrus DA (2007) Extending assembly of short DNA sequences to handle error. Bioinformatics 23(21):2942–2944

Kamath GM, Shomorony I, Xia F, Courtade TA, David NT (2017) HINGE: long-read assembly achieves optimal repeat resolution. Genome Res 27(5):747–756

Khost DE, Larracuente AM (2016) Single molecule long read sequencing resolves the detailed structure of complex satellite DNA loci in Drosophila melanogaster. bioRxiv:054155. https://doi.org/10.1101/054155

Kiechle FL, Zhang X (2002) The postgenomic era. Implications for the clinical laboratory. Arch Pathol Lab Med 126:255–262

Klein JD, Ossowski S, Schneeberger K, Weigel D, Huson DH (2011) LOCAS–A low coverage assembly tool for resequencing projects. PLoS One 6(8):e23455

Koren S, Harhay GP, Smith TP (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14:R101

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27(5):722–736

Kulski JK, Kenworthy W, Bellgard M (2005) Gene expression profiling of Japanese psoriatic skin reveals an increased activity in molecular stress and immune response signals. J Mol Med (Berl) 83:964–975

Lanc I, Emrich S (2013) An unsupervised learning approach to assembly validation. In: IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences (ICCABS)

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Lenoir T, Giannella E (2006) Case study. The emergence and diffusion of DNA microarray technology. J Biomed Discov Collab 1:11

LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, Caruthers MH (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res 38:2522–2540

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Venter JC (2007) The diploid genome sequence of an individual human. PLoS Biol 5(10):e254

Li X, Waterman MS (2003) Estimating the repeat structure and length of DNA sequences using L-tuples. Genome Res 13:1916–1922

Li H, Ruan J, Durbin R (2008a) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858

Li R, Li Y, Kristiansen K, Wang J (2008b) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714

Li R, Zhu H, Ruan J (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Law M (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364

Luo R, Liu B, Xie Y (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18

Ma S, Saaem I, Tian J (2012a) Error correction in gene synthesis technology. Trends Biotechnol 30(3):147–154

Ma S, Tang N, Tian J (2012b) DNA synthesis, assembly and applications in synthetic biology. Curr Opin Chem Biol 16(3–4):260–267

Mardis ER (2008) The impact of next generation sequencing technology on genetics. Trends Genet 24:133–141

Margulies M, Egholm M, Altman W (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74:560–564

Mazzoni G, Kogelman LJ, Suravajhala P, Kadarmideen HN (2015) Systems genetics of complex diseases using RNA-sequencing methods. Int J Biosci Biochem Bioinform 5:264

Medvedev P, Pham S, Chaisson M (2011) Paired de Bruijn graphs: a novel approach for incorporating mate pair information into genome assemblers. J Comput Biol 18:1625–1634

Melsted P, Pritchard JK (2011) Efficient counting of k-mers in DNA sequences using a bloom filter. BMC Bioinf 12:333
Metzker ML (2005) Emerging technologies in DNA sequencing. Genome Res 15:1767–1776
Metzker ML (2009) Sequencing technologies: the next generation. Nat Rev Genet 11:31–46
Miller JR, Delcher AL, Koren S (2008) Aggressive assembly of pyrosequencing reads with mates. Bioinformatics 24(24):2818–2824
Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95(6):315–327
Mullikin JC (2002) The phusion assembler. Genome Res 13(1):81–90
Myers EW, Sutton GG, Delcher AL (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204
Nagarajan N, Pop M (2010) Sequencing and genome assembly using next-generation technologies. In: Computational Biology. Humana Press, Totowa, pp 1–17
Nagarajan N, Pop M (2013) Sequence assembly demystified. Nat Rev Genet 14:157–167
Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res 34(19):5623–5630
Padovani de Souza K, Setubal JC, de Carvalho P d LF, AC, Oliveira G, Chateau A, Alves R (2018) Machine learning meets genome assembly. Brief Bioinform. https://doi.org/10.1093/bib/bby072
Palmer LE, Dejori M, Bolanos R (2010) Improving de novo sequence assembly using machine learning and comparative genomics for overlap correction. BMC Bioinf 11(1):33
Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52(4):413–435
Pell J, Hintze A, Canino-Koning R (2012) Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. Proc Natl Acad Sci 109:13272–13277
Peltonen L, McKusick VA (2001) Dissecting human disease in the postgenomic era. Science 291:1224–1229
Phillippy AM (2017) New advances in sequence assembly. Genome Res 27(5):xi–xiii
Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311(5759):392–394
Pop M (2004) Shotgun sequence assembly. Adv Comput 60:193–248
Pop M, Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. Trends Genet 24(3):142–149
Potamias G, Lakiotaki K, Katsila T, Lee MTM, Topouzis S, Cooper DN, Patrinos GP (2014) Deciphering next-generation pharmacogenomics, an information technology perspective. Open Biol 4(7):140071
Ribeiro FJ, Przybylski D, Yin S (2012) Finished bacterial genomes from shotgun sequence data. Genome Res 22:2270–2277
Richter DC, Ott F, Auch AF, Schmid R, Huson DH (2008) MetaSim: a sequencing simulator for genomics and metagenomics. PLoS One 3(10):e3373
Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B et al (2013) Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. PLoS One 8:e66621
Rothberg JM, Hinz W, Rearick TM (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467
Schmidt B, Sinha R, Beresford-Smith B, Puglisi SJ (2009) A fast hybrid short read fragment assembly algorithm. Bioinformatics 25(17):2279–2280
Scholz MB, Lo CC, Chain PSG (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol 23:9–15
Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145
Shidhi PR, Suravajhala P, Nayeema A, Nair AS, Singh S, Dhar PK (2014) Making novel proteins from pseudogenes. Bioinformatics 31(1):33–39

Simpson JT (2014) Exploring genome characteristics and sequence quality without a reference. Bioinformatics 30:1228–1235
Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. Genome Res 22:549–556
Simpson JT, Wong K, Jackman SD (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123
Smith LM, Sanders JZ, Kaiser RJ (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674–679
Stein L (2001) Genome annotation. From sequence to biology. Nat Rev Genet 2:493–503
Stricklin SL, Griffiths-Jones S, Eddy SR (2005) C. elegans noncoding RNA genes. In: WormBook: the online review of C. elegans biology. WormBook, Pasadena, CA; 2005–2018
Sutton GG, White O, Adams MD, Kerlavage AR (1995) TIGR assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci Technol 1:9–19
Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226:1211–1213
Thompson JF, Steinmann KE (2010) Single molecule sequencing with a HeliScope genetic analysis system. Curr Protoc Mol Biol 92:7–10
Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G (2004) Accurate multiplex gene synthesis from programmable DNA microchips. Nature 432:1050–1054
Tian J, Ma K, Saaem I (2009) Advancing high-throughput gene synthesis technology. Mol BioSyst 5:714–722
Ting C-K (2016) A genetic algorithm for diploid genome reconstruction using paired-end sequencing. PLoS One 11(11):e0166721
Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38(15):e159
Tripathi R, Chakraborty P, Varadwaj PK (2017) Unraveling long non-coding RNAs through analysis of high-throughput RNA-sequencing data. Non-Coding RNA Res 2(2):111–118
Venter JC, Adams MD, Myers EW (2001) The sequence of the human genome. Science 291:1304–1351
Wang Y, Yang Q, Wang Z (2005) The evolution of nanopore sequencing. Front Genet 5:1–20
Warren RL, Sutton GG, Jones SJM, Holt RA (2006) Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23(4):500–501
Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738
Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Petterson O, Suh A et al (2017) Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. Genome Res 27:697
Wheeler DA, Srinivasan M, Egholm M (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876
Xu J, Bai J, Zhang X, Lv Y, Gong Y, Liu L, Zhao H, Yu F, Ping Y, Zhang G, Lan Y, Xiao Y, Li X (2017) A comprehensive overview of lncRNA annotation resources. Brief Bioinform 18(2):236–249
Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13(5):329–342
Ye C, Ma ZS, Cannon CH (2012) Exploiting sparseness in de novo genome assembly. BMC Bioinf 13:S1
Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829
Zhang W, Chen J, Yang Y, Tang Y, Shang J, Shen B (2011) A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. PLoS One 6(3):e17915
Zhu T (2003) Global analysis of gene expression using GeneChip microarrays. Curr Opin Plant Biol 6:418–425

Zimin AV, Marcais G, Puiu D (2013) The MaSuRCA genome assembler. Bioinformatics 29:2669–2677

Zimin A, Stevens KA, Crepeau MW (2014) Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics 196:875–890

Recent Advances, Challenges, and Opportunities in Synthetic Genomes

13

Vijai Singh

Abstract

Synthetic genomics is a newly emerging area of synthetic biology that allows us to design, synthesize, and construct a desired genome to create new synthetic life forms. It has been possible because of recently developed advanced cloning and DNA synthesis technologies. Synthetic genomes of bacteria, viruses, and yeast have been chemically synthesized and characterized. Synthetic viruses have been designed and constructed in a way that alters the pathogenicity but maintains antigenicity for vaccination purpose. *Mycoplasma mycoides* JCVI-syn1.0, JCVI-syn2.0, and JCVI-syn3.0 and yeast *Saccharomyces cerevisiae* chromosome have been chemically synthesized and grown as fully functional cells. Synthetic genome technology allows the installation of complete biosynthetic pathways, deletion of non-essential genes, and replacement of the defective chromosome with synthetic chromosome for proper cell functioning. The chapter highlights the recent progress, challenges and opportunities for accelerating synthetic genomics toward biomedical, therapeutic, and industrial applications.

Keywords

Synthetic genome · Therapeutic · Vaccine · Genomics

13.1 Introduction

Synthetic biology is a rapidly growing area which allows us to design and build a new biological system with desired functions. Newly designed systems have been used for a wide range of applications including diagnosis, therapy, vaccine, genome

V. Singh (✉)
Department of Biosciences, School of Science, Indrashil University, Mehsana, Gujarat, India
e-mail: vijai.singh@indrashiluniversity.edu.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_13

editing, and many more (Purnick and Weiss 2009; Khalil and Collins 2010; Singh 2014; Singh et al. 2017, 2018; Khambhati et al. 2019; Bhattacharjee et al. 2019, 2020). In the past decade, a number of synthetic parts, devices, and circuits have been designed and characterized in a broad range of organisms (Singh 2014). At the same time, synthetic biology toolbox has been extended for generating more robust organisms or for building biosynthetic pathways (Gibson et al. 2009; Kosuri et al. 2010; Kim et al. 2012; Gohil et al. 2017; Panchasara et al. 2018).

Regular sticky and blunt end cloning are widely used for the construction of recombinant DNA molecules for enzymes, vaccines, and therapeutic production. However, recently developed cloning and assembly techniques including sequence- and ligation-independent cloning (SLIC) (Li and Elledge 2007), in-fusion PCR cloning (Sleight et al. 2010), Golden Gate cloning or Golden Gate assembly (Engler et al. 2008), Gibson assembly (Gibson et al. 2009), modular cloning (MoClo) system (Weber et al. 2011), and GoldenBraid2.0 (Sarrion-Perdigones et al. 2011) have been extensively used in synthetic biology and metabolic engineering to rapidly construct and exchange a gene in a wide range of organisms. Combinations of cloning and assembly methods have been used for building the complete synthetic genome of *Mycoplasma mycoides* (Gibson et al. 2010).

Viruses are small particles that cause serious human and animal diseases. In order to understand the basic mechanism and also identify virulence factors for pathogenesis, a pressing need has arisen to design and build a synthetic viral genome with desired function, say, for live vaccine development. Cello et al. (2002) chemically synthesized the first poliovirus genome which has been already tested for vaccination purpose. Subsequently, the chemically synthesized ΦX174 phage genome has the ability to infect *Escherichia coli* (Smith et al. 2003). Influenza A virus is a serious pathogen and is responsible for the high rate of mortality in animals including humans. Non-structural protein 1 (NS1) and nuclear export protein (NEP) have been codon de-optimized for humans, which are regarded safe for vaccination purpose (Nogales et al. 2014). Gibson assembly has been used for building of *M. mycoides* JCVI-syn1.0, which is the first reported self-replicating organism on Earth with a synthetic genome (Gibson et al. 2010). Subsequently, *M. mycoides* JCVI-syn1.0 was used as a template for minimization of genome using transposon mutagenesis that resulted in two versions, *M. mycoides* JCVI-syn2.0 and JCVI-syn3.0. Currently, *M. mycoides* JCVI-syn3.0 is one of the smallest self-replicating organisms on Earth with a genome size of 531 kb and harbors 473 known genes (Hutchison et al. 2016).

Following the chemical synthesis of the first *S. cerevisiae* chromosome (synIII) (Annaluru et al. 2014), a number of yeast chromosomes including synII, synV, synVI, synX, and synXII were designed and chemically synthesized (Richardson et al. 2017; Mitchell et al. 2017; Shen et al. 2017; Wu et al. 2017; Xie et al. 2017; Mercy et al. 2017; Zhang et al. 2017). Yeast 2.0 is the world's first synthetic functional genome to be constructed (Pretorius and Boeke 2018). In this chapter, we emphasize recent gene cloning and assembly methods accelerating synthetic genomics research. We highlight the recent advances of the synthetic genome, challenges, and opportunities toward therapeutic, biological, and biotechnological applications.

13.2 Modern Methods for Multiple DNA Assembly

Recent advances in cloning techniques allow rapid, easy, and efficient multiple gene assembly for constructing small gene network, biosynthetic pathway, and complete genome. In the past few decades, gene cloning has been extensively used for a wide range of synthetic biology applications. There are mainly two types of cloning strategies, i.e., conventional and advanced cloning techniques, which are described below.

13.2.1 Conventional Restriction Enzyme-Based Gene Cloning Methods

A restriction endonuclease, also called a restriction enzyme, has the ability to cleave DNA into small fragments by binding upon a recognition site which is known as restriction site (Roberts, 1976). In 1970, the first restriction enzyme (HindIII) was isolated from *Haemophilus influenzae* and characterized (Smith and Wilcox 1970; Kelly and Smith 1970). For the discovery of restriction enzyme HindIII, in 1978, Hamilton O. Smith, Thomas Kelly, and Kent Wilcox were awarded the Nobel Prize in Physiology or Medicine. There are currently 600 commercially available restriction enzymes (Roberts et al. 2007) for gene cloning and molecular biology applications.

Normally, sticky and blunt end cloning are extensively used for assembly of a gene into plasmid for different applications. Sticky end cloning is more precise, faster, and accurate where both the 5′ and 3′ ends of DNA generate an overhanging fragment that can easily combine with same restriction pattern in plasmid DNA for insertion with the help of DNA ligase enzyme (it can seal the nicked gap of two DNA molecules). If two different restriction enzymes are used afterward, it will easily get the directional cloning for gene expression. A blunt end cloning on the other hand can generate a blunt end in both insert and plasmid which can combine together with less efficiency. TA cloning is a method which is commercially available for rapid cloning. The plasmid contains "T" overhang and thus while preparing the insert we can add "A" overhang, in order to easily combine T:A for cloning purpose. These cloning methods allowed us to clone stepwise multiple genes for generating recombinant DNA, synthetic networks, and biosynthetic pathways. Furthermore, the efficiency of TA cloning can vary depending upon the polymerase, and secondly, "A" overhangs can get easily degraded over time which can reduce ligation efficiency.

13.2.2 Advanced Gene Cloning Methods

13.2.2.1 Ligase-Independent Cloning Methods

Ligase-independent cloning (LIC) is a rapid, simple, and easy technique for joining two DNA molecules without the use of restriction enzymes. T4 DNA polymerase is used for generating a 10–15-base single overhang in plasmid and then the DNA molecule is inserted; which tends to easily join together. The technique has been used for inter-Alu fragment gene cloning from the hybrid cell and human cells (Aslanidis and de Jong 1990; Haun et al. 1992). Similarly, a new technique has been developed for multiple fragments joining in a single tube reaction using homologous recombination. In this method, the double-stranded break in the DNA generates single-stranded DNAs (ssDNA) by using exonucleases. Sequence- and ligation-independent cloning (SLIC) is another cloning technique which allows multiple gene assembly in a single reaction using in vitro homologous recombination. SLIC allows for cloning of 5–10 fragments simultaneously with ease and the results are reproducible. Homology search by recombinases allows the repairing of overhangs and gaps. Exonuclease chews back one strand of DNA and creates ssDNA overhangs in both plasmid and insert that can be joined by T4 DNA polymerase in the absence of dNTPs. Both are mixed together and incubated with and without RecA and ATP for catalyzing homologous recombination (Li and Elledge 2007). This is a flexible and versatile platform for rapid multiple gene assembly.

13.2.2.2 In-Fusion PCR Cloning

In-fusion PCR cloning is a homology-based restriction enzyme-free cloning method developed by Clontech Laboratories, USA. It is a simple, rapid, efficient, and sensitive technique to minimize the cost and time of research. In this technique, the primers are designed with 15–20 base homology with the sequence at destination region. Both insert and plasmid are amplified by using high-fidelity DNA polymerase and subsequently purified. Insert and vector (3:1 ratio) are used together with the "in-fusion" enzyme for 15 min at 50 °C and then transformed into competent cells. Clones can be analyzed through blue-white screening. The in-fusion PCR cloning has been used for assembling of BioBricks parts into a plasmid for construction and alteration of parts (Sleight et al. 2010). In the study, cDNA and plasmid could be separately amplified with high-fidelity DNA polymerase. Both the PCR products contained 16-base homology and were assembled with the help of in-fusion enzymes. Ligated product was transformed into the cell and desired clones were observed (Li et al. 2009).

13.2.2.3 Gibson Assembly

Gibson assembly is a single-tube-based isothermal reaction. It is a simple, rapid, and efficient cloning method for joining of multiple DNA molecules. It was developed by Daniel G. Gibson (Vice President of DNA Technology, Synthetic Genomics, Inc.) in collaboration with J. Craig Venter Institute (Gibson et al. 2009). This is among the well-known and widely used cloning methods. It contains three

enzymes including T5 exonuclease (chews back DNA from 5′), DNA polymerase (incorporates nucleotides to fill gaps), and Taq DNA ligase (seals the gap). It is commercially available in the form of a master mixture (Gibson Assembly® Master Mix, New England Biolabs, USA) and can be directly used for multiple DNA fragment assembly. However, it requires homology of 20–40 bases in both insert and plasmid. Both the products are mixed in a single tube along with Gibson master mixture and incubated at 50 °C for 30–60 min. This reaction mixture can be directly transformed into competent cells and screened for the desired clones (Gibson et al. 2009). Gibson assembly has a number of advantages including seamless construction of synthetic and natural genes, biosynthetic pathways, and complete genomes that can also be used for simultaneous assembly of multiple genes (Gibson 2011; Silayeva and Barnes 2018).

13.2.2.4 Golden Gate/MoClo Cloning

Golden Gate cloning or Golden Gate assembly is a simple, rapid, and specific gene cloning method that allows us to simultaneously and directionally assemble DNA molecules. It requires type II restriction enzymes (BsaI, BsmBI, and BbsI) and T4 DNA ligase for assembling an insert into the plasmid (Engler et al. 2008). The type II enzymes cut the dsDNA at a specified distance from recognition sequences. It can create a customized overhang which is not possible with conventional restriction enzymes. The digestion and ligation are accomplished in a single tube, allowing the joining of overlapping sequences in the desired direction. Currently, the Golden Gate assembly is one of the most popular cloning methods for multiple DNA assembly. It is now divided into two tiers. In the first tier, it can construct a single gene by adding in genetic elements including promoter, RBS, CDS, and terminator. And in the second tier, it is combined with several constructs that give rise to multigene constructs (Casini et al. 2015). In order to achieve second-tier assembly, modular cloning (MoClo) system (Weber et al. 2011) and GoldenBraid2.0 (Sarrion-Perdigones et al. 2011) have been developed and extensively used. These Golden Gate assemblies are widely used for the construction of gene network, biosynthetic pathways, and many more.

13.3 Design and Assembly of the Synthetic Genome

In order to understand the molecular mechanism of a cell, a pressing need has arisen to design and create an artificial cell. It can be used to build a completely new genome or to modify a pre-existing living organism's genome. It allows to produce a new form of life with expected features. A number of viruses, bacteria, and eukaryotic chromosomes have been designed and tested for biotechnological, therapeutic, and medical applications.

13.3.1 Design and Construction of Viral Genome

A number of viruses have been designed and chemically synthesized with the help of genetic engineering, synthetic biology, and gene synthesis technology. The major goal of synthetic viruses is to utilize them for medical treatments, reviving organisms, cancer therapy, and genetic toolbox (Wimmer et al. 2009; Guenther et al. 2014). Poliomyelitis, also called polio or infantile paralysis, is a disease caused by poliovirus which belongs to family *Picornaviridae*. It causes muscle weakness and inability to walk. Poliovirus contains single-stranded positive-sense RNA genome and capsid protein. The genome size is about 7500 nucleotides long (Hogle 2002), and it was first isolated by Karl Landsteiner and Erwin Popper in 1909 (Paul 1971), and in 1981, the poliovirus genome was announced (Racaniello and Baltimore 1981; Kitamura et al. 1981). Poliovirus is a well-characterized and useful model system enabling better understanding of RNA viruses.

Viruses are small and have the ability to enter into a cell and produce more infectious particles. Poliovirus binds with the CD155 receptor on human cells that facilitates its entry into the cell and later causes infection. It cannot infect other organisms due to lack of host CD155. Therefore, it is restricted to human infection and causes polio (Mueller and Wimmer 2003; Suzuki 2006). A more serious form of this infection is paralytic poliomyelitis (<1%), where the virus enters into the central nervous system (CNS). It replicates in motor neuron within the spinal cord, motor cortex, and brain stem and causes alteration in motor neuron, leading to permanent or temporary paralysis (Koeller and Shih 2017). In order to control poliovirus infection, recombinant DNA technology has been used to clone poliovirus, wherein the poliovirus RNA genome was introduced into mammalian cells that produced infectious poliovirus (Racaniello and Baltimore 1981). This study has opened up a new avenue for understanding of poliovirus and been considered as a tool for the study of many viruses. In 2002, groundbreaking virus research was done by Cello et al. (2002). They chemically synthesized the first complete poliovirus. This was the first synthetic virus in the world that attracted virologist's attention and interest. It was 7741 bases long and was assembled by the Gene Synthesis Company. A total of 19 markers was incorporated that allowed them to distinguish the synthetic virus from the natural one. Synthetic poliovirus was injected into transgenic mice and was found to have the ability to cause disease. It could replicate, infect, and cause paralysis or death of mice. The synthetic virus was 1/1000 to 1/10000th lethal as the original poliovirus (Couzin 2002; Wimmer and Paul 2011).

The phi X 174 (or ΦX174) bacteriophage (phage) virus that infects *E. coli*, is a single-stranded DNA (ssDNA) virus. The genome is completely sequenced (Sanger et al. 1977) which is 5386 nucleotides long. It is currently used as a model organism for the study of evolution and as a positive control for whole-genome sequencing. It was first used for DNA synthesis in test tube using purified enzymes that could produce all characters of natural phage (Goulian et al. 1967). In 2003, the genome of ΦX174 phage was chemically synthesized and assembled in vitro using oligonucleotides. The genome size was determined to be 5386 nucleotides

long and it has the ability to infect *E. coli*. This could be only made possible due to the advances in DNA synthesis technology and assembly (Smith et al. 2003). Subsequently, the ΦX174 virus particle was assembled in vitro based on structural and in vivo data (Cherwa and Fane 2001). In addition to this, synthetic bacteriophage øX174 genome with highly overlapping coding proteins has been fully decompressed. It was propagated in yeast and has shown to exhibit lytic activity on *E. coli* strains (Jäschke 2012). Currently, phage therapy is being used for controlling multidrug-resistant pathogenic bacteria. For this purpose, phage tail fiber has been modified and evolved for increasing the broad-spectrum activity. *E. coli*-based phage can now target and kill other members of *Enterobacteriaceae* family such as *Yersinia* and *Klebsiella* (Ando et al. 2015; Costa et al. 2018).

Influenza viruses are widely known to cause epidemics and pandemics. It can be prevented by effective vaccination to reduce the transmission and infection (Krammer 2015). The non-structural protein 1 (NS1) and nuclear export protein (NEP) of influenza A virus are important for replication and virulence that can be an attractive target for development of live attenuated vaccine. In a study, researchers have demonstrated codon de-optimization for de novo synthesis of NS1 fragment in mammalian cells. Three different recombinant influenza viruses were generated which were composed of codon-deoptimized synonymous mutations in coding sequences of NS1 and NEP with no alterations in any other activity. The synthetic influenza virus was attenuated in vivo but it retained immunogenic features. This study has welcomed a new possibility for using synthetic virus as a safe vaccine to prevent influenza virus infections (Nogales et al. 2014).

Herpes simplex virus (HSV) is among the highly widespread pathogens globally. It causes oral and genital ulcerations. Recently, herpes simplex virus type 1 (HSV-1) has been synthesized using genomic tools. Yeast transformation-associated recombination was used for cloning of 11 fragments of HSV-1 strain KOS. The reconstituted HSV was later transfected into mammalian cells. Modular assembly technology has been used for incorporating a number of changes in a single gene or changes in two genes simultaneously. It has also been used in genome-wide editing for potential benefits including enhanced functional studies, oncolytic property, and development of a platform for vaccine or therapy (Guenther et al. 2014; Oldfield et al. 2017). This research can be further expanded to enhance the functionality and use as a nano-device for biomedicine and therapeutic applications (Guenther et al. 2014). The building of synthetic virus by constructing viral genome is facing a lot of challenges and issues that need to be resolved. Recent advances in synthetic biology, gene synthesis technology, and microfluidics allow us to accelerate the viral genomics for therapeutic, biomedical, and biotechnological applications.

13.3.2 Design and Construction of Bacterial Genome

In the past decade, increasing data and knowledge of complete genome from a wide range of organisms has allowed us to understand molecular mechanism, biochemistry, physiology, and genomics. It has endowed us to see the defined look

of a bacterial synthetic genome. *Mycoplasma* is the simplest of organisms that has the ability to autonomously grow and self-replicate. It is considered as a model organism for understanding the basic principles of life. *Mycoplasma* belongs to risk group 2 pathogens. The complete genome of *Mycoplasma genitalium* has been sequenced and 525 genes were found. It has smallest size of genomes on Earth. Of these 525 genes, 250 genes are essential, and following the development of transposon mutagenesis, a number of non-essential genes have been identified by altering the associated function of the corresponding gene responsible for growth of *M. genitalium* (Hutchison et al. 2016).

In 2010, Gibson et al. (2010) published the groundbreaking research unveiling the first synthetic life on Earth. They designed, chemically synthesized, and assembled the complete genome of *Mycoplasma mycoides* using Gibson assembly for rapid cloning of genes toward the construction of a complete genome. The synthetic genome was 1.08 megabase pair long with 901 known genes. It was transplanted into *M. capricolum* which served as a recipient cell to create a new *M. mycoides* JCVI-syn1.0 cell. It was the first synthetic viable cell. In order to verify the synthetic genome, they incorporated a "watermark" for easy recognition. The newly created cells were self-replicating and depicted similar phenotypic properties as natural *M. mycoides* strain. The genome design of *M. mycoides* JCVI-syn1.0 is based on completely published genome sequences of *M. mycoides* subspecies Capri GM12 (Lartigue et al. 2007; Lartigue et al. 2009; Benders et al. 2010). The genome was designed based on homologous joining of overlapping sequences (80 bp) of 1080-bp size in three steps. First 1080-bp sequences were constructed by overlapping oligonucleotides that could be recombined in a set of ten to generate 109 10-kb assemblieds. These sequences were then used for recombining in a set of ten to generate 11 100-kb assemblies. Finally, all 11 fragments were recombined for creating a complete synthetic genome of 1.08 mega bp through in vivo homologous recombination in yeast. The genome was verified through multiplex PCR using 11 primer pairs. The 48 clones were screened and it was found that one clone could produce 11 positive amplicons. Errors that occurred during the assembly process were corrected by re-assembly of genome in yeast in order to make functional cells with the desired phenotype. Synthetic genome-containing cells are self-replicating and have the ability of logarithmic growth with 60 min of cell division time (Gibson et al. 2010). The major goal of building the genome is to determine the molecular and biological function of each gene and also to create microbial cell factory for future bio-refinery (Kondo et al. 2013; Gustavsson and Lee 2016; Hutchison et al. 2016).

After the creation of synthetic *M. mycoides* JCVI-syn1.0 cells, a need has further arisen for better and in-depth understanding of cells. Hutchison et al. (2016) used transposon mutagenesis to alter the gene function and applied the design-build-test (DBT) cycle for refinement of cell growth. The complete DBT represents how the genome can be designed, built, and tested. After each cycle, the genome was assembled into yeast cell and then tested by transplanting the synthetic genome into recipient *M. capricolum* cells. Global Tn5 transposon was used and the non-essential genes from genome were removed. After three DBT cycles, the non-

essential genes were deleted which led to the improvement of eight more segments, thus giving the second synthetic cell of *M. mycoides* JCVI-syn2.0. The genome size was estimated to be 576 kb with 517 known genes that could encode 478 protein-coding and 38 RNA-coding regions. The synthetic JCVI-syn2.0 cell can multiply in 92 min. Hutchison et al. (2016) have used JCVI-syn2.0 cells as a template for further minimization of genome. They improved the transposon mutagenesis method and revealed a quasi-essential gene which is required for robust cell growth. They repeated DBT thrice and were able to synthesize *M. mycoides* JCVI-syn3.0 by removing 42 genes from the genome. The genome size was reduced to 531 kb with 473 genes that encoded 439 protein-coding and 35 RNA-coding sequences. This is one of the smallest living autonomously replicating cells in nature. This cell has almost all essential genes required for synthesis and processing of biological macromolecules which are essential for cell reproduction and growth. Surprisingly, 149 genes were found with unknown functions (Hutchison et al. 2016).

As compared to JCVI-syn1.0 cells, in JCVI-syn3.0 cells, 428 genes were missed and about 63% were found to have unknown functions including 73 mobile elements and genes required for restriction and lipoproteins. This cell has been optimized for growth in complete media, but it lacks genes responsible for transport and metabolism which are majorly through carbon utilization. The JCVI-syn3.0 genome has been divided into four categories: (1) genes required for gene expression (195 genes) and genome maintenance (34 genes), (2) components of cell membrane (84 genes), (3) genes for protein metabolism (81 genes), and (4) genes that have not been assigned to any biological function (79 genes). Though the basic functions of these genes are known, out of 79 genes, 55 genes have unknown functions. These genes share homology with genes present in other organisms and their functions need to be explored in the future. The JCVI-syn3.0 cells have showed similar colony morphology with JCVI-syn1.0 cells, but they have smaller colony size and 180 min of division time which is quite longer than that of JCVI-syn1.0. It may be possible because of deletion of a number of genes from genome (Hutchison et al. 2016).

JCVI-syn3.0 is currently considered as a versatile platform for discovery of basic functions of life and can be used for expanding the whole-genome design and synthesis. This approach can be applied in other medically important microorganisms for development of an attenuated live vaccine or therapeutic applications. This approach can also be applied for designing cells by installing complete biosynthetic pathways or altering genetic codes or expanding genetic codes for microbial cell factory.

13.3.3 A Blueprint for Synthetic Eukaryotic Genome

The Human Genome Project (HGP) was one of the biggest genome projects that was started with the aim to identify and map sequences present in the genome from both physical and functional viewpoint. The completion of project was declared in 2003. It was reported to cover 99% of human genome (euchromatin) with 99.99% accuracy. A total of 2.91 billion base pair consensus sequences were generated from

the euchromatic region using whole-genome shotgun sequencing techniques. The human genome holds a wide range of benefits including early disease diagnostics, genetics, personalized medicine, therapy, and human evolution (Venter et al. 2001). A number of chromosomal defects, mutations, and non-functional regions have been discovered that need to be addressed in order to improve human life.

Recently, the synthetic genome is gaining more attention and interest in the scientific community. The complete genome sequencing of yeast *S. cerevisiae* has revealed the size to be 12,068 kilobases long, with 5885 protein-encoding genes. There are about 140 genes responsible for ribosomal RNA, 40 genes for small nuclear RNA, and 275 genes for tRNA. Yeast has 16 chromosomes (Goffeau et al. 1996) and is one of the best eukaryotic model systems. Yeast is easy to manipulate, enabling understanding of its physiology, genetics, and evolution (Botstein et al. 1997). Why do we need to chemically synthesize genome? The possible answer to this question would be the statement of Richard Feynman "What I cannot create, I do not understand." The large-scale de novo design of genome has allowed to systematically understand genotype to phenotype, genome structure, organization, and many more (van der Sloot and Tyers 2017). Recent advances in DNA synthesis technology, assembly, and testing methods allow us to rapidly build a designer genome or rewire genome. An international consortium has been formed for "The Synthetic Yeast Genome Project (Sc2.0)" with the aim to chemically synthesize all 16 chromosomes of yeast. Different chromosome numbers were assigned to different teams. In order to build a complete yeast synthetic genome, a stepwise approach is required from chromosome by chromosome (van der Sloot and Tyers 2017; Wang et al. 2018).

In 2014, the first yeast synthetic chromosome (synIII) was completely chemically synthesized, assembled, and tested by replacing the same wild-type chromosome in yeast. The size of the chromosome (synIII) was 272,871 base pairs based on the wild type of *S. cerevisiae* chromosome III (316,617 bp). Annaluru et al. (2014) replaced the TAG/TAA stop codon; deleted the sub-telomeric portion, tRNAs, introns, transposons, and silent mating loci; and inserted loxPsym into the new chromosome. This study provides a foundation for building a designer eukaryotic chromosome. Subsequently, in the Sc2.0 project, more chromosomes such as synII, synV, synVI, synX, and synXII were targeted which correspond to one-third of *S. cerevisiae* genome (Richardson et al. 2017; Mitchell et al. 2017; Shen et al. 2017; Wu et al. 2017; Xie et al. 2017; Mercy et al. 2017; Zhang et al. 2017).

A 770-kilobase (kb) synthetic yeast chromosome II (synII) has been designed, constructed, and characterized. The characterization has been done using phenomics, transcriptomics, replication, and chromosome segregation analysis. By both SCRaMbLE (synthetic chromosome recombination and modification by loxP-mediated evolution) and complementation assay, researchers have corrected the origin of growth defect at 37 °C (in glycerol media) which tends to mis-regulate at high-osmolarity glycerol response (Shen et al. (2017). In another study, chromosome synV was designed and 536,024 bp were de novo synthesized by correcting isolated synV that could match with designed sequences using CRISPR-Cas9 system and also integrative co-transformation. The complete genome was edited

in 22 steps and high fitness under various culture conditions was determined. Ring synV derivative was constructed which is fully functional in *S. cerevisiae* in different conditions, but during meiosis, it shows lower spore viability. This study provides a foundation for understanding the genomic rearrangement, ring chromosome disorders, and evolution (Xie et al. 2017).

A synthetic yeast Sc2.0 chromosome VI (synVI) has been designed, assembled, and characterized. There is a mitochondrial defect in synVI strain mapped identical to coding changes in *PRE4* which encodes an essential proteasome subunit that is responsible for high accumulation of Pre4. Mitchell et al. (2017) have also identified the alteration in HIS2 transcription start because of tRNA deletion and insertion of loxPsym site during the proteomic analysis. They reduced up to 6% of synthetic genome (Mitchell et al. 2017).

Mutations and defects in wild type of chromosomes need to be corrected by designing and building of synthetic chromosome. Yeast synthetic chromosome X (synX) has been chemically synthesized having a size of 707,459 bp. Strong fitness has been observed under a variety of conditions. Researchers have developed an efficient mapping approach which is called as PCRTag mapping (PoPM) that helps to identify defects. A growth defect was corrected and mapped to a recorded sequence in FIP1 and loxPsym that may affect ATP2 PoPM promoter function. The study provided a powerful tool for correction of synthetic genome that can also be useful for mapping of genotype and phenotype (Wu et al. 2017). A 976,067-bp-long linear yeast chromosome XII (synXII) has been designed and chemically synthesized based on yeast wild-type chromosome XII. A defect in minor growth has been detected in synXII due to the deletion of tRNA genes for which the phenotype was later rescued. The ribosomal DNA has been used for species identification using barcoding procedures (Zhang et al. 2017).

Recently, the eukaryotic yeast genome (Sc2.0) has been designed and chemically synthesized that is 1.1 megabases long and 8% reduced in size as compared to the wild-type yeast genome. Insertion, deletion, and alteration have been done in synthetic genome for robustness. Richardson et al. (2017) used BioStudio for eukaryotic chromosome design that allows hassle-free modifications from nucleotide to genome scale. In order to completely design a synthetic genome (Sc2.0), individual teams have designed and chemically synthesized chromosomes and incorporated them into a single yeast strain by "endoreduplication intercross." The chemically synthesized Sc2.0 genome can be useful for structure, function, and evolution study with bottom-up design approach (Richardson et al. 2017).

Compared to the wild-type yeast chromosome, the new yeast synthetic genome (Sc2.0) is highly conserved with deleted number of repeated sequences. It has successfully modified thousands of locations that may affect cell physiology and growth rate. Mercy et al. (2017) have determined a three-dimensional (3D) pattern of Sc2.0, wherein they observed that repeats can lead to smoother contact pattern, tractable conformations, and large-scale organization of genome which remains unaffected in synthetic chromosomes (Mercy et al. 2017). Yeast 2.0 is the world's first synthetic functional genome, the construction of which has turned into reality (Pretorius and Boeke 2018). In summary, the synthetic yeast chromosomes have

opened up new avenues for a better understanding of life, genetics, gene function, correction in the defective chromosome, and modification as per choice and sites.

13.4 Conclusion and Future Remarks

Recent advances in gene synthesis technology allow us to rapidly synthesize a gene in a cost-effective manner. A number of cloning and assembly methods have been developed for building of gene network and biosynthetic pathways and fine-tuning the gene by bringing about some precision in either a single gene and/or a complete genome. Currently, a number of viruses including poliovirus, influenza virus, HSV, and phages have been designed, chemically synthesized, and tested in a way that alters their pathogenicity, yet could be used as live attenuated vaccines and in therapeutic applications. In 2010, the first synthetic self-replicating *M. mycoides* JCVI-syn1.0 genome was designed and chemically synthesized that could show similar phenotypes to wild type (Gibson et al. 2010). The yeast *S. cerevisiae* is the first eukaryotic model organism to have its chromosome designed, chemically synthesized, and tested. A number of mutations or chromosomal defects have been corrected or improved by designing, building, transplanting, and testing in higher eukaryotic organisms. There are still a number of key challenges and issues that need to be overcome in the near future, and more research work needs to be done for synthetic genomics to come with full potential in biomedical, therapeutic, and industrial applications.

Competing Interests There is no competing interest.

References

Ando H, Lemire S, Pires DP et al (2015) Engineering modular viral scaffolds for targeted bacterial population editing. Cell Syst 1:187–196

Annaluru N, Muller H, Mitchell LA et al (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344:55–58

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

Benders GA, Noskov VN, Denisova EA et al (2010) Cloning whole bacterial genomes in yeast. Nucleic Acids Res 38:2558–2569

Bhattacharjee G, Khambhati K, Singh V (2019) RNA-guided CRISPR-Cas9 system for removal of microbial pathogens. In: Tripathi V, Kumar P, Tripathi P, Kishore A (eds) Microbial genomics in sustainable agroecosystems. Springer, Singapore, pp 227–248

Bhattacharjee G, Mani I, Gohil N et al (2020) CRISPR technology for genome editing. In: Faintuch J, Faintuch S (eds) Precision medicine for investigators, practitioners and providers. Academic Press, London, pp 59–69

Botstein D, Chervitz SA, Cherry M (1997) Yeast as a model organism. Science 277:1259–1260

Casini A, Storch M, Baldwin GS et al (2015) Bricks and blueprints: methods and standards for DNA assembly. Nat Rev Mol Cell Biol 16:568–576

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

Cherwa JE, Fane BA (2001) Microviridae: microviruses and gokushoviruses. eLS. https://doi.org/10.1002/9780470015902.a0000781.pub2
Costa AR, Brouns SJ, Nobrega FL (2018) Complete genome sequences of two T4-like *Escherichia coli* bacteriophages. Genome Announc 6:e00586–e00518
Couzin J (2002) Virology. Active poliovirus baked from scratch. Science 297:174–175
Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647
Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349–361
Gibson DG, Smith HO, Hutchison CA et al (2010) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7:901–903
Gibson DG, Young L, Chuang RY et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 genes. Science 274:563–567
Gohil N, Panchasara H, Patel S et al (2017) Book review: recent advances in yeast metabolic engineering. Front Bioeng Biotechnol 5:71
Goulian M, Kornberg A, Sinsheimer RL (1967) Enzymatic synthesis of DNA, XXIV. Synthesis of infectious phage phi-X174 DNA. Proc Natl Acad Sci U S A 58:2321–2328
Guenther CM, Kuypers BE, Lam MT et al (2014) Synthetic virology: engineering viruses for gene delivery. Wiley Interdiscip Rev Nanomed Nanobiotechnol 6:548–558
Gustavsson M, Lee SY (2016) Prospects of microbial cell factories developed through systems metabolic engineering. Microb Biotechnol 9:610–617
Haun RS, Serventi IM, Moss J (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. Bio Techniques 13:515–518
Hogle J (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. Annu Rev Microbiol 56:677–702
Hutchison CA, Chuang RY, Noskov VN et al (2016) Design and synthesis of a minimal bacterial genome. Science 351:aad6253
Jäschke A (2012) Genetically encoded RNA photoswitches as tools for the control of gene expression. FEBS Lett 586:2106–2111
Kelly TJ, Smith HO (1970) A restriction enzyme from *Haemophilus influenzae*. J Mol Biol 51:393–409
Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11:367–379
Khambhati K, Bhattacharjee G, Singh V (2018) Current progress in CRISPR-based diagnostic platforms. J Cell Biochem 120(3):2721–2725
Kim H, Han H, Ahn J et al (2012) Shotgun DNA synthesis' for the high throughput construction of large DNA molecules. Nucleic Acids Res 40:e140
Kitamura N, Semler BL, Rothberg PG et al (1981) Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature 291:547–553
Koeller KK, Shih RY (2017) Viral and prion infections of the central nervous system: radiologic-pathologic correlation: from the radiologic pathology archives. Radiographics 37:199–233
Kondo A, Ishii J, Hara KY et al (2013) Development of microbial cell factories for bio-refinery through synthetic bioengineering. J Biotechnol 163:204–216
Kosuri S, Eroshenko N, Leproust EM et al (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat Biotechnol 28:1295–1299
Krammer F (2015) Emerging influenza viruses and the prospect of a universal influenza virus vaccine. Biotechnol J 10:690–701
Lartigue C, Glass JI, Alperovich N et al (2007) Genome transplantation in bacteria: changing one species to another. Science 317:632–638
Lartigue C, Vashee S, Algire MA et al (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 325:1693–1696
Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256

Li YL, Li H, Li AN et al (2009) Cloning of a gene encoding thermostable cellobiohydrolase from the thermophilic fungus *Chaetomium thermophilum* and its expression in *Pichia pastoris*. J Appl Microbiol 106:1867–1875

Mercy G, Mozziconacci J, Scolari VF et al (2017) 3D organization of synthetic and scrambled chromosomes. Science 355:eaaf4597

Mitchell LA, Wang A, Stracquadanio G et al (2017) Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science 355:eaaf4831

Mueller S, Wimmer E (2003) Recruitment of nectin-3 to cell-cell junctions through trans-heterophilic interaction with CD155, a vitronectin and poliovirus receptor that localizes to alpha(v)beta3 integrin-containing membrane microdomains. J Biol Chem 278:31251–31260

Nogales A, Baker SF, Ortiz-Riaño E et al (2014) Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. J Virol 88:10525–10540

Oldfield LM, Grzesik P, Voorhies AA et al (2017) Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. Proc Natl Acad Sci U S A 114:E8885–E8894

Panchasara H, Patel S, Gohil N et al (2018) Biosynthetic pathway for production of renewable biofuels. In: Mittal V (ed) Biofuels. Central West Publishing, Orange, pp 63–86

Paul JR (1971) A history of poliomyelitis. (Yale studies in the history of science and medicine). Yale University Press, New Haven

Pretorius IS, Boeke JD (2018) Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. FEMS Yeast Res 18(4):foy032. https://doi.org/10.1093/femsyr/foy032

Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422

Racaniello V, Baltimore D (1981) Cloned poliovirus complemenatry DNA is infectious in mammalian cells. Science 214:916–919

Richardson SM, Mitchell LA, Stracquadanio G et al (2017) Design of a synthetic yeast genome. Science 2017(355):1040–1044

Roberts RJ (1976) Restriction endonucleases. CRC Crit Rev Biochem 4:123–164

Roberts RJ, Vincze T, Posfai J et al (2007) REBASE–enzymes and genes for DNA restriction and modification. Nucleic Acids Res 35:D269–D270

Sanger F, Air GM, Barrell BG et al (1977) Nucleotide sequence of bacteriophage ΦX174 DNA. Nature 265:687–695

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

Shen Y, Wang Y, Chen T et al (2017) Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science 355:eaaf4791

Silayeva O, Barnes AC (2018) Gibson assembly facilitates bacterial allelic exchange mutagenesis. J Microbiol Methods 144:157–163

Singh V (2014) Recent advancements in synthetic biology: current status and challenges. Gene 535:1–11

Singh V, Braddick D, Dhar PK (2017) Exploring the potential of genome editing CRISPR-Cas9 technology. Gene 599:1–18

Singh V, Gohil N, Ramírez García R et al (2018) Recent advances in CRISPR-Cas9 genome editing technology for biological and biomedical investigations. J Cell Biochem 119(1):81–94

Sleight SC, Bartley BA, Lieviant JA et al (2010) Designing and engineering evolutionary robust genetic circuits. J Biol Eng 4:12

van der Sloot A, Tyers M (2017) Synthetic genomics: rewriting the genome chromosome by chromosome. Mol Cell 66:441–443

Smith HO, Hutchison CA, Pfannkoch C et al (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A 100:15440–15445

Smith HO, Wilcox KW (1970) A restriction enzyme from *Haemophilus influenza* I. Purification and general properties. J Mol Biol 51:379–391

Suzuki Y (2006) Ancient positive selection on CD155 as a possible cause for susceptibility to poliovirus infection in simians. Gene 373:16–22
Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351
Wang L, Jiang S, Chen C et al (2018) Synthetic genomics: from DNA synthesis to genome design. Angew Chem Int Ed Engl 57:1748–1756
Weber E, Engler C, Gruetzner R et al (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765
Wimmer E, Mueller S, Tumpey TM et al (2009) Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat Biotechnol 27:1163–1172
Wimmer E, Paul AV (2011) Synthetic poliovirus and other designer viruses: what have we learned from them? Annu Rev Microbiol 65:583–609
Wu Y, Li BZ, Zhao M et al (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. Science 355:eaaf4706
Xie ZX, Li BZ, Mitchell LA et al (2017) "Perfect" designer chromosome V and behavior of a ring derivative. Science 355:eaaf4704
Zhang W, Zhao G, Luo Z et al (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. Science 355:eaaf3981

Expansion of the Genetic Code

14

Nisarg Gohil, Gargi Bhattacharjee, and Vijai Singh

Abstract

The genetic code consists of codons that code for 20 canonical amino acids and 3 termination signals and is an indispensable element of life. It controls all cellular activities through expression of genes. Advances in the genetic engineering techniques have enabled the incorporation of novel non-canonical amino acids (ncAAs) into proteins to confer distinct physical and chemical features. Currently, over 150 ncAAs have been site-specifically introduced into proteins with high efficiency using an orthogonal tRNA synthetase pair in various organisms. In this chapter, genetic code engineering and its advances, challenges and future opportunities have been explored towards biological, biomedical, therapeutic, industrial and biotechnological applications.

Keywords

Non-canonical amino acids · Genetic code expansion · Genetic code engineering · Orthogonal tRNA · Orthogonal synthetase

14.1 Introduction

Amino acids, the building blocks of protein, play a crucial role in growth, development and homeostasis of prokaryotes and eukaryotes. Not just constricted to that, these organic compounds are also associated with multiple regulatory processes including signalling, immunity, reproduction, blood flow, metabolic regulation, body composition, pigmentation, oxidative stress response and many more (Wu

N. Gohil · G. Bhattacharjee · V. Singh (✉)
Department of Biosciences, School of Science, Indrashil University, Mehsana, Gujarat, India
e-mail: vijai.singh@indrashiluniversity.edu.in

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_14

2010). It is quite astonishing that for all of the complex activities of life, it just requires 20 standard amino acids! Amongst these, nine are essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine) that cannot be synthesized by organisms and therefore must be obtained from diet. The remaining are nonessential amino acids. A question which arises here is that is it possible to expand the genetic code in a way that would allow the addition of non-canonical (also known as non-standard or unnatural) amino acids (ncAAs) into proteins, rendering the proteins distinct biological, chemical or physical properties? Interestingly, it is of course possible.

The ncAAs can be site-specifically incorporated into proteins by exercising novel orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair. The chemically aminoacylated novel tRNA delivers new ncAAs by accessing nonsense (most typically the rarest amber stop codon (UAG)) or four base codons as an insertion signal (Fig. 14.1). For successful incorporation of ncAAs, the designed tRNA must work proficiently without being recognized by existing aaRS. Similarly, new aaRS also solely functions by recognizing and aminoacylating orthogonal tRNAs (Wang et al. 2001; Chin 2014; Wang 2017; Mukai et al. 2017). The incorporation of ncAAs into the native protein allows the extension of genetic code for increasing the

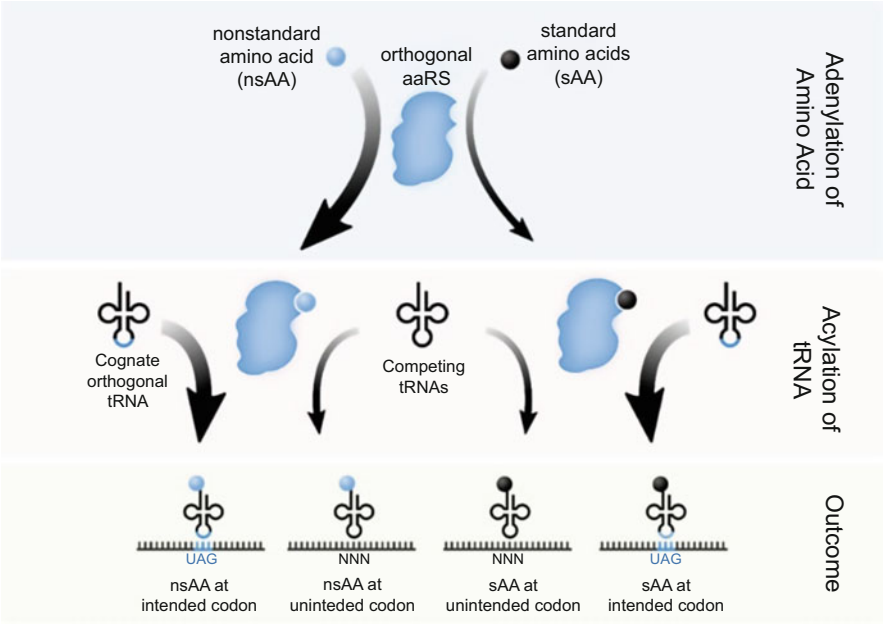


Fig. 14.1 Expansion of genetic code via incorporation of ncAAs. It requires an orthogonal pair of tRNA-aaRS and a unique codon. This pair must not be recognized by the endogenous tRNA or aaRS. The orthogonal aaRS aminoacylates the ncAA on the cognate tRNA in response to encountering a nonsense or quadruplet codon. The ncAA should be transported into cytoplasm when added to growth medium (Figure adopted from Arranz-Gibert et al. (2019) © MDPI)

protein diversity (Acevedo-Rocha and Budisa 2016; Kubyshkin and Budisa 2017). To date, genetic code has been successfully extended using more than 150 different ncAAs in *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, zebrafish, mouse and mammalian cell line for various applications (Wang 2017; Arranz-Gibert et al. 2018). In this chapter, we explore genetic code engineering and its progress, recent advances and opportunities.

14.2 Orthogonal tRNA/Aminoacyl-tRNA Synthetase Pairs

An orthogonal tRNA-aaRs pair must be particularly defined with respect to host as aaRs and tRNAs in different organisms may vary. It should not be cross-talked with internal transcriptional and translational cellular machinery. It also means that a tRNA-aaRs pair that is orthogonal for one host may not be cross-talked to others (Chin 2014, 2017). For example, *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj* TyrRS)-$tRNA_{CUA}$ pair is orthogonal for *E. coli* and other bacteria but not for eukaryotes (Wang et al. 2001). Similarly, *E. coli*-derived TyrRS (*Ec* TyrRS)-$tRNA_{CUA}$ and leucyl-tRNA synthetase (*Ec* LeuRS)-$tRNA_{CUA}$ pairs are orthogonal for eukaryotic cells but not for *E. coli* (Chin 2014, 2017). Methanogens (most commonly *Methanosarcina barkeri* and *M. mazei*)-derived pyrrolysyl-tRNA synthetase (PylRS)-$tRNA_{CUA}$ pair is orthogonal for bacteria and eukaryotes (Wan et al. 2014).

Ideal orthogonal tRNAs should have high levels of read-through of nonsense codons. However, most of them have lower tendencies, which can be increased by about 10- to 1000-fold in the presence of a cognate synthetase (Chin 2017). Incorporation of ncAAs has been made possible by engineering orthogonal pairs using positive and negative selection approach or discovering orthogonal synthetases that have the ability to incorporate structurally allied ncAAs (Voloshchuk and Montclare 2010). Advanced parallel positive selections using sequencing and statistical analysis have boosted the screening methods to find appropriate synthetases that can incorporate ncAAs with high efficiency and fidelity (Zhang et al. 2017). The orthogonal pairs can also be optimized by directed evolution approach (Ellefson et al. 2014; Zhang et al. 2017) towards industrial and biotechnological applications.

14.3 Expanding the Genetic Code in Organisms

14.3.1 Extending Genetic Code in Bacteria

In 2001, the genetic code was expanded in *E. coli* by incorporating the synthetic amino acid *O*-methyl-L-tyrosine at the amber nonsense codon (UAG) using an archaebacterial orthogonal *Mj* TyrRS-$tRNA^{Tyr}$ pair (Wang et al. 2001). Thereafter, the same pair has also been employed in diverse bacteria such as *Shigella*

flexneri, *Salmonella enterica*, *Yersinia ruckeri* and *Acinetobacter baylyi* for the incorporation of 3-nitro-L-tyrosine or 3-iodo-L-tyrosine (Tack et al. 2016); *Salmonella typhimurium* LT2 for a photocrosslinker, benzoyl-phenylalanine (Gan et al. 2016); *Pseudomonas syringae* for *p*-benzoyl-*L*-phenylalanine or *p*-azido-*L*-phenylalanine (Raber 2015); *Mycobacterium tuberculosis* for *p*-azidophenylalanine, *p*-benzoylphenylalanine, *p*-nitrophenylalanine, *p*-iodophenylalanine and *p*-boronophenylalanine (Wang et al. 2010); and *Streptomyces venezuelae* ATCC 15439 for *p*-iodo-*L*-phenylalanine and *p*-azido-*L*-phenylalanine (He et al. 2016) ncAAs in response to the same codon.

Many other orthogonal pairs such as *S. cerevisiae* phenylalanyl-tRNA synthetase (*Sc* PheRS)-tRNA^{Phe} (Furter 1998; Kwon et al. 2003, 2006; Kwon and Choi 2016), *Sc* tryptophanyl-tRNA synthetase (*Sc* TrpRS)-tRNA^{Trp} (Chatterjee et al. 2013), *Methanosarcina acetivorans* TyrRS (*Ma* TyrRS)-tRNA^{Tyr} (Ikeda-Boku et al. 2013), *Methanococcus maripaludis* phosphoseryl-RS (*Mm* SepRS)-*Mj*tRNA^{Cys} (Park et al. 2011; Rogerson et al. 2015; Gan et al. 2016), *Desulfitobacterium hafniense* PylRS-tRNA (*Dh* PylRS)-tRNA^{Pyl} (Katayama et al. 2012), *Methanosarcina barkeri* or *mazei* PylRS (*M* PylRS)-tRNA^{Pyl} (Blight et al. 2004; Neumann et al. 2008; Yanagisawa et al. 2008; Zeng et al. 2014; Fan et al. 2015) and *Ec* TyrRS-tRNA^{Tyr} (Iraha et al. 2010; Italia et al. 2017) have been discovered and utilized efficiently in various bacteria for inserting ncAAs.

Apart from the nonsense codons, unique quadruplet codons (e.g. UAGN, AGGA) have also been used for insertion of ncAAs (Neumann et al. 2010; O'Donoghue et al. 2012; Wang et al. 2014; Chatterjee et al. 2014; Lee et al. 2017). Evolution of tRNA is another proven strategy for efficient incorporation (Neumann et al. 2008; Fan et al. 2015). Ribosome evolution can also enable multiple ncAAs (Neumann et al. 2010). List of successfully expanded genetic codes in bacteria has been given in Table 14.1.

14.3.2 Extending Genetic Code in Yeast

Besides bacteria, genetic codes have also been expanded in many yeast cells, including *S. cerevisiae* (Chin et al. 2003; Wu et al. 2004; Chen et al. 2007; Hancock et al. 2010); a methylotrophic yeast, *Pichia pastoris* (Young et al. 2009); a human pathogenic fungi, *Candida albicans* (Palzer et al. 2013); and a fission yeast, *Schizosaccharomyces pombe* (Shao et al. 2015) (see Table 14.1). *Ec* TyrRS-tRNA^{Tyr} pair is the most efficiently exercised orthogonal pair in yeast (Chin et al. 2003; Chen et al. 2007; Young et al. 2009; Palzer et al. 2013; Shao et al. 2015).

Nonsense-mediated mRNA decay (NMD), a cellular mRNA surveillance mechanism of eukaryotes, degrades mRNAs containing a premature termination codon (Aznarez et al. 2018). Inhibition of NMD can increase ncAA incorporation efficiency in yeast (Wang 2017). Unlike bacteria, yeast cells have not been explored much with other codons (nonsense and quadruplet) and different orthogonal pairs. The future studies may include other yeasts, novel orthogonal pairs and unique quadruplet codons.

Table 14.1 Expanded genetic codes in bacteria and yeast

Host organisms	Orthogonal pair	Codon	ncAA(s)	Reference
Bacteria				
Escherichia coli	*Mj* TyrRS-tRNA^{Tyr}	UAG	*O*-Methyl-*L*-tyrosine	Wang et al. 2001
		UAGN	*p*-Acetylphenylalanine	Chatterjee et al. 2014
		AGGA	*p*-Azido-*L*-phenylalanine and *N*6-(2-propynyloxy)carbonyl)-*L*-lysine	Neumann et al. 2010
			p-Azidophenylalanine, *O*-propargyl-tyrosine, *p*-acetylphenylalanine and *p*-benzoylphenylalanine	Lee et al. 2017
	Sc PheRS-tRNA^{Phe}	UAG	*p*-Fluoro-phenylalanine	Furter 1998
			p-Bromophenylalanine	Kwon et al. 2006
		UUU	L-3-(2-Naphthyl)alanine	Kwon et al. 2003
		UUG	2-Naphthylalanine	Kwon and Choi 2016
	Sc TrpRS-tRNA^{Trp}	UAG	3-(1-Naphthyl)-alanine, 1-methyl-tryptophan, 3-benzothienyl-alanine and 6-methyl-tryptophan	Chatterjee et al. 2013
	Ma TyrRS-tRNA^{Tyr}	UAG	3-Azidotyrosine	Ikeda-Boku et al. 2013
	Mm SepRS-*Mj* tRNA^{Cys}	UAG	Phosphoserine	Rogerson et al. 2015
	Dh PylRS-tRNA^{Pyl}	UAG	Lysine derivatives	Katayama et al. 2012
	M PylRS-tRNA^{Pyl}	UAG	Synthetic *L*-pyrrolysine	Blight et al. 2004
			N^{ε}-(Tert-butyloxycarbonyl)-L-lysine and N^{ε}-allyloxycarbonyl-L-lysine	Yanagisawa et al. 2008
		AGG	N^{ε}-Alloc-lysine	Zeng et al. 2014
			L-Homoarginine and *L*-*N*6-(1-iminoethyl)lysine	Mukai et al. 2015
		AGGA	N^{ε}-(Tert-butyloxycarbonyl)-*L*-lysine and N^{ε}-cyclopentyloxycarbonyl-*L*-lysine	O'Donoghue et al. 2012

(continued)

Table 14.1 (continued)

Host organisms	Orthogonal pair	Codon	ncAA(s)	Reference
E. coli Δ(*Ec* TyrRS-tRNA^{Tyr})	*Ec* TyrRS-tRNA^{Tyr}	UAG	3-Iodo-*L*-tyrosine and 3-azido-*L*-tyrosine	Iraha et al. 2010
E. coli Δ(*Ec* TrpRS-tRNA^{Trp})	*Ec* TrpRS-tRNA^{Trp}	UGA	5-Hydroxytryptophan, 5-bromotryptophan, 5-methylthiopropyl, 5-azidotryptophan, 5-propargyloxytryptophan, 5-aminotryptophan and azido-lysine	Italia et al. 2017
Enteropathogenic *E. coli*	*M* PylRS-tRNA^{Pyl}	UAG	(((3-(3-Methyl-3H-diazirin-3-yl)propamino)carbonyl)-N^{ε}-*L*-lysine) and N^{ε}-((((1R,2R)-2-azidocyclopentyl)oxy)carbonyl)-*L*-lysine	Lin et al. 2011
Shigella sp.				
Salmonella sp.				
S. flexneri	*Mj* TyrRS-tRNA^{Tyr}	UAG	3-Nitro-*L*-tyrosine and 3-iodo-*L*-tyrosine	Tack et al. 2016
S. enterica				
Yersinia ruckeri				
Acinetobacter baylyi				
S. typhimurium LT2	*Mj* TyrRS-tRNA^{Tyr} and *Mm* SepRS-*Mj* tRNA^{Cys}	UAG	Benzoyl-phenylalanine	Gan et al. 2016
Pseudomonas syringae	*Mj* TyrRS-tRNA^{Tyr}	UAG	*p*-Benzoyl-*L*-phenylalanine and *p*-azido-*L*-phenylalanine	Raber 2015
Mycobacterium tuberculosis	*Mj* TyrRS-tRNA^{Tyr}	UAG	*p*-Azidophenylalanine, *p*-benzoylphenylalanine, *p*-nitrophenylalanine, *p*-iodophenylalanine and *p*-boronophenylalanine	Wang et al. 2010
Streptomyces venezuelae ATCC 15439	*Mj* TyrRS-tRNA^{Tyr}	UAG	*p*-Iodo-*L*-phenylalanine and *p*-azido-*L*-phenylalanine	He et al. 2016
Synechococcus elongatus	*M* PylRS-tRNA^{Pyl}	UAG	N^{ε}-Propargyl-*L*-lysine and N^{ε}-boc-*L*-lysine	Chemla et al. 2017
Bacillus cereus	*M* PylRS-tRNA^{Pyl}	UAG	N^{ε}-(Tertbutoxycarbonyl)-*L*-lysine, N^{ε}-allyloxycarbonyl-*L*-lysine and N^{ε}-prop-2-ynyloxycarbonyl-*L*-lysine	Luo et al. 2016

Yeast				
Saccharomyces cerevisiae	*Ec* TyrRS-tRNA^{Tyr}	UAG	*p*-Benzoylphenylalanine, *p*-azidophenylalanine, *p*-acetylphenylalanine, *p*-propargyloxyphenylalanine and *p*-methoxyphenylalanine	Chen et al. 2007
			p-Acetyl-*L*-phenylalanine, *p*-benzoyl-*L*-phenylalanine, *p*-azido-*L*-phenylalanine, *O*-methyl-*L*-tyrosine and *p*-iodo-*L*-phenylalanine	Chin et al. 2003
	Ec LeuRS-tRNA^{Leu}	UAG	*O*-Methyl tyrosine, α-aminocaprylic acid and *o*-nitrobenzyl cysteine	Wu et al. 2004
	M PylRS-tRNA^{Pyl}	UAG	N^{ε}-((2-Propynyloxy)carbonyl)-*L*-lysine, N^{ε}-((1-(6-nitrobenzo(*d*)(1,3)dioxol-5-yl)ethoxy)carbonyl)-*L*-lysine, N^{ε}-((2-(3-methyl-3*H*-diazirin-3-yl)ethoxy)carbonyl)-*L*-lysine, N^{ε}-acetyl-*L*-lysine and N^{ε}-trifluoroacetyl-*L*-lysine	Hancock et al. 2010
Pichia pastoris	*Ec* TyrRS-tRNA^{Tyr} and *Ec* LeuRS-tRNA^{Leu}	UAG	p-Benzoylphenylalanine, p-azidophenylalanine, p-(propargyloxy)-phenylalanine, p-methoxyphenylalanine, p-iodophenylalanine, 4,5-dimethoxy-2-nitrobenzylserine, 3-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid	Young et al. 2009
Candida albicans	*Ec* TyrRS-tRNA^{Tyr}	UAG	*p*-Azido-*L*-phenylalanine	Palzer et al. 2013
Schizosaccharomyces pombe	*Ec* TyrRS-tRNA^{Tyr}	UAG	Azido-phenylalanine	Shao et al. 2015

Mj Methanocaldococcus jannaschii (formerly *Methanococcus jannaschii*), *Sc Saccharomyces cerevisiae*, *Ma Methanosarcina acetivorans*, *Mm Methanococcus maripaludis*, *Dh Desulfitobacterium hafniense*, *M Methanosarcina barkeri* or *mazei*, *Ec Escherichia coli*

14.3.3 Extending Genetic Code in Mammalian Cells

The cutting-edge research in the area of synthetic biology has made it possible to incorporate a number of ncAAs into the mammalian cells which consequently inculcates novel characters into their existing proteins. The basic understanding of genetic code expansion has helped to site-specifically introduce ncAAs within the proteins which is generated by the mammalian system (Dumas et al. 2015; Xiao and Schultz 2016; Nödling et al. 2019). Just a few sets of orthogonal synthetases and their cognate tRNAs are sufficient to incorporate a whole set of structurally diverse amino acids. For instance, $tRNA^{Pyl}$ (pyrrolysyl-tRNA) derived from archaea species *M. barkeri* and *M. mazei* is orthogonal in both *E. coli* and mammalian cells. It is inherently capable of identifying the UAG codon and thus this tRNA does not require to be engineered. It is possible to engineer its complementary tRNA synthetase (PylRS) in *E. coli* and thereafter utilize the engineered PylRS mutants to add ncAAs into mammalian cells.

Quite a few endogenous *Ec* tRNA-aaRS pairs have been implemented as orthogonal pairs in the mammalian system of which the most prominent ones include *E. coli* leucine, tryptophan and tyrosine pairs (Vargas-Rodriguez et al. 2018). But prior to utilizing them to serve their designated function, it is crucial to eliminate their inherent ability to recognize the canonical amino acid and instead decode a blank codon. Also technically it is infeasible to evolve the orthogonal pairs in mammalian cells which have given their low transfection potency. It is better to engineer the synthetases in *E. coli* (Italia et al. 2017, 2018) or yeast (Chin et al. 2003; Liu et al. 2007; Wang et al. 2007) to enable effortless screening of numerous mutant libraries.

Multiple approaches have been suggested to facilitate the expansion of genetic code which includes translational control through amber suppression, light-mediated induction and small-molecule-induced control (Nödling et al. 2019). Translational control through means of amber suppression is much quicker as compared to the conventional inducible system (tetracycline-controlled transactivation) that is used for mammalian protein synthesis (Gossen et al. 1995; Das et al. 2016), since the lag time invested in translation and folding of protein of interest is much shorter than the cumbersome processes that involved transcription, mRNA processing, translation and folding. Additionally, one can easily halt the further addition of ncAAs just by eliminating the targeted amino acid from the medium and subsequently the translation process will get turned off. However, a drawback of using amber suppression for translation control is that its response is quite delayed (several minutes to hours). The duration right from the addition of ncAAs into the medium until its uptake, expression and protein folding is time-intensive. Apart from this, in order to turn off the production of novel proteins, it is necessary to deplete all of ncAAs present in growth medium; however, the functional characteristics of newly synthesized protein will still prevail until all of the already formed proteins get used up or degrade. It basically means that the response to switching off the protein function predominantly depends on its half-life.

Another approach is to use light for bringing in distinctive features within ncAAs thereby modifying protein function. The ncAA harbouring a photolabile group (photocaged) can be used to either activate or silence a protein function (Wu et al. 2004; Lemke et al. 2007). A synthetic photocaged amino acid is inserted in place of a functionally important amino acid leading to inactivation of protein. Upon irradiation, the photocage gets removed and the protein gains its function again. Several ncAAs including photocaged tyrosines (Hino et al. 2005; Valentin-Hansen et al. 2014; Murray et al. 2016), cysteines (Köhrer et al. 2004; Xiao et al. 2013; Nguyen et al. 2014; Zheng et al. 2017; Baumdick et al. 2018), lysines (Ren et al. 2015; Gramespacher et al. 2017) and perhaps serine (Lemke et al. 2007) have been employed for the regulation of cell-signalling cascades, enzymatic reactions and localization of proteins within the cells (Young and Schultz 2018; UniProt Consortium 2018). An upper hand of using light-mediated control over translational control is that the earlier approach allows highly ordered spatiotemporal resolution, so that the protein expression even in the subcellular locations can be controlled by illuminating even the smallest cellular regions through sharp focusing of light beam, which is almost unattainable through the translational approach.

Besides the translational and light-mediated control, certain small molecules can be used to bioorthogonally replace the masking or protecting groups over resident amino acid situated within deep tissue spaces where light beam seems unapproachable; on the other hand, translational regulation limits the spatiotemporal control. A number of examples that support the use of bioorthogonal reactions in intracellular regulations include Staudinger reactions (Luo et al. 2016), Diels-Alder reactions (Li et al. 2014a; Fan et al. 2016; Zhang et al. 2016), palladium-catalysed propargyl removal (Li et al. 2014b) and 1,3-dipolar cyclo additions (Ge et al. 2016) to name a few. Whilst all of the three approaches are remarkable in one way or another, almost all of the studies reported so far are based on proof of concept. The extension of theoretical knowledge into practical value shall open up newer avenues for in-depth functional studies of proteins in higher organisms.

14.4 Conclusion and Future Remarks

Twenty diverse amino acids are genetically expressed by organisms to produce proteins. Each of the constituent amino acid is responsible for a protein's size, shape and function. Advancements in contemporary molecular biology techniques have helped greatly in broadening the genetic code beyond the 20 standard amino acids encoded by 61 designated triplet codons. Modifications in the backbone of existing translation elements have helped impart distinct characters by replacing one or more ncAAs, leading to expression of mutant proteins. This way it is now possible to analyse the role of each and every amino acid in a protein structure. Keeping pace with the current demand for generating novel proteins with unprecedented functions, the insistence to develop proteins carrying ncAAs is at the forefront of synthetic biology. Over 150 amino acids with distinctive structural and chemical features have already been expressed in many different organisms. In the future, a

series of permutation and combination between modified orthogonal tRNA and its cognate synthetase pairs may fuel the hopes to extend this knowledge into advanced applications such as biocontainment of strains with rewired genomes, production of viral vaccines and synthesis of efficient cargoes to facilitate gene therapy and to decipher many other applications associated with protein-host interactions. It has the potential to decode and utilize the rare data from the genetic code, whilst the strides towards extending the genetic code will bestow bountiful opportunities to explore latent aspects of the language of life.

Acknowledgements The authors gratefully acknowledge the Gujarat State Biotechnology Mission (GSBTM Project ID: 5LY45F) for financial support. The authors also thank Indrashil University for providing infrastructure and support.

References

Acevedo-Rocha CG, Budisa N (2016) Xenomicrobiology: a roadmap for genetic code engineering. Microb Biotechnol 9:666–676

Arranz-Gibert P, Patel JR, Isaacs FJ (2019) The role of orthogonality in genetic code expansion. Life 9:58

Arranz-Gibert P, Vanderschuren K, Isaacs FJ (2018) Next-generation genetic code expansion. Curr Opin Chem Biol 46:203–211

Aznarez I, Nomakuchi TT, Tetenbaum-Novatt J et al (2018) Mechanism of nonsense-mediated mRNA decay stimulation by splicing factor SRSF1. Cell Rep 23:2186–2198

Baumdick M, Gelléri M, Uttamapinant C et al (2018) A conformational sensor based on genetic code expansion reveals an autocatalytic component in EGFR activation. Nat Commun 9:3847

Blight SK, Larue RC, Mahapatra A et al (2004) Direct charging of tRNA CUA with pyrrolysine *in vitro* and *in vivo*. Nature 431:333–335

Chatterjee A, Lajoie MJ, Xiao H et al (2014) A bacterial strain with a unique quadruplet codon specifying non-native amino acids. Chembiochem 15:1782–1786

Chatterjee A, Xiao H, Yang PY et al (2013) A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli*. Angew Chem Int Ed 52:5106–5109

Chemla Y, Friedman M, Heltberg M et al (2017) Expanding the genetic code of a photoautotrophic organism. Biochemistry 56:2161–2165

Chen S, Schultz PG, Brock A (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. J Mol Biol 371:112–122

Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83:379–408

Chin JW (2017) Expanding and reprogramming the genetic code. Nature 550:53–60

Chin JW, Cropp TA, Anderson JC et al (2003) An expanded eukaryotic genetic code. Science 301:964–967

Das AT, Tenenbaum L, Berkhout B (2016) Tet-on systems for doxycycline-inducible gene expression. Curr Gene Ther 16:156–167

Dumas A, Lercher L, Spicer CD et al (2015) Designing logical codon reassignment–expanding the chemistry in biology. Chem Sci 6:50–69

Ellefson JW, Meyer AJ, Hughes RA et al (2014) Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat Biotechnol 32:97–101

Fan C, Xiong H, Reynolds NM et al (2015) Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. Nucleic Acids Res 43:e156

Fan X, Ge Y, Lin F et al (2016) Optimized tetrazine derivatives for rapid bioorthogonal decaging in living cells. Angew Chem Int Ed 55:14046–14050

Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. Protein Sci 7:419–426

Gan Q, Lehman BP, Bobik TA et al (2016) Expanding the genetic code of salmonella with non-canonical amino acids. Sci Rep 6:39920

Ge Y, Fan X, Chen PR (2016) A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. Chem Sci 7:7055–7060

Gossen M, Freundlieb S, Bender G et al (1995) Transcriptional activation by tetracyclines in mammalian cells. Science 268:1766–1769

Gramespacher JA, Stevens AJ, Nguyen DP et al (2017) Intein zymogens: conditional assembly and splicing of split inteins via targeted proteolysis. J Am Chem Soc 139:8074–8077

Hancock SM, Uprety R, Deiters A et al (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J Am Chem Soc 132:14819–14824

He J, Van Treeck B, Nguyen HB et al (2016) Development of an unnatural amino acid incorporation system in the actinobacterial natural product producer *Streptomyces venezuelae* ATCC 15439. ACS Synth Biol 5:125–132

Hino N, Okazaki Y, Kobayashi T et al (2005) Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. Nat Methods 2:201–206

Ikeda-Boku A, Ohno S, Hibino Y et al (2013) A simple system for expression of proteins containing 3-azidotyrosine at a pre-determined site in *Escherichia coli*. J Biochem 153:317–326

Iraha F, Oki K, Kobayashi T et al (2010) Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in *Escherichia coli* for genetic code expansion. Nucleic Acids Res 38:3682–3691

Italia JS, Addy PS, Wrobel CJ et al (2017) An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol 13:446–450

Italia JS, Latour C, Wrobel CJ et al (2018) Resurrecting the bacterial tyrosyl-tRNA Synthetase/tRNA pair for expanding the genetic code of both *E. coli* and eukaryotes. Cell Chem Biol 25:1304–1312

Katayama H, Nozawa K, Nureki O et al (2012) Pyrrolysine analogs as substrates for bacterial pyrrolysyl-tRNA synthetase *in vitro* and *in vivo*. Biosci Biotechnol Biochem 76:205–208

Köhrer C, Sullivan EL, Rajbhandary UL (2004) Complete set of orthogonal 21st aminoacyl-tRNA synthetase-amber, ochre and opal suppressor tRNA pairs: concomitant suppression of three different termination codons in an mRNA in mammalian cells. Nucleic Acids Res 32:6200–6211

Kubyshkin V, Budisa N (2017) Synthetic alienation of microbial organisms by using genetic code engineering: why and how? Biotechnol J 12:1600097

Kwon I, Choi ES (2016) Forced ambiguity of the leucine codons for multiple-site-specific incorporation of a noncanonical amino acid. PLoS One 11:e0152826

Kwon I, Kirshenbaum K, Tirrell DA (2003) Breaking the degeneracy of the genetic code. J Am Chem Soc 125:7512–7513

Kwon I, Wang P, Tirrell DA (2006) Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. J Am Chem Soc 128:11778–11783

Lee BS, Kim S, Ko BJ et al (2017) An efficient system for incorporation of unnatural amino acids in response to the four-base codon AGGA in *Escherichia coli*. Biochim Biophys Acta Gen Subj 1861:3016–3023

Lemke EA, Summerer D, Geierstanger BH et al (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. Nat Chem Biol 3:769–772

Li J, Jia S, Chen PR (2014a) Diels-Alder reaction–triggered bioorthogonal protein decaging in living cells. Nat Chem Biol 10:1003–1005

Li J, Yu J, Zhao J et al (2014b) Palladium-triggered deprotection chemistry for protein activation in living cells. Nat Chem 6:352–361

Lin S, Zhang Z, Xu H et al (2011) Site-specific incorporation of photo-cross-linker and bioorthogonal amino acids into enteric bacterial pathogens. J Am Chem Soc 133:20581–20587
Liu W, Brock A, Chen S et al (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Methods 4:239–244
Luo J, Liu Q, Morihiro K et al (2016) Small-molecule control of protein function through Staudinger reduction. Nat Chem 8:1027–1034
Mukai T, Lajoie MJ, Englert M et al (2017) Rewriting the genetic code. Annu Rev Microbial 71:557–577
Mukai T, Yamaguchi A, Ohtake K et al (2015) Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. Nucleic Acids Res 43:8111–8122
Murray CI, Westhoff M, Eldstrom J et al (2016) Unnatural amino acid photo-crosslinking of the IKs channel complex demonstrates a KCNE1: KCNQ1 stoichiometry of up to 4: 4. elife 5:e11815
Neumann H, Peak-Chew SY, Chin JW (2008) Genetically encoding Nε-acetyllysine in recombinant proteins. Nat Chem Biol 4:232–234
Neumann H, Wang K, Davis L et al (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464:441–444
Nguyen DP, Mahesh M, Elsässer SJ et al (2014) Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. J Am Chem Soc 136:2240–2243
Nödling AR, Spear LA, Williams TL (2019) Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells. Essays Biochem 63:237–266
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 586:3931–3933
Palzer S, Bantel Y, Kazenwadel F et al (2013) An expanded genetic code in Candida albicans to study protein-protein interactions *in vivo*. Eukaryot Cell 12:816–827
Park HS, Hohn MJ, Umehara T (2011) Expanding the genetic code of Escherichia coli with phosphoserine. Science 333:1151–1154
Raber SD (2015) Investigations into the molecular mechanisms of bacterial pathogen-host interactions: construction of a dual plasmid system for incorporation of unnatural amino acids into *Pseudomonas syringae* pv. tomato DC3000. MS thesis, University of Nebraska
Ren W, Ji A, Ai HW (2015) Light activation of protein splicing with a photocaged fast intein. J Am Chem Soc 137:2155–2158
Rogerson DT, Sachdeva A, Wang K (2015) Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol 11:496–503
Shao N, Singh NS, Slade SE et al (2015) Site specific genetic incorporation of azidophenylalanine in *Schizosaccharomyces pombe*. Sci Rep 5:17196
Tack DS, Ellefson JW, Thyer R et al (2016) Addicting diverse bacteria to a noncanonical amino acid. Nat Chem Biol 12:138–140
UniProt Consortium (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515
Valentin-Hansen L, Park M, Huber T et al (2014) Mapping substance P binding sites on the neurokinin-1 receptor using genetic incorporation of a photoreactive amino acid. J Biol Chem 289:18045–18054
Vargas-Rodriguez O, Sevostyanova A, Söll D et al (2018) Upgrading aminoacyl-tRNA synthetases for genetic code expansion. Curr Opin Chem Biol 46:115–122
Voloshchuk N, Montclare JK (2010) Incorporation of unnatural amino acids for synthetic biology. Mol BioSyst 6:65–80
Wan W, Tharp JM, Liu WR (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta 1844:1059–1070
Wang F, Robbins S, Guo J et al (2010) Genetic incorporation of unnatural amino acids into proteins in *Mycobacterium tuberculosis*. PLoS One 5:e9354
Wang K, Sachdeva A, Cox DJ et al (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. Nat Chem 6:393–403

Wang L (2017) Engineering the genetic code in cells and animals: biological considerations and impacts. Acc Chem Res 50:2767–2775
Wang L, Brock A, Herberich B et al (2001) Expanding the genetic code of *Escherichia coli*. Science 292:498–500
Wang W, Takimoto JK, Louie GV et al (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies. Nat Neurosci 10:1063–1072
Wu G (2010) Functional amino acids in growth, reproduction, and health. Adv Nutr 1:31–37
Wu N, Deiters A, Cropp TA et al (2004) A genetically encoded photocaged amino acid. J Am Chem Soc 126:14306–14307
Xiao H, Chatterjee A, Choi SH et al (2013) Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. Angew Chem Int Ed 52:14080–14083
Xiao H, Schultz PG (2016) At the interface of chemical and biological synthesis: an expanded genetic code. Cold Spring Harb Perspect Biol 8:a023945
Yanagisawa T, Ishii R, Fukunaga R et al (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode Nε-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. Chem Biol 15:1187–1197
Young DD, Schultz PG (2018) Playing with the molecules of life. ACS Chem Biol 13:854–870
Young TS, Ahmad I, Brock A et al (2009) Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. Biochemistry 48:2643–2653
Zeng Y, Wang W, Liu WR (2014) Towards reassigning the rare AGG codon in *Escherichia coli*. Chembiochem 15:1750–1754
Zhang G, Li J, Xie R et al (2016) Bioorthogonal chemical activation of kinases in living systems. ACS Cent Sci 2:325–331
Zhang MS, Brunner SF, Huguenin-Dezot N et al (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods 14:729–736
Zheng S, Fan X, Wang J et al (2017) Dissection of kinase isoforms through orthogonal and chemical inducible signaling cascades. Chembiochem 18:1593–1598

Expanding the Potential of CRISPR-Cas9 Technology for Crops Improvement

15

Pankaj Kumar, Anshu Alok, and Jitesh Kumar

Abstract

Cereal crops like rice, maize, barley, and wheat are the major staple food crops, and their nutrition and processing quality requires a dedicated long-term breeding program. However, genome editing (GE) could be a potential approach for the targeted improvement of cultivated crop plants via specifically selected alternation in their genome. Various GE approaches have been previously used, which make use of zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short identical and inverted repeats CRISPR/Cas9 (CRISPR-associated protein 9). Genome modification methods have been implemented for the desired trait improvement in numerous crops as well in some gymnosperms and angiosperm. CRISPR/Cas9-based methods have provided a great capacity to genetic engineering due to their competency, simplicity, and the ability to adopt in comparison to other GE techniques. These methods have provided capacity to gain an exact and deeper understanding of GE. But in plant research, crop system specifically aims at genetic improvement by targeting desired biosynthesis pathways. This urges to develop new technologies for specific mutagenesis and accurate genome editing of the key genes involved in different biosynthetic pathways, and targeted control of their gene expression at transcriptional or post-transcriptional level.

P. Kumar
Central Potato Research Institute, Shimla, Himachal Pradesh, India

A. Alok
University Institute of Engineering and Technology, Panjab University, Chandigarh, India

J. Kumar (✉)
Center of Innovative and Applied Bioprocessing, Department of Biotechnology, (DBT), S.A.S. Nagar, Mohali, Punjab, India

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_15

Keywords
Cereal crops · CRISPR/Cas · Genome editing · Nucleases · Synthetic biology

15.1 Introduction

Nearly more than billions of populations are hungry in the world due to deficiency of food. The reasons are very complicated, and it may be due to poverty, management, and crop losses. Environment plays a significant role in yield and crop loss due to involvement of various stresses such as biotic and abiotic. Biotic stresses include the loss due to the attack of various pests and diseases. Pests are the major cause for diseases in plants during pre- and post-harvesting. Abiotic stress is also responsible for the loss of the crop yield due to various environmental effects. Worldwide, the total annual loss of crop yield is approximately 20% due to abiotic stresses. Nowadays, food demand of the world is also increasing day-by-day due to the growing population, which will expectedly gain 34% increment by 2050 (http://www.fao.org). Population boom causes a tremendous decline in fertile land and different resources. This has led to immediate demand for a sustainable crop improvement and management. The expansions of new and advance genotypes, with several improved agronomic and nutritional qualities, through classical plant breeding approach, are very challenging and take lots of time. Unfortunately, a number of other technologies are available for crop improvement at present but they suffer from lack of precision and are highly random events (Miflin 2000). Thus, there is a huge need to have precise tools for crop improvement that targets the desired DNA manipulation of the endogenous genes for important traits. Current progress in the NGS (next-generation sequencing), proteomics, and metabolomics has facilitated the identification of functional genes. Such advancements in functional genomics have opened the doors in exploring the allelic variations and functional genes for crop improvement. Furthermore, a powerful site-specific method of genome editing CRISPR-Cas9 has opened a new horizon to accelerate crop improvement through rapid and precise modifications of these alleles and functional genes.

The emergence of CRISPR technology takes advantages of ZFNs and TALENs, which are very popular as a novel approach in recent years from 2011 to 2015. But nowadays, a genome editing CRISPR/Cas tool has advantages over RNAi and used commonly for knockdown or silencing. CRISPR/Cas9 targets the segment of DNA that is difficult to manipulate by using previous technology like RNAi or other techniques with accuracy and simplicity. CRISPR/Cas9 targets the gene of interest with the uses of a RNA sequence, which act as guide RNA (SgRNA) that is very specific to the target gene with an average length of ~100 nucleotides (nt). The (PAM) protospacer region is the segment of host DNA where the Cas9 nuclease recognizes and cleaves targeted gene 3 nucleotide upstream of it (Tsai et al. 2015). The Cas9 endonuclease introduces double-strand breaks (DSBs), which further activates the DNA repair mechanism in the specific gene by utilizing endogenous repair system through non-homologous end joining (NHEJ) and homologous recombination (HR). Nowadays, CRISPR has wide applications in plant system biology and is commonly

Table 15.1 List of selected genes in which CRISPR/Cas9 approaches have been used

Wheat	TaMLO	Wang et al. (2014)
Rice	*Os*SWEET	Zhou et al. (2014)
	*Os*Waxy	Ma et al. (2015)
	Gn1a, DEP1, GS3, IPA1	Li et al. (2016)
Zea mays	LIG1	Svitashev et al. (2015)
	ALS	Char et al. (2017)
Potato	GBSS	Anderson et al. (2018)
Solanum lycopersicum	ANT1	Čermák et al. (2015)
	SIAGO7	Brooks et al. (2014)
Cotton	GhclA1	Chen et al. (2017)
	GhVP	
Sorghum color	DsRed	Jiang et al. (2013)
Populus tomentosa	PDS	Fan et al. (2015)
Glycine max	ALS1	Li et al. (2015)
Brassica oleracea	BolC.GA4	Lawrenson et al. (2015)
Lotus japonicus	SNF	Wang et al. (2016)
Camelina sativa	FAD2	Jiang et al. (2017)

used to produce important agricultural traits and products via targeting different genes belonging to different pathways (Tables 15.1 and 15.2). Nowadays, there are many plants, which are used for genome editing purpose through CRISPR/Cas9 for, e.g., tomatoes, corn, and wheat. The latest example is corn, which is edited to have quality of drought resistance with high yield. Similarly, *Triticum aestivum* have been engineered against bacterial disease and powdery mildew. In case of mushrooms genes are targeted to reduce the melanin content via CRISPR/Cas9 (Tieman et al. 2017; Wang et al. 2014; Shi et al. 2017). The plant system can utilize editing through CRISPR-Cas9, which could have major advantage in terms of production of non-transgenic plants and furthermore more genetic engineering can be done in such a way to overcome traditional classical breeding programs. In conclusion, CRISPR-Cas9 technology is the future demand in crops biology to attain vigorous traits.

15.2 Previous and Latest Tools for Genome Editing in Plant System

15.2.1 Zinc Finger Nucleases (ZFNs)

ZFNs are the first generation tool, which were used previously for genome editing in plant genome till 2005 (Lloyd et al. 2005). ZFNs are designed nucleases with targeted gene-specific zinc finger DNA-binding domain for the cleavage of DNA at non-specific ends. ZFN consists of type II restriction endonuclease, F*ok*I for causing nicks in the DNA (Kim et al. 1996), and a DNA-binding domain that

Table 15.2 List of plant components and their potential health benefits that can be used to targeted via CRISPR/Cas9

Component	Source	Health benefit
Carotenoids		
α-Carotene	Carrot	Inhibitor of free radicals
β-Carotene	Fruits, vegetables	
Lutein	Vegetables	Eye sight
Lycopene	Tomato	Low down risk of prostate cancer
Zeaxanthin	Citrus, maize	Eye sight
Dietary fiber		
Insoluble fiber	Wheat bran	Low down the risk of breast cancer
β-Glucan	Oat	Cardiovascular disease (CVD)
Soluble fiber	Psyllium	
Whole grains	Cereal grains	
Collagen hydrolysate	Gelatin	Osteoarthritis
Fatty acids		
ω-3 fatty acids (DHA/EPA)	Marine oils	Cardiovascular disease CVD Mental and visual functions
Flavonoids		
Anthocyanidins	Berries	Inhibitor of free radicals
Hydroxycinnamates	Wheat	Degenerative diseases
Flavanols: Catechins, tannins	Tea	Inhibitor of free radicals
Flavanones	Citrus	
Flavones: Quercetin	Fruits/vegetables	
Glucosinolates, indoles, isothiocyanates		
Sulforaphane	Vegetables broccoli	
Phenolics		
Stilbenes (resveratrol)	Grape	Heart related disease, and cancer
Caffeic acid, ferulic acid	Fruits, vegetables	
Epicatechin	Cacao	Antioxidant-like activities
Plant stanols/sterols		
Stanol/sterol esters	Maize, soy, wheat	Blood cholesterol
Prebiotic/probiotics		
Fructans, inulins, fructo-oligosaccharides	Onion and garlic	Stomach related problems
Lactobacillus	Dairy product	
Saponins	Soybean	LDL cholesterol
Soybean protein	Soybean	Heart related problems

(continued)

Table 15.2 (continued)

Component	Source	Health benefit
Phytoestrogens		
Isoflavones (daidzein, genistein)	Soybean	Osteoporosis and CVD
Lignans	Flax, rye, vegetables	LDL cholesterol, total cholesterol, and triglycerides
Sulfides/thiols		
Diallyl sulfide	Onion, garlic, olive, leek, scallion	LDL cholesterol, immune system
Allyl methyl trisulfide, dithiolthiones	Cruciferous vegetables	
Tannins		
Proanthocyanidins	Cranberry	Urinary tract health, may reduce risk of CVD and high blood pressure

consists of linked zinc finger (ZF) motifs, which recognizes a three-base pair DNA sequence. *Fok*I nuclease requires dimerization event with DNA for the double-stranded cleavage. The technique needs a combination of ZFNs and *Fok*I to create a double-strand break at the desired genomic position (Bitinaite et al. 1998). For the dimerization event of *Fok*I, ZFNs should bind on the forward and reverse strands of the DNA strands, respectively, and the two target sequences (forward and reverse) should be five to seven base pair away from each other. Soon after its discovery the ZFNs approach is validated in various plant systems, such as tobacco, corn, soybean, and *Arabidopsis* (Shukla et al. 2009; Urnov et al. 2010). The main disadvantage of this technology is difficulty in its construction and assembly, and also requires professional with strong technical knowledge and specialized resources (advance lab facility) (Ramirez et al. 2008). This shortcoming has greatly affected its popularity for adoption by the scientific community.

15.2.2 Transcription Activator-like Effector Nucleases (TALENs)

Although TALENs were discovered earlier in 2011, but it is recognized as the second-generation genome editing tools (Cermak et al. 2011). Like ZFNs, TALENs are nucleases that are generated by adding TAL effectors (TALE) DNA-binding domain to non-specific DNA F*ok*I endonuclease (Miller et al. 2011). TAL effector was mainly isolated from *Xanthomonas* bacteria, which is secreted during infection stage in their host. Tandem repeats in the host DNA are the binding sites for TAL (proteins) through a domain that consist of tandem repeats of amino acid with an approx. length of 34–35 (Boch and Bonas 2010). TALE protein generally contains N-terminal and C terminal responsible for translocation signal, and acidic transcription-activation domain, respectively, together with a DNA-binding domain (Bogdanove et al. 2010). While the DNA-binding domain consists of stretch of

30–35 amino acid with several tandem repeats of conserved amino acids, known as RVDs (two adjacent residues at positions 12 and 13) (Bogdanove et al. 2010). DNA-binding domain binds to single nucleotide at the RVDs (RVDs His-Asp (HD), Asn-Ile (NI), Asn-Gly (NG), and Asn-Asn (NN)) repeats and shows a strong affinity for the base pair like Cytosine, Adenine, Thymine, and Guanine (Boch et al. 2009). The complex of protein and DNA helps in the modification of specific DNA. On that basis, TALENs can be designed according to the target genomic sequence. But one of the major disadvantages is that TALE and DNA-binding are more sensitive to methylation which affects its binding to the chosen target sequence (Deng et al. 2012). Except sensitivity to methylation, the assembly for TALENs is easier to apply than ZFNs, resulting in a more commonly used for the genome editing in plants and other eukaryotic systems (Cermak et al. 2011; Li et al. 2012; Wendt et al. 2013; Zhang et al. 2013; Haun et al. 2014; Jung and Altpeter 2016).

15.2.3 CRISPR/Cas9

The most advanced third-generation approach of genome editing is CRISPR/Cas9 which was for first time used in 2013 (Nekrasov et al. 2013; Shan et al. 2013). CRISPR/Cas9 is a prokaryotic acquired adapted immunity against a specific phage (Horvath and Barrangou 2010; Marraffini and Sontheimer 2010). The presently used modified type II endonuclease is based on CRISPR/Cas9 for genome editing, consists of the following two main constituents: the CRISPR-associated protein 9 (Cas9) proteins, and non-coding crRNAs: a combination of trans-activating crRNA (tracrRNA) and a precursor crRNA (pre-crRNA) (Bhaya et al. 2011). The Cas9 protein is an endonuclease, which consisted of a HNH and a RuvC-like nuclease domain, for their cleavage activity (Bhaya et al. 2011; Horvath and Barrangou 2010). The tracrRNA contains the sequence, which is complementarily adjacent to a nucleotide within the pre-crRNA that helps in the formation of an RNA duplex required for crRNA maturation and crRNA-mediated DNA cleavage (Bhaya et al. 2011; Horvath and Barrangou 2010). The pre-crRNAs are synthesized from the CRISPR sequences that contain the repeat/spacer sequences. These repeats are of 23 to 47 base pairs and typically identical in length. The hairpin-like structures are formed by the palindrome sequences acting as secondary structures. While the spacers consisted of 21–72 bp sequences and are originated from the foreign DNA that helps to guide Cas9 protein to cleave a foreign protospacer, upon invasion.

The pre-crRNA is comprised of the CRISPR repeat interspaced by spacer sequence from invading foreign DNA and post-transcription tracrRNA base pairs with the pre-crRNA to form RNA complex. Further this complex in the presence of Cas9 endonuclease protein is cleaved by RNase III to synthesize the mature crRNAs. The mature duplex consists of 20 nucleotides sequence and the crRNA stretch at the 5′ end, which directs the Cas9 protein to attach and perform cleavage on the target DNA sequence with protospacer adjacent motifs (PAM). Finally, the Cas9HNH nuclease domain recognizes and creates nicks in the DNA strand complementary to the guide RNA, whereas RuvC-like nuclease domain of Cas9 creates a DSB at

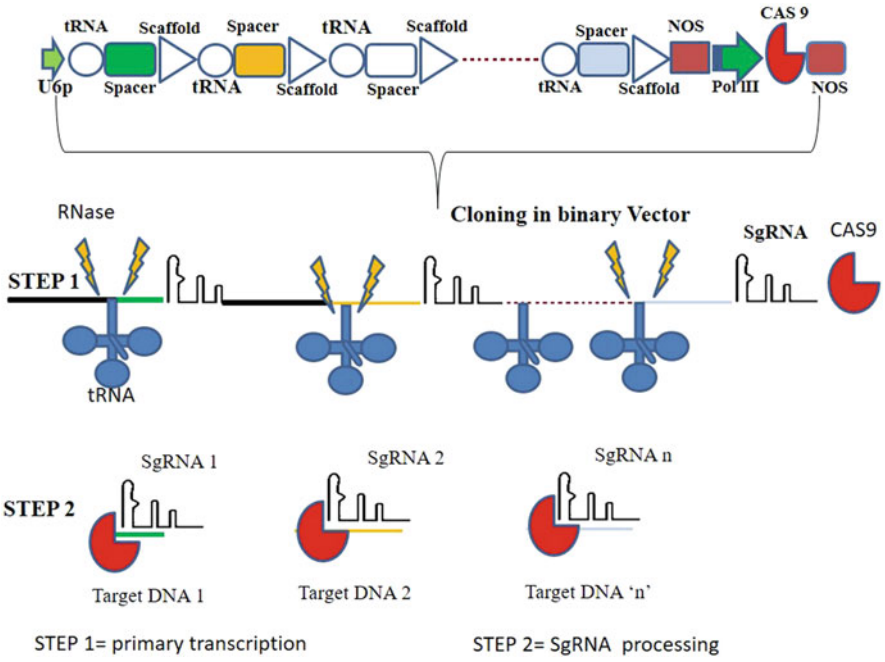


Fig. 15.1 Strategy of multigene targeting in the plant genome. Appropriate cassette of multiguide RNA (multiplex) with different sgRNAs is shown. The spacers sequence and sgRNA scaffold can be incorporated between the tRNAs and terminator. The cassette is driven under the Pol-III promoter in the binary vector

3 nucleotides upstream within the protospacer. The crRNA and tracrRNA complex has been engineered into a single guide RNA (sgRNA), resulting in two-component for the functioning of CRISPR/Cas9, that determines the specificity of cleavage (Cong et al. 2013) (Fig. 15.1). Therefore, CRISPR/Cas9 is a much advanced and easily available tool to engineer, according to the desired need and most importantly it has a much higher ability to generate targeted orientated mutations. In the case of CRISPR/Cas system multiple sgRNAs can be used for the selected target DNA sequences for multiplexing. CRISPR/Cas9 has been used in a various number of plants and crops including rice, wheat, tobacco, *Arabidopsis*, etc. (Jiang et al. 2013).

15.3 Advancement in CRISPR/Cas system

15.3.1 CRISPR/Cpf1

CRISPR/Cpf1 is a more advanced genome editing tool. Currently, there are number of Cpf1 systems available: *Fn*Cpf1 from *Francisella novicida*, *As*Cpf1 from *Acidaminococcus sp.,* and *Lb*Cpf1 from *Lachnospiraceae bacterium*. All these Cpf1s have been used for crop genome editing in various plant species, including model

plant and crop plants recently (Endo et al. 2016; Wang et al. 2017a, b). In the year 2015, CRISPR/Cpf1 the system was discovered as a new genome editing tool and in the year 2016, it was for the first time used for plant genome editing (Zetsche et al. 2015; Endo et al. 2016). The CRISPR/Cpf1 consists of two main components, the Cpf1 enzyme and the crRNA which are responsible for the activity and specificity. Although Cpf1 and Cas9 systems have same mode of action (Shmakov et al. 2017), but the first major difference is CRISPR/Cpf1, system does not require a trans-acting crRNA/tracrRNA for crRNA maturation in the Cas9 system. The second difference is that mature crRNA is about 42–44 nt long and contains 19 nucleotide sequence repeat and 23–25 nucleotide spacer sequence. On the other hand, CRISPR/Cas9 contains gRNA of ∼100 nts (Zetsche et al. 2015). The third major difference is the nuclease activity for the cleavage of the double-stranded DNA. CRISPR/Cpf1 system needs only a single promoter to start a cascade of more than various small crRNAs, which is a more simplified tool than CRISPR/Cas9 which needs expression of individual sgRNAs for editing multiple targets of genes. Fourth, unlike Cas9 containing nuclease domains (RuvC and HNH), which cause nicks in DNA strands at the same position upstream of 3–4 bp of the PAM region to produce blunt ends. Cpf1 cuts upstream of PAM to produce sticky ends (Zetsche et al. 2015). Mutation caused by Cpf1 usually is larger than Cas9-induced mutations. The sticky ends produced lead to an increase in the efficiency of HDR-recombination caused by Cpf1 insertion in the cleaved position (Zetsche et al. 2015). Fifth, CRISPR/Cas9 requires a PAM sequence with G-rich repeats (5′-NGG-3′) at the 3′ end and CRISPR/Cpf1 requires a T-rich PAM (5′-TTTN-3′ or 5′-TTN-3′) at the 5′ end of the target sequence.

15.3.2 Data Mining Target Selection (sgRNA Design) and Selection of CRISPR/Cas9 System

The initial step for the genome editing is the selection of desired target genomic region in the plant system or any system for their manipulation. The whole genomic features of the targeted host include genome polymorphism, off-targets, presence of intronic region that may get further arise due to alternative splicing, and single-nucleotide polymorphisms (SNPs) are the main considerable things, which may affect the genome editing. However, nowadays many online tools are available to tackle with such restrictions for, e.g., off-target and sgRNA designing for possible less off-target as well as prediction of cleavage sites. After the successful plant genome editing and engineering, several freely available online tools such as CCTop, MIT CRISPR, ATUM, Alt-R CRISPR-Cas9 System, CHOPCHOP, CROP-IT, GT-Scan, sgRNA Designer and Cas-Finder, etc. can play a significant important role in finding gRNA and PAM region from the genomic sequences. After the successful selection of gRNA and PAM region in the genome, next step is to express bacterial Cas9 protein into host (plant) cells. For the optimized expression in plant system codon's optimization, an appropriate promoter is required for CRISPR/Cas9 functionality in crop plants (Belhaj et al. 2015). In plant system, various types of

promoters are validated like LTR, UBO U3, CaMV 35S, CMV, EF1A, and U6 and are mainly used for Cas9 expression (Belhaj et al. 2015).

Likewise another point to be considered is selection of a suitable endonuclease Cas9 system. In, eukaryotic system, *Sp*Cas9, has a powerful gene editing tool commonly in eukaryotes. Meanwhile, in many reports, a plant codon-optimized Cas9 is used to achieve high efficiency. Such type of advancement in the genome editing in future laid the foundation of high-throughput genetic engineering and open up new horizon using CRISPR/Cas9 system.

15.3.3 CRISPR/Cas9 Cassette Delivery into Crops and Screening of Targeted Mutations in Crops

Steps of cloning gene cassette containing sgRNA and Cas9 in plant expression vector are used to deliver into the plant genome. For the successful delivery of the cassette into crop plants with precision will be a challenging task due to polyploidy events in plants like wheat, potato, and mustard having higher polyploidy level.

Various approaches can be used to deliver CRISPR/Cas9 cassette transformation into crops plant system via protoplast, biolistic inoculations, and transit peptides. Nowadays various plant virus-based vector backbones are used to achieve high efficiency genome editing (Butler et al. 2016). However, for the stable genetic transformation in crops, two main approaches are *Agrobacterium*-mediated and particle bombardment mediated genetic transformation of callus, embryos, used for the successful methods for delivering of CRISPR/Cas9 in crops (Ma et al. 2015). After the successful transfection, the crops plant system (T_0) can be selected on selectable media like hygromycin and kanamycin, which are used for the confirmation of the gene editing using different approaches such as real-time PCR and Sanger sequencing.

Phenotypic variation caused by the specific gene editing is further confirmed using various approaches. A rapid way for screening the efficacy of gene editing is the use of reporter genes such as YFP, GFP, RFP, and GUS. The targeted gene can be confirmed using PCR or their transcript abundance using real-time PCR. Another classical approach to screen the successful genetic engineered plants process via CRISPR/Cas9 system can be checked by using (PAGE) polyacrylamide gel electrophoresis or using high-throughput Li-COR analyzer. Next-generation or high-throughput sequencing of the PCR amplicon is another advanced method for detection of mutation in the specific segment. But this approach is quite expensive and time taking. One of the major advantages of NGS and high-throughput PCR is that it is highly accurate to detect the rate of deletion and off-target mutations in the targeted region or as well as in whole genome (Fauser et al. 2014) (Fig. 15.2). NGS approach is required and very useful for confirming the mutations (Zhou et al. 2014).

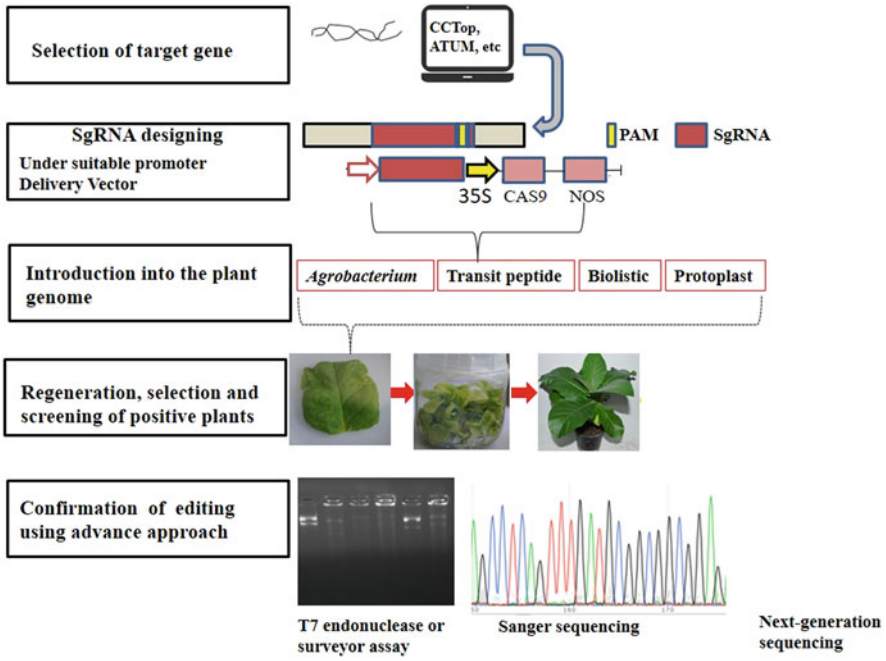


Fig. 15.2 Summary of the CRISPR/Cas9 based steps of plant genome editing. In the first step identification of best target sites is done using various online tools and primers are designed. The target-specific sgRNA with Cas9 cassettes are cloned in binary vector. In next step cassettes are then co-transformed in vivo into the plants. The transgenic plants are confirmed for the target-specific mutations using various approaches like high-throughput sequencing approaches. The positive transformed plant is then selected for further applications and analysis

15.3.4 Gene Stacking Using Multiplex CRISPR/Cas9 Model in Crop System

For the accurate modification of multiple genes responsible for a particular agronomical trait, the major concern to plant scientist is improving the agronomical traits by using forward and reverse genetics (Kumar et al. 2018). Nowadays, most advanced approach is being used for multigene modification in various crops like rice and wheat, where manipulation of tRNA-maturing cascade process in cells was utilized to express multiple sgRNAs in a single construct (Xie et al. 2015). In CRISPR/Cas9 multigene targeting system, cassettes are arranged in a polycistronic tRNA–sgRNA (PTG) fashion, which is composed of multiple sgRNAs units with scaffolds driven by a single promoter. During the process of transcription, a long precursor polycistronic tRNA–sgRNA template is cleaved by host's endogenous RNases leaving more than one free sgRNAs that further directs the endonuclease (Cas9) to recognize the targeted sites. This mechanism helps in achieving a higher level of mutation in the genetically engineered plants as compared to other system

and can be used for various crops for achieving complete and multigene disruption (Xie et al. 2015). The successful implications of PTG-based method with complex genome have been already employed in allopolyploid wheat and *Zea mays*, where increased mutation frequency was observed using PTG system (Paul and Qi 2016). Such universal tRNA–gRNA-based approach can be used to target multiple genomic positions at a time in crop plants for different biotic stress (Table 15.3).

Table 15.3 List of diseases related to plant and their causal agent

Diseases	Causal agent	Plants
Bacterial		
Aster yellows	Phytoplasma bacterium	Vegetables, cereals, garden plants
Bacterial wilt	Corynebacterium, Erwinia, Pseudomonas, and Xanthomonas	Vegetables, cereals, garden plants
Blight	Erwinia amylovora (fire blight)	Rose family, tomatoes, potatoes, and apples,
	Xanthomonas oryzae (Rice bacterial blight)	*Oryza sativa* and *O. glaberrima*
Crown gall	Agrobacterium tumefaciens	Grape, rose, nut trees
Basal rot	Pectobacterium carotovorum and Pseudomonas viridiflava	Vegetables, cereals, garden plants, and wild species
Scab	Streptomyces scabies and related species	Apples, crabapples, cereals, cucumbers, peaches, pecans
Fungal		
Anthracnose	Colletotrichum or Gloeosporium	Sycamore, ash, oak, and maple
Black knot	Apiosporina morbosa	Plums, cherries, apricots, chokecherries
Blight	Cryphonectria parasitica (chestnut blight)	Chestnut, post oak, live oak
	Phytophthora infestans (late blight)	Potato or tomato
Canker	Botryosphaeria dothidea	Tree
Clubroot	Plasmodiophora brassicae	Mustard family
Damping-off	Rhizoctonia solani, Aphanomycescochlioides, Phytophthora, Botrytis, Fusarium, Cylindrocladium, Diplodia, Phoma, and Alternaria	Plant seedlings
Dutch elm disease	Ophiostoma, Ceratocystis ulmi	Tree
Ergot	Ascomycete fungus Claviceps	Cereal grasses
Fusarium wilt	Fusarium oxysporum	Sweet potatoes, tomatoes, legumes, melons, and bananas
Leaf blister	T. deformans	Peaches, nectarines, and almonds
Mildew	Downy mildew, Sclerospora, Bremia, Peronospora, Phytophthora, Plasmopara, and Pseudoperonospora	Stems, flowers, and fruits

15.3.5 Countering off-Targets in Plant Genome Editing

The advantage of CRISPR/Cas9 is fascinating and most popular tool used for genome editing, but the major challenge is off-targets. However, in some eukaryotic system due to large number of ploidy level like wheat and potato, off-targets can be caused by selecting sgRNA of less size, which has resemblance to unrelated sequences (isoforms) in the genome. To tackle this problem manipulation of genes can be a challenging task. Various studies have showed that Cas9 (nuclease) has the potential to cleave the target regions in the genome, even if various conflicts are already preexisting between sgRNA and the target sequence, which is known as off-target. Previous reports showed 20 base pair long stretch of sgRNA which has high level of specificity (Pattanayak et al. 2013), while some reports suggest that advanced genetic manipulations can be increased by using less than 20 bp length sgRNA (Fu et al. 2014). It has been well proven that minimum of three nucleotide discrepancies, at 5′ prime end, between sgRNA and target sequence can be adaptable by CRISPR/Cas9 system (Ran et al. 2013). Off-targets can also be significantly minimized by incorporating a number of (G) guanidine residues at 5′ end or by limiting the length of the sgRNA up to 17 base pairs. But less extension of base pairing may provide the RNA–DNA hybrid more prone to discrepancies, and also affects the binding energy of the sgRNA–DNA hybrid.

Another way to minimize off-targets is by using modified Cas9 variants (VQR and VRER) or Cas9 orthologs like *Stl*Cas9, which could be also used for creating efficient mutation with altered engineered PAM sites (Honda et al. 2014; Kleinstiver et al. 2015). The single-stranded nick is not sufficient for creating mutations, due to base excision repair mechanism existing in the cells. To get rid of this a pair of nickases for cleaving the target DNA sequence up to 20–40 base pair has been demonstrated by the use of two sgRNAs (Ran et al. 2013). Over the time, the Cas9 specificity can be increased by taking approach of creating variations in the Cas9 by structure-guided method by replacing the positively charged amino acid with the capacity to affect the interlinking of Cas9 with non-target DNA strand, thereby increasing the specificity by improving the quality of Cas9 activity on the sgRNA to the selected desired sequence (Slaymaker et al. 2016). These strategies will allow us to obtain required mutation free of any non-specific (off-targets) mutations through CRISPR/Cas9in complex crop genome like wheat.

15.3.6 Future Applications of Genome Editing CRISPR/Cas9 in Crop System

From the ancient time of human civilization agricultural-based cultivation is being done for the survival of human race and economic benefit. In the modern era, the use of high-throughput technology opens the way to manipulate the genome. The crop system has undergone through lots of evolution and various biotic and abiotic pressures to evolve (Bekheet and Hanafy 2011). The whole genome map of

crops also makes it possible to other studies including transcriptomics and microarray, mRNAs expression profiling, comparative analysis of other crops reference-based genome assembly, genetic variability using single-nucleotide polymorphisms (SNPs).

Currently, more than 50% of the recently discovered genes show homologous sequence similar to formerly reported genes in other plants. However, *Insilco* analysis is not exact way to define gene function, and further experimental validation is needed in most cases. The cellular or developmental function and related indirect information can be achieved from spatial and temporal expression experiment of mRNA and/or protein in different cell types, during various developmental stages, or during biotic and abiotic infection. Transcription and post-translation modification also affect its integrity and can be informative as well. Functional genomics tools used especially for RNA biology by knocking out or over-expressing the targeted gene to resolve the gene function linked to a phenotype or other agronomic traits. The developed methods for the study of gene function in plants and crops are associated with benefits and few drawbacks.

15.3.7 Abiotic and Biotic Limitations in Crop Cultivation

Biotic and abiotic stresses play a major role in genetic resource development through natural selection (Arzani and Ashraf 2016). In the modern era of comprehension and scientific knowledge interlinked to gene and genomics helps a lot in expanding the horizon of crop development and improvement (Arzani and Ashraf 2016). Nowadays CRISPR/Cas9 has become popular among the scientific community for the genetic manipulations. These GE tool CRISPR/Cas9 systems have provided new horizon in basic plant research and crop improvement. In plants system the transfer DNA (T-DNA) and knockout are mainly delivered by *Agrobacterium* resulting in preferentially integrated and induction of DSBs and successfully achieved in various plants like *Nicotiana tabacum* (Gao et al. 2015), *Arabidopsis* (Li et al. 2013), and crops system also, such as wheat (Fauser et al. 2014), maize, rice, sorghum (Jiang et al. 2013), tomato (Ron et al. 2014), and sweet orange (Jia and Wang 2014). An established protocol and procedure of selected targeted mutations in crops like rice and wheat using the CRISPR/Cas9 system is well defined by Shan et al. (2013). CRISPR/Cas9 system can be also used for developing resistance against plant diseases (Jia et al. 2017; Peng et al. 2017). CRISPR/Cas9 in some gymnosperms and angiosperm, like citrus, populus (Song et al. 2016), and apple (Nishitani et al. 2016) is well established. This type of crops system is also known as lighthouse of life, which is a source of nutrition for human diet. In the past decades, situation such as genetic vigorous loss occurred to many productive many crops is due to different biotic and abiotic conditions. Traditional biological procedures, like PTC (plant tissue culture), MAS (marker-assisted selection and breeding), and DNA finger printing, have been used in plant at genomics and proteomic level, but remain partially successful in bringing major impact to improvement in crop development.

There is increasing demand to develop new variety, which are highly tolerant to biotic and abiotic stresses. This cannot be only achieved by classical breeding approaches but certainly requires advanced functional genomic tools for modifying the plant genome resistant to a wide range of biotic and abiotic stresses. Another approach is to engineer tolerance to various pathogens and diseases using CRISPR/Cas9. Various stresses like salinity, drought, and heat are the major abiotic stresses affecting plants system but there is less focus on biotic stress as compared to abiotic stress. In spite of much research, biotic stress is a very vast growing field and the more knowledge is still required to understand the molecular mechanism of biotic stress. However, a genomic and proteomic analysis will help to understand the cascade mechanism and pathway for the biotic stress and overcome this problem.

Plant diseases are an abnormal state of a plant that disturbs its essential functions. All species of plants are infected by disease once in their lifecycle, including wild and cultivated plants. The occurrence and impact of plant diseases vary from environmental conditions, presence of the pathogen, and flora and fauna that may affect the plant health. Some plant genotypes are prone to outbreaks of diseases, while some are resistant to them. However, the development of new elite, high yielding, and resistant cultivar demands a high accurate and trustable methodology. The CRISPR/Cas9-based methodology can provide the target genomes of plants and pathways of pathogens directly or indirectly involved in developing the infection in the plants. In crops, biotic and abiotic mediated pathways genes can be selected by CRISPR/Cas9-based approaches for developing resistance (Fig. 15.3).

15.3.8 Disease Resistance and Herbicide Tolerance

Plant diseases are the major concern and cause crop yield loss, which ultimately affect the nutritional value and food processing quality (Savary et al. 2012). Genome modification and editing approaches have been applied to develop disease immunity by editing the pathway genes involved. One of the good examples is from rice, where a host disease-susceptibility gene (*OsSWEET14*) is activated by pathogen *Xanthomonas oryzae* pv. Genome editing approach was used for the disruption of the sequences in the promoter region for increased affinity towards the resistance to bacterial blight (Li et al. 2012). Similarly, *CsLOB1* in citrus plays an important role in host disease-susceptibility and promotes pathogen interaction to plants, has been targeted using CRISPR/Cas9-targeted modification to develop canker-resistance cultivars (Peng et al. 2017). Similarly, *TaMLO* knockout in wheat leads to enhanced resistance against a large variety of pathogens and durable resistance to obligate fungus, *Blumeria graminis* f. sp. *tritici* (Bgt). Nowadays plant viruses are also an important concern beside the bacterial and fungal pathogens. For example, RNA viruses (Turnip mosaic potyvirus) of plant require an eIF4 eukaryotic translation initiation factor, which acts as a template for their replication. Manipulation in eIF4E has produced broad virus resistance in many crops such as in cucumber (Chandrasekaran et al. 2016), whereas in tomato, resistance has been acquired

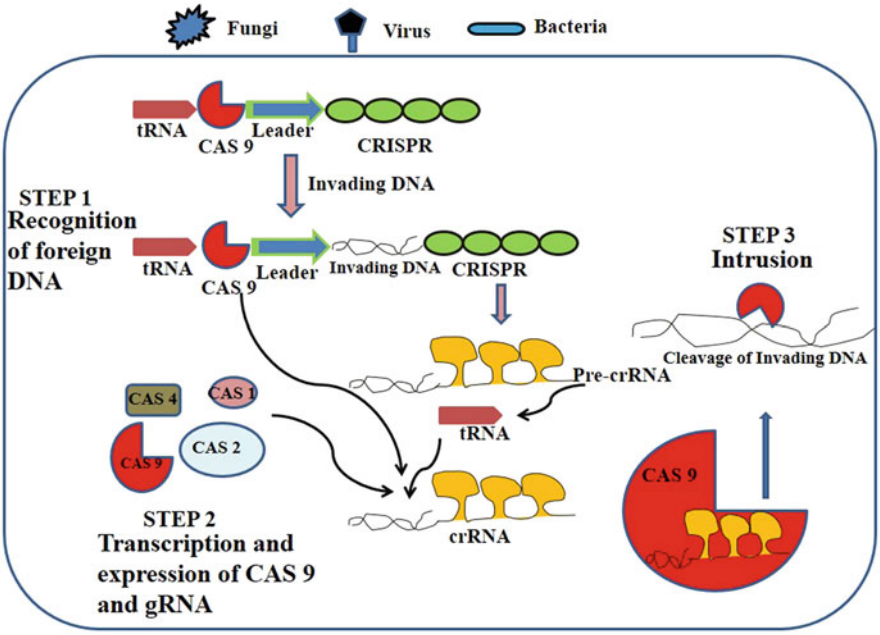


Fig. 15.3 Mechanism of resistance (cartoon representation) based on CRISPR/Cas9-based for plant biotic stress. The identification and removal of pathogen DNA/RNA occur in the three steps: reorganization, transcription and expression, and intrusion. In the first step reorganization of the invading DNA/RNA is introduced into the CRISPR array at the leading sequence. In next step the transcription and expression of pre-CRISPR/RNA (Pre-crRNA) occur. The final steps includes the attachment to the complementary sequence of the foreign DNA/RNA, which is identified by Cas9 and nicks are introduced

directly by targeting the viral genome directly using advance genomic tool like CRISPR/Cas9.

Herbicide tolerant crop production by using genome editing opens a new path to create vigorous and high yielding crops (Lombardo et al. 2007). This opens the new gate to edit endogenous plant genes such as *EPSPS* and *ALS* for the herbicide resistance. *ALS* cause the biosynthesis of various amino acids like valine, leucine, and isoleucine (Chipman and Shaanan 2001) and its activity is inhibited by herbicides (Svitashev et al. 2015). Similarly another gene *EPSPS* (5-enolpyruvylshikimate-3-phosphate synthase) is a precursor molecule for the synthesis of aromatic amino acids in plants (Kishore and Shah 1988). The main target of EPSPS is glyphosate, a widely used herbicide (Kishore and Shah 1988). The common approach to introduce glyphosate tolerance in crop plants is by disrupting its structure that may affect the binding sites resulting in increased tolerance to plants (Sammons and Gaines 2014). Various herbicide-resistance genes can be targeted via CRISPR/Cas9 in many crops like maize, soybean, and rice.

15.3.9 Nutrition and Processing Quality in Potatoes

Potatoes are good source of starch and consumed worldwide. The tuber (potato) is stored in cold storage to stop sprouting, which affects the nutrition and processing quality. Sprouting results in the accumulation of reducing sugars in the tuber. At high temperature, these sugars react with free radicals (amino acids) to produce acrylamide, which is highly toxic and undesirable for human health. Another gene from potato encodes an invertase (*VINV)* that converts sucrose into glucose and fructose, which is mainly responsible for the production of reduced sugars. The invertase pathways can also be targeted for the development of cis-genic lines to remove problem associated with acrylamide production.

15.3.10 Functional Studies of Biotic and Abiotic Stress-Related Genes

Wheat is another important staple crop, in which functional validation of *TaMLO* gene using CRISPR/Cas9 revealed resistance against the mildew resistance locus (Shan et al. 2014). Kim et al. (2018) reported that abiotic stress-related genes were targeted in wheat (*TaDREB2*) and their cis-responsive element showed expression in drought resistance. In another report Wang et al. (2018) showed that in wheat three genes (*TaGW2, TaLpx-1* lipoxygenase, and *TaMLO*) were negatively correlated with grain traits, as well as resistance against *Fusarium graminearum,* respectively, when targeted with CRISPR.

In model plant *Arabidopsis,* Feng et al. (2013) showed for the first time genome editing using CRISPR/Cas9. In A. *thaliana* protoplast *PDS3 (PHYTOENE DESATURASE), AtFLS2 (FLAGELLIN SENSITIVE 2), CYCD3 (CYCLIN D-TYPE 3), RACK1 (RECEPTOR FOR ACTIVATED C KINASE 1-AtRACK1b* and *AtRACK1c)* genes were edited (Li et al. 2013). Out of these gene showed resistance towards turnip mosaic virus (TuMV). However resistance is also shown in cotton against *Verticillium dahliae* after targeting the *Gh14–3-3d* gene via CRISPR/Cas9 (Zhang et al. 2018b).

Soybean (*Glycine max*) is one of the most important oil producing seed crops grown in many parts of the world. The fertilized ovary (seed) also consists of variety of biologically active compounds that are helpful to human health. Cai et al. (2015) successfully obtained the firstly genome edited plants of soybean using CRISPR/Cas9 by a single sgRNA; meanwhile, *GmFEI2*and *GmSHR* were targeted with six guide RNA showed phenotype in development of hairy root system. CRISPR/Cas9 was also used to target the pathogen virulence gene *Avr*4/6 in soil borne pathogen *Phytophthora sojae* (Fang and Tyler 2016) in soybean. Tomato (*Solanum lycopersicum* L.) is a commercially important crop and consumed worldwide that is an ideal model for testing CRISPR/Cas9 gene editing methods. In one study gene encoding mitogen-activated protein kinases (MAPKs) and

(*SIMAPK*3) mutants was targeted to generate tolerance lines against drought stress and wilting (Wang et al. 2017b).

Citrus is an economically and commonly available important fruit crop. CRISPR approach is also used to improve the various traits like citrus canker resistance by targeting the *LATERAL ORGAN BOUNDARIES* (*CsLOB1*) gene. *CsLOB1* showed the susceptibility to citrus canker. Different alleles are responsible for *CsLOB1,* which also contain the special effector for binding, one of the classical examples is EBEPthA4 to support the resistance towards the citrus canker which can be targeted via CRISPR to high degree of resistance (Peng et al. 2017). In the past decades breeders are trying to obtain grape fruit with good quality agronomical traits and resistance against several abiotic stress via classical approaches. Nowadays modern genomic approaches are also used to obtain the desired traits via CRISPR/Cas9. In one of such approaches ribonucleoproteins (RNPs) delivery in the grape protoplasts showed more effective results against the powdery mildew. In other case targeted mutagenesis were done in *VvWRKY*52 for biotic stress responses and that resulted in increased disease resistance to fungal infection *Botrytis cinerea* (Wang et al. 2018).

15.3.11 Future Application of CRISPR/Cas9 in Improving the Agronomical Trait

Targeted genome modification is pivotal for precise modification of the targeted genomic regions to avoid the laborious task of backcrossing, linkage drag effects, and footprints of transgenes. The majority of both biotic and abiotic stress-related traits are quantitative in nature and are controlled by multiple genes. CRISPR/cas9 system can be of particular interest in modifying the multiple genes through insertions or homologous recombination. The versatility and economics of CRISPR/Cas9 system gave a new hope for the plant breeders to save time and labor for improving the elite cultivars. The simplicity and low cost of the latter have drawn the attention of molecular breeders. The meteoric rise in research articles on the usage of CRISPR/Cas9 system in plants during the year 2014–2019 has further given the boost of adapting this technology for different crop species. For instance, three homologous alleles of powdery mildew resistance locus (*MLO*) were successfully knocked out using TALEN and CRISPR/Cas9 system in hexaploid wheat and demonstrated the NHEJ mediated mutations in wheat. Similarly, transient expression and stable transgenic lines were recovered using CRISPR/Cas9 system in *Arabidopsis*, tobacco, rice, maize, tomato, sorghum, and soybean. Researchers have attempted for homologous recombination mediated genome modifications using CRISPR/Cas9 in *Arabidopsis*, rice, and tobacco. However, in *Arabidopsis* the replacement efficiency was very low. Multiplexing of gene modification has also shown to be a promising approach in plants to accelerate the use of CRISPR/Cas9 system (Xie et al. 2015). Recently, simple and powerful platform to enhance the targeting of multigene editing capabilities using CRISPR/Cas9 system was developed using endogenous tRNA-processing system. Here, they demonstrated the use of synthetic gene containing randomly arrayed tRNA–gRNA, later processed

into individual gRNAs targeting multiple genes in rice (Minkenberg et al. 2017). CRISPR/Cas9 has been applied in various herbaceous plant species also. However, it was not clear whether this technology can be used in woody plants which have long life span and complex genome.

As biotic and abiotic stresses are controlled by multiple genes, the manipulations of several genes to improve resistance/tolerance are challenging. However, site-specific designer nuclease allows stacking of multiple genes in the close vicinity to avoid the recombination between them and helps in co-segregation without leaving any footprints behind. Also the unpredicted events created through other methods of gene transfer like *Agrobacterium*-mediated gene transfer and gene gun methods can be avoided through site-specific nucleases. In particular, targeted phenotype can be developed against biotic and abiotic stresses with the help of site directed nucleases (Podevin et al. 2013). The CRISPR/Cas9 method of genome engineering can provide potential platform for improving the agronomic traits such as yield, nutritional quality, and stress tolerance in crop plants (Jaganathan et al. 2018). The problems associated with off-targets can be easily overcome in crop plants by developing segregating population and precise selection criteria. Hence, CRISPR/Cas9 system is a tremendous possibility for developing elite cultivars to meet the global food demands, and to overcome many of the environmental stresses without contaminating the original genetic constitution of the plant genome.

CRISPR-Cas9 genome editing system was used in wheat, maize, soybean, flax, and rice for trait development and crop improvement. CRISPR-Cas9 was employed to generate wheat plants that carry mutations in all the three homoeoalleles of *TaMLO* to validate its involvement in conferring heritable resistance against powdery mildew. DuPont Pioneer deployed CRISPR-Cas9 system for selected mutagenesis, accurate gene editing at site-specific gene insertion in crops like maize. In maize five genes viz., liguleless-1 gene (*LIG*), male fertility genes (*MS26 and MS45*), and acetolactate synthase (*ALS1 and ALS2*) were targeted, to improve traits like liguleless, male fertility, and herbicide tolerance (Luo et al. 2016). The progenies were reported to segregate according to Mendelian inheritance for mutations, and site-specific targeted gene insertions. *ALS*1 gene was edited in soybean using CRISPR-Cas9 system to make the crop plant resistant to chlorsulfuron (Luo et al. 2016). The Cas9-gRNA system generated herbicide glyphosate tolerant flax plants by editing the *5′-enolpyruvylshikimate-3-phosphate synthase* (*EPSPS*) genes. The edited plants showed higher tolerance towards glyphosate under greenhouse conditions. CRISPR-Cas9 mediated targeted knockout of the transcription factor *OsERF922,* improved resistance of rice blast. The number of blast lesions formed after pathogen invasion was significantly reduced in mutant T_2 lines compared to counterpart wild type. In tomato RIN gene responsible for MADS-box transcription, which control fruit ripening, was targeted using CRIPSR-Cas9. The mutants produced incomplete ripening of fruits, suggesting the involvement of RIN gene in fruit ripening in tomato (Ito et al. 2015). Herbicide tolerant rice crop plants were successfully produced through CRISPR/Cas9-related recombination of acetolactate synthase gene (Sun et al. 2016), showing the feasibility and effectiveness of CRISPR-Cas9 for gene replacement in plants.

15.4 The Limitations of Genome Editing in Crop Plants

CRISPR/Cas9 is a tool of great interest for genome, genetic editing, and engineering. But genetic manipulation in the genome of crop plants can have certain bottlenecks. One of the major challenges in this path is complex genome; it may be due to heterozygosity and polyploidy, which affects the efficiency of genome editing. The crossing event leads to high genetic variation, polymorphism, and complex genome in the crop plant genome like wheat and many tetraploid and hexaploid crops. The current scenario of the increase in population on earth could require lots of new improved traits, which is not possible through previous breeding approaches. While fast plant breeding approaches through genetic engineering mainly in crops and cereal plants with long life cycles is possible. CRISPR/Cas9 system can be executed to knockout genes that drive regulation in cereal crops for good agronomical traits and arrange their molecular traits at desired position with low segregation. Insertion of CRISPR/Cas9 system into the crop genome through genetic transformation of somatic tissues may lead to chance of arbitrary blending to bacterial plasmid genetic material and may lead to current ethical concern of GMO in the action. However, the CRISPR/Cas9 approaches using ribonucleoproteins (RNPs) may open new opportunities to address current challenge of GMO-regulations in cereal crops (Kanchiswamy 2016).

15.5 Concluding Remarks and Future Prospects

A number of genome editing platforms are available; however, the CRISPR/Cas9 system has emerged as one of the best tools due to its simplicity, preciseness, and specificity. The system can serve as a new plant breeding technique as it offers same advantages as conventional breeding methods but with extra benefits of predicted outcomes in short duration and without linkage drag effects. The applications of CRISPR/Cas9 system ranges from site directed insertions, deletions, replacements (yet to be proved in plants) with homologous recombination, tandem gene stacking, multiplex genome editing and beyond genome editing in transcriptional reprogramming, as programmable DNA-binding proteins, epigenome editing and RNA cleavage. All together with astonishing applications, one could foresee the CRISPR/Cas technology to become a regular tool in fast track crop improvement in the era of modern agriculture to meet food scarcity. The CRISPR system offers multiplex genome editing advantage, which facilitates high-throughput forward and reverse genetics screening. This further boosts the development of knockout libraries and helps in enhancing the basic research. This technology also offers a promising platform for improving the agronomic traits like yield, nutritional quality, and improved stress tolerance in crop plants. In addition, pyramiding/stacking multiple genes involved in stress responses with HDR mediated could be breakthrough. Having said huge promising potentialities, this technology however is challenged with some technical and regulatory uncertainties. However, due to

rapid advancements in NGS technologies the availability of genome information is increasing its pace in recent past for many crop plants. In addition to this, more and more exploitation of data from system biology studies such as transcriptomics, proteomics, and metabolomics and genome wide association studies are speeding up the identification of functional candidate genes and their allelic variants in plants. Such advancements have opened new opportunities for crop improvement using genome editing tools. So far, a number of transformation methods were used to deliver CRISPR-Cas9 system in crop plants for efficient genome editing, due to its genotype dependency still one has to standardize suitable delivery system for host of interest. Researchers may also look for new viral vectors, which can facilitate efficient DNA delivery in plants.

The extent of off-target mutations of CRISPR technology could be an issue as well. However, off-target effects can be overcome in different ways, if at all there are some, they would be less problematic compared to human or animals as they can be separated out in segregating generations. Taking all these aspects of genome editing technologies into consideration, more active research is needed in the field of crop genome editing. Genome modifications created in tailor-made situation using CRISPR are identical to the modifications created by conventional plant breeding strategies and plants with transgene could be separated from plants having targeted modification in segregating generations. Therefore, crop plants developed by using genome editing technology may potentially be classified as non-transgenic. However, it is still unresolved whether genome edited plants should be treated as transgenic or non-transgenic. Thus there is a call for urgent discussions worldwide among the scientific community on creating a regulatory framework and legislations. In the immediate future, not the genome editing but the availability of suitable genes with well characterized stress resistance functions is a serious limitation in plant improvement.

Acknowledgements Authors wish to thank Institutes for the facilities.

Competing Interests There is no competing interest.

References

Anderson KR, Haeussler M, Watanabe C, Janakiraman V, Lund J, Modrusan Z, Thamminana SR (2018) CRISPR off-target analysis in genetically engineered rats and mice. Nature Meth 15(7):512

Arzani A, Ashraf M (2016) Smart engineering of genetic resources for enhanced salinity tolerance in crop plants. Crit Rev Plant Sci 35(3):146–189

Bekheet SA, Hanafy MS (2011) Towards sex determination of date palm. In: Date palm biotechnology. Springer, Dordrecht, pp 551–566

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. Curr Opin Biotechnol 32:76–84

Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45:273–297

Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci 95(18):10570–10575

Boch J, Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 48:419–436

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326(5959):1509–1512

Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol 13(4):394–401

Brooks C, Nekrasov V, Lippman ZB, Van Eck J (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol 166(3):1292–1297

Butler NM, Baltes NJ, Voytas DF, Douches DS (2016) Geminivirus-mediated genome editing in potato (Solanum tuberosum L.) using sequence-specific nucleases. Front Plant Sci 7:1045

Cai Y, Chen L, Liu X, Sun S, Wu C, Jiang B et al (2015) CRISPR/Cas9-mediated genome editing in soybean hairy roots. PLoS One 10(8):e0136064

Čermák T, Baltes NJ, Čegan R, Zhang Y, Voytas DF (2015) High-frequency, precise modification of the tomato genome. Gen Biol 16(1):232

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39(12):e82–e82

Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C, Pearlsman M, Gal-On A (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. Mol Plant Pathol 17(7):1140–1153

Char SN, Neelakandan AK, Nahampun H, Frame B, Main M, Spalding MH, Yang B (2017) An agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. Plant Biotechnol J 15(2):257–268

Chen Y, Wang Z, Ni H, Xu Y, Chen Q, Jiang L (2017) CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. Sci China Life Sci 60(5):520–523

Chipman DM, Shaanan B (2001) The ACT domain family. Curr Opin Struct Biol 11(6):694–700

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823

Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK et al (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335(6069):720–723

Endo A, Masafumi M, Kaya H, Toki S (2016) Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. Sci Rep 6:38169

Fan D, Liu T, Li C, Jiao B, Li S, Hou Y, Luo K (2015) Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. Scientific Rep 5:12217

Fang Y, Tyler BM (2016) Efficient disruption and replacement of an effector gene in the oomycete Phytophthora sojae using CRISPR/C as9. Mol Plant Pathol 17(1):127–139

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/C as-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J 79(2):348–359

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res 23(10):1229

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnol 32(3):279

Gao J, Wang G, Ma S, Xie X, Wu X, Zhang X et al (2015) CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. Plant Mol Biol 87(1–2):99–110

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E et al (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol J 12(7):934–940

Honda A, Hirose M, Sankai T, Yasmin L, Yuzawa K, Honsho K et al (2014) Single-step generation of rabbits carrying a targeted allele of the tyrosinase gene using CRISPR/Cas9. Exp Anim 64(1):31–37

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327(5962):167–170

Ito Y, Nishizawa-Yokoi A, Endo M, Mikami M, Toki S (2015) CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. Biochem Biophys Res Commun 467(1):76–82

Jaganathan D, Ramasamy K, Sellamuthu G, Jayabalan S, Venkataraman G (2018) CRISPR for crop improvement: an update review. Front Plant Sci 9:985

Jia H, Wang N (2014) Targeted genome editing of sweet orange using Cas9/sgRNA. PLoS One 9(4):e93806

Jia H, Zhang Y, Orbović V, Xu J, White FF, Jones JB, Wang N (2017) Genome editing of the disease susceptibility gene Cs LOB 1 in citrus confers resistance to citrus canker. Plant Biotechnol J 15(7):817–823

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41(20):e188–e188

Jiang WZ, Henry IM, Lynagh PG, Comai L, Cahoon EB, Weeks DP (2017) Significant enhancement of fatty acid composition in seeds of the allohexaploid, Camelina sativa, using CRISPR/Cas9 gene editing. Plant Biotechnol J 15(5):648–657

Jung JH, Altpeter F (2016) TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. Plant Mol Biol 92(1–2):131–142

Kanchiswamy CN (2016) DNA-free genome editing methods for targeted crop improvement. Plant cell Rep 35(7):1469–1474

Kim D, Alptekin B, Budak H (2018) CRISPR/Cas9 genome editing in wheat. Funct Integr Genomics 18(1):31–41

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci 93(3):1156–1160

Kishore GM, Shah DM (1988) Amino acid biosynthesis inhibitors as herbicides. Annu Rev Biochem 57(1):627–663

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z et al (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523(7561):481

Kumar P, Mishra A, Sharma H, Sharma D, Rahim MS, Sharma M et al (2018) Pivotal role of bZIPs in amylose biosynthesis by genome survey and transcriptome analysis in wheat (Triticum aestivum L.) mutants. Sci Rep 8(1):17240

Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N et al (2015) Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biol 16(1):258

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J et al (2013) Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31(8):688

Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30(5):390

Li M, Li X, Zhou Z, Wu P, Fang M et al (2016). Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. Front Plant Sci 7:377

Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang YJ et al (2015) Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. Metabol Eng 31:13–21

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci 102(6):2232–2237

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Holmes MC (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25(11):1298

Luo ML, Leenay RT, Beisel CL (2016) Current and future prospects for CRISPR-based tools in bacteria. Biotechnology Bioeng 113(5):930–943

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R et al (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8(8):1274–1284

Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11(3):181

Miflin B (2000) Crop improvement in the 21st century. J Exp Bot 51(342):1–8

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29(2):143

Minkenberg B, Xie K, Yang Y (2017) Discovery of rice essential genes by characterizing a CRISPR-edited mutation of closely related rice MAP kinase genes. Plant J 89(3):636–648

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31(8):691

Nishitani C, Hirai N, Komori S, Wada M, Okada K, Osakabe K et al (2016) Efficient genome editing in apple using a CRISPR/Cas9 system. Sci Rep 6:31481

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31(9):839

Paul JW, Qi Y (2016) CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. Plant Cell Rep 35(7):1417–1427

Peng A, Chen S, Lei T, Xu L, He Y, Wu L et al (2017) Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene Cs LOB 1 promoter in citrus. Plant Biotechnol J 15(12):1509–1519

Podevin N, Davies HV, Hartung F, Nogue F, Casacuberta JM (2013) Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. Trends Biotechnol 31(6):375–383

Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, Cathomen T (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5(5):374

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE et al (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154(6):1380–1389

Ron M, Kajala K, Pauluzzi G, Wang D, Reynoso MA, Zumstein K et al (2014) Hairy root transformation using agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiol 166(2):455–469

Sammons RD, Gaines TA (2014) Glyphosate resistance: state of knowledge. Pest Manag Sci 70(9):1367–1377

Savary S, Ficke A, Aubertot JN, Hollier C (2012) Crop losses due to diseases and their implications for global food production losses and food security. Food Secur 4(4):519–537

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9(10):2395

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8):686

Shi TQ, Liu GN, Ji RY, Shi K, Song P, Ren LJ et al (2017) CRISPR/Cas9-based genome editing of the filamentous fungi: the state of the art. Appl Microbiol Biotechnol 101(20):7435–7443

Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W et al (2017) Diversity and evolution of class 2 CRISPR–Cas systems. Nat Rev Microbiol 15(3):169

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE et al (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459(7245):437

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. Science 351(6268):84–88

Song G, Jia M, Chen K, Kong X, Khattak B, Xie C et al (2016) CRISPR/Cas9: a powerful tool for crop genome editing. Crop J 4(2):75–82

Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H et al (2016) Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. Mol Plant 9(4):628–631

Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM (2015) Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. Plant Physiol 169(2):931–945

Tieman D, Zhu G, Resende MF, Lin T, Nguyen C, Bies D et al (2017) A chemical genetic roadmap to improved tomato flavor. Science 355(6323):391–394

Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V et al (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 33(2):187

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11(9):636

Wang L, Chen L, Li R, Zhao R, Yang M, Sheng J, Shen L (2017b) Reduced drought tolerance by CRISPR/Cas9-mediated SlMAPK3 mutagenesis in tomato plants. J Agric Food Chem 65(39):8674–8682

Wang M, Mao Y, Lu Y, Tao X, Zhu JK (2017a) Multiplex gene editing in rice using the CRISPR-Cpf1 system. Mol Plant 10(7):1011–1013

Wang W, Simmonds J, Pan Q, Davidson D, He F, Battal A et al (2018) Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat. Theor Appl Genet 131(11):2463–2475

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol 32(9):947

Wang Z, Pan Q, Gendron P, Zhu W, Guo F, Cen S et al (2016) CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. Cell Rep 15(3):481–489

Wendt T, Holm PB, Starker CG, Christian M, Voytas DF, Brinch-Pedersen H, Holme IB (2013) TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. Plant Mol Biol 83(3):279–285

Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci 112(11):3570–3575

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P et al (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759–771

Zhang Z, Ge X, Luo X, Wang P, Fan Q, Hu G et al (2018b) Simultaneous editing of two copies of Gh14-3-3d confers enhanced transgene-clean plant defense against Verticillium dahliae in allotetraploid upland cotton. Front Plant Sci 9:842

Zhang Z, Xiang D, Heriyanto F, Gao Y, Qian Z, Wu WS (2013) Dissecting the roles of miR-302/367 cluster in cellular reprogramming using TALE-based repressor and TALEN. Stem Cell Rep 1(3):218–225

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W (2014) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature 509(7501):487

Synthetic Biology at the Hand of Cell-Free Systems

16

Amir Pandi, Olivier Borkowski, and Jean-Loup Faulon

Abstract

Cell-free systems are emerging as membraneless transcriptional and translational tools to study biology and for prototyping, characterization, and engineering novel biological systems. Cell-free systems allow in vitro transcription and translation using the cellular machinery prepared from a variety of prokaryotic and eukaryotic cells. These non-GMO tools enable rapid and high-throughput characterization by their rapid gene expression, simple building of large combinatorial libraries, easier cloning, lower noise, less susceptibility to toxicity, and high tunability. In this chapter, different types of cell-free systems, different techniques to obtain them, and their applications in synthetic biology and bioengineering are presented.

A. Pandi (✉)
Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

Present Address: Department of Biochemistry & Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany
e-mail: amir.pandi@inra.fr; amir.pandi@mpi-marburg.mpg.de

O. Borkowski
iSSB Laboratory, Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

J.-L. Faulon (✉)
Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

iSSB Laboratory, Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

SYNBIOCHEM Center, School of Chemistry, University of Manchester, Manchester, UK
e-mail: jean-loup.faulon@inra.fr

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_16

Keywords
Cell-free systems · Cell-free protein translation (CFPT) · Cell-free transcription-translation (TX-TL) · Cell-free synthetic biology · In vitro synthetic biology

16.1 Introduction

Cell-free protein synthesis is an alternative tool to applying cellular protein production machinery beyond the living cells' growth and maintenance (Perez et al. 2016; Villarreal and Tan 2017). This tool has been used since the dawn of molecular biology for the discovery of genetic codes, mechanisms of the "genetic central dogma" (Chong 2014). Moreover, cell-free protein synthesis has provided an alternative for recombinant protein and toxic protein production (Carlson et al. 2012; Katzen et al. 2005).

Cell-free systems may perform only the protein translation from mRNA called "uncoupled" translation with regard to "coupled" in which both transcription and translation are processed in vitro from DNA (Katzen et al. 2005). Cell-free systems could be prepared either from cell lysate or using defined purified components depending on the applications.

In the last few years, cell-free systems have opened doors in the field of synthetic biology as a potential host/expression platform. Although cell-free systems are not sustainable as living systems, they provide advantages over whole-cell systems such as:

- Abiotic and non-GMO tools for bioproduction and diagnostic kits
- Fast and high-throughput prototyping and characterization of biological circuits and pathways because of the quicker gene expression and the ease of building combinatorial libraries without transforming cells (Jiang et al. 2018)
- High tunability provided by the membraneless system in which the components can be easily tuned by pipetting (Voyvodic et al. 2019)
- Easier cloning as there is no need to assemble multiple genes in one or two vectors. In cell-free systems, multiple plasmids with the same origin of replication and antibiotic resistance can be expressed at the same time. Linear DNA generated by PCR can also be used (Sun et al. 2014; Marshall et al. 2017; Nomoto and Tada 2018)
- Lower noise in gene expression and susceptibility to toxicity (Chang et al. 2017)
- Less limitation on the number of genes used in the pathways and circuits since there is no growth hence, no burden due to resource competition with the host (Pandi et al. 2019a).

All the above-mentioned advantages of cell-free systems have attracted the community's attention to describing a variety of protocols for different organisms (Table 16.1). Apart from the facilities to study biological mechanisms and rapid prototyping, cell-free systems provide a promising platform for metabolic engineering

Table 16.1 Different types of cell-free systems

	Organisms	Reference
Prokaryotic	*Escherichia coli*	Sun et al. (2013), Voloshin and Swartz (2005)
	Streptomyces venezuelae	Li et al. (2017)
	Vibrio natriegens	Wiegand et al. (2018), Failmezger et al. (2018), Des Soye et al. (2018), Wiegand et al. (2019)
	Bacillus subtilis	Kelwick et al. (2016)
	Bacillus megaterium	Kelwick et al. (2016), Moore et al. 2018)
	Pseudomonas putida	Wang et al. (2018), Yim et al. (2019)
	Escherichia fergusonii	Yim et al. (2019)
	Pantoea agglomerans	Yim et al. (2019)
	Corynebacterium glutamicum	Yim et al. (2019)
	Salmonella enterica	Yim et al. (2019)
	Klebsiella oxytoca	Yim et al. (2019)
	Lactococcus lactis	Yim et al. (2019)
Eukaryotic	*Wheat germ*	Takai et al. (2010)
	Rabbit reticulocyte	Jackson and Hunt (1983)
	Spodoptera frugiperda	Ezure et al. (2014)
	Leishmania tarentolae	Johnston and Alexandrov (2014)
	Human cells	Mikami et al. (2008)
	Saccharomyces cerevisiae	Hodgman et al. (2013)

and diagnosis as well as to construct synthetic cell in a bottom-up approach. In this chapter, we focus on the current protocols used to provide the extract as well as a variety of applications of the cell-free systems in the field on synthetic biology and bioengineering (Fig. 16.1).

16.2 Different Techniques to Obtain/Use Cell-Free Systems

The cell extract can be obtained through mechanical disruption of the membrane using sonication (Kwon and Jewett 2015; Shrestha et al. 2012), beads beating (Sun et al. 2013; Caschera and Noireaux 2014), or French press (Kim et al. 2006). Nevertheless, these methods required specific equipment: sonicator, shaker, or French press that are not available in every laboratory. Moreover, samples obtained using such method can be damaged by the high temperature/pressure reached during the process leading to protein denaturation. A chemical approach using lysozyme can also be used to obtain a lysate (Fujiwara and Doi 2016) avoiding steps in the protocols with high temperature/pressure. In a recent approach, autolysis strain has been developed with a protocol based on a constitutively expressed phage lambda endolysin coupled with a − 80 °C step to weaken the membrane (Didovyk et al. 2017).

An in vitro transcription and translation system can also be constructed in a bottom-up approach from defined required components called "purified recombi-

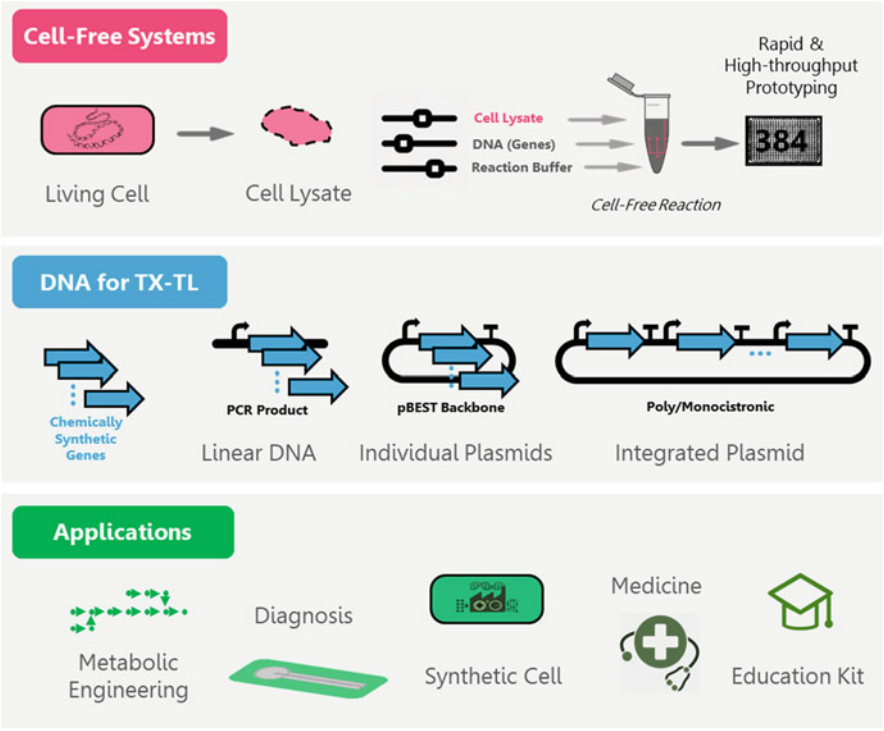


Fig. 16.1 Overview of the research field of cell-free synthetic biology. *Top panel*: Schematic representation of the preparation of TX-TL cell-free reaction. Extract is obtained from living cells and is used along with DNA (see middle panel) and reaction buffer (energy mix, amino acids, and nucleotides) to perform the cell-free reaction which can be characterized in a rapid and high throughput manner. *Middle panel*: DNA molecules used in transcription-translation cell-free reaction. DNA is expressed using transcription and translation machineries present in the cell-free reaction. The chemically synthesized gene (or amplified from an already existing DNA molecule) can be used as linear DNA (to save time and cost of cloning), individual plasmids (to have higher expression of genes with regard to linear DNA), or assembled plasmid (in cases such as incorporating a set of genes for synthetic cell application). Bottom panel: Schematic representation of the main applications of cell-free systems in the field of synthetic biology and bioengineering

nant elements" (PURE) system (Shimizu et al. 2001; Lavickova and Maerkl 2019). However, the PURE system is costly since 32 components of the transcription (RNA polymerase) and translation (translation initiation factors, elongation factors, release factor, aminoacyl-tRNA synthetases, methionyl-tRNA transformylase, ribosomes) machineries must be purified independently to make the functional system.

Cell-free components are mostly stocked in a liquid form at −20/−80 °C and the reaction is started at 30 °C or 37 °C when DNA is added to the mix. Liquid cell-free mixes cannot be conserved at room temperature but can be freeze-dried on paper and remain functional even after a year at room temperature (Pardee et al. 2014). Such, paper-based cell-free system has been developed using PURE system

or lysate-based cell-free system (Didovyk et al. 2017; Pardee et al. 2014). Cell-free reaction is then activated by adding DNA and water on paper making it easy to use, stock, and transport.

16.3 Applications of Cell-Free Systems

Cell-free systems, which emerged as tools to discover multiple biological mechanisms in the twentieth century, are becoming platforms for rapid and high-throughput characterization and prototyping of biological systems. Moreover, cell-free systems are applied for diagnostic devices and for the bottom-up synthetic cell construction. Here we present different categories of applications (Bottom panel, Fig. 16.1) with a few examples for each.

16.3.1 Metabolic Engineering

One of the applications of cell-free protein synthesis is the prototyping of metabolic pathways (Jiang et al. 2018; Karim and Jewett 2018). Easier cloning and the possibility of using PCR products make prototyping faster and more efficient. Since in vivo synthetic pathways can be toxic, cell-free systems have privileges over whole-cell systems through enabling expression of higher number of genes (as there is no resource competition with the host) and decreasing the harmful effects of intermediates (as there is no impact on cell physiology). In such an open system, multiple parameters such as the level of gene expression, the combination of different genes, and the concentration of different elements (reaction buffer composition) can be adjusted in an efficient design-build-test-learn cycle (Dudley et al. 2015). This ease of use makes cell-free a powerful tool for synthetic biologists and metabolic engineers to find new synthetic pathways as well as to optimize metabolic pathways (Dudley et al. 2015).

The prototyped metabolic pathway candidates with a higher performance can be transformed into whole-cell systems. In vivo, the genes should be cloned in a limited number of plasmids or integrated into the genome of the host. Since the properties of the host cell and the cell-free system are different, developing computational models will enable more predictable transfer from cell-free into in vivo chassis (Koch et al. 2018). For highly valuable/toxic products, the cell-free system itself might be used as the production chassis (Guo et al. 2017).

To perform a metabolic production in cell-free system, genes encoding enzymes can be added to a TX-TL cell-free extract supplied by the reaction buffer (Dudley et al. 2015). The enzymes can also be provided with a doped extract, a cell lysate prepared with the cells harboring a plasmid encoding a specific enzyme (Karim et al. 2018). However, doping of the extract with a multi-enzyme pathway reduces the growth and causes the burden in the cells used to prepare the extract. To avoid such issues, each enzyme can be expressed in a separate cell line and a mix of different extracts can be used to provide the multi-enzyme pathway in

cell-free (Karim et al. 2018). In a similar approach, purified enzymes can also be directly added to the reaction. Cell-free systems with DNA, doped extract, or supplied by purified enzymes have been used for bioproduction of psicose, violacein, 1,4-butanediol, polyhydroxyalkanoates, mevalonate, n-butanol, raspberry ketone, and limonene (Karim et al. 2018; Nguyen et al. 2016; Wu et al. 2015; Kelwick et al. 2018; Dudley et al. 2016; Moore et al. 2017; Dudley et al. 2019; Pandi et al. 2019). In a recent study, a biosensor screening method was developed to monitor the cell-free bioproduction (Dudley et al. 2019). Biosensors provide monitoring tools in metabolic engineering for pathway/enzyme optimization and screening through sensing the final products or intermediates (Koch et al. 2019). The biosensor development further speeds up the design-build-test-learn cycle of metabolic engineering using cell-free systems (Koch et al. 2019).

16.3.2 Biosensors and Diagnosis

Biosensor development for medical and environmental diagnosis is where the potential of cell-free systems in building portable abiotic kits plays a principal role (Chang et al. 2017; Soltani et al. 2018; Karig 2017; Gräwe et al. 2019). Cell-free systems allow building of abiotic and portable diagnosis kits that are safer and simpler to maintain and distribute. These kits keep their functionality after months when freeze-dried (Pardee et al. 2014). The low susceptibility of cell-free systems to the toxicity of chemicals and lower noise in gene expression with regard to living cells are other advantages of cell-free diagnostic devices. In addition, biosensors optimization can be facilitated by rapid prototyping and high-throughput characterization that these systems offer (delos Santos et al. 2016).

During the last decade, cell-free protein synthesis has been used to develop medical diagnostic devices. In an early study, the Collins' lab introduced a cell-free transcription-translation approach to build paper-based gene circuits (Pardee et al. 2014). Pardee et al. described a modular strategy to design and build toehold switches (gene circuits that respond to a short sequence of RNA when the small sequence of RNA opens the designed loop around the RBS and start codon) in cell-free systems (Pardee et al. 2014). As a proof of concept, they built multiple gene circuits for Ebola virus detection which were able to distinguish between viruses from two distinct populations. The same research group later extended their methodology and build cell-free devices for Zika virus as well (Pardee et al. 2016). A few years later they developed an in vitro method called SHERLOCK by employing high potential of Cas proteins (CRISPR machinery) to detect RNA and DNA sequences (Gootenberg et al. 2017; Gootenberg et al. 2018). In a recent study, toehold circuits were also applied to detect human gut microbiome composition in fecal samples (Takahashi et al. 2018).

The other approach to build cell-free biosensors is through transcriptional regulators. Wen et al. constructed a biosensor responding to quorum molecule of *Pseudomonas aeruginosa* along with its cognate transcription factor to detect this pathogen in clinical samples (Wen et al. 2017). In a recent study, Voyvodic et

al. proposed a modular way to extend the number of detectable molecules using metabolic enzymes in the cell-free system (Voyvodic et al. 2019). The enumerated pathway using computer-aided tools (Delépine et al. 2016) enables the conversion of an undetectable molecule to another which is a transcriptional or translational regulator. They optimized cell-free biosensors by adjusting the concentrations of DNA plasmids encoding the transcription factor, the GFP reporter gene, and the metabolic enzymes. Eventually, they used these sensors to detect cocaine and hippuric acid in clinical samples and benzoic acid in beverages (Voyvodic et al. 2019). Taking two or more biomarkers into consideration will increase the precision of the medical diagnosis. A sophisticated device called "metabolic perceptron" allows the integration of multiple signals for multiplex detection (Pandi et al. 2019a). The metabolic perceptron also brings an alternative approach to perform biological computation using biological circuits (Pandi et al. 2019a).

Cell-free biosensors can also be used for industrial and environmental applications (Karig 2017). The non-GMO diagnostic kit can be distributed to a wide geographical area as a cheap and easy way of detecting hazardous and pollutant molecules in the environment and industry. In a recent work, Alam et al. used RNA output sensors activated by ligand induction (ROSALIND) to detect pollutants in environmental water samples (Alam et al. 2019). They developed a modular strategy for the detection of different water pollutants such as antibiotics, toxic molecules, and metals (Alam et al. 2019). As an industrial/food example, Pandi et al. (Dudley et al. 2019) demonstrated that a repressor-based transcriptional sensor that suffers from low fold repression in the cell-free system can be optimized in several ways. Without optimizations a repressor-based system may exhibit weak fold change in the cell-free system. They introduced three strategies to do so: doping the extract with a transcription factor, preincubation of the extract with the components which is needed to be in excess (the repressor), and reinitiation of the cell-free reaction when the system's ability in gene expression diminishes. They then used the optimized biosensor of psicose to monitor its bioproduction from fructose using a metabolic enzyme.

16.3.3 Studying Biological Mechanisms

The ability of cell-free systems to perform minimal biological functions without the need to express a full genome makes them a promising tool to study specific mechanisms independently. In a recent attempt, *E. coli* TX-TL system was used to predict the cost of protein expression in living cells (Borkowski et al. 2018). The authors proposed a standard cell-free assay to relatively measure the resource consumption of the expression of a protein sequence. In this approach, the in vivo burden of growing cells expressing a variety of proteins and multigene operons can be predicted (Borkowski et al. 2018). In another study, the cell-free system was used to study the CRISPR mechanisms such as characterization of gRNAs and anti-CRISPR proteins (Marshall et al. 2018).

16.3.4 Building a Synthetic Cell

Building synthetic cells is one of the main goals of synthetic biology to understand the minimal elements necessary for life. The synthetic cell can be applied as a universal minimal chassis in systems and synthetic biology and for medical applications such as drug delivery (Xu et al. 2016). There are two approaches to build a minimal cell: top-down and bottom-up (Fritz et al. 2010). In the top-down path, the genome of an existing organism, preferably an organism which is well-known and/or with a small genome such as *E. coli* or *Mycoplasma mycoides* is reduced (Hutchison 3rd et al. 2016). In the bottom-up construction, the minimal components are assembled from scratch to build a system which is sustainable, can divide and interact with its environment (Göpfrich et al. 2018). The encapsulation of the minimal system harboring the genetic material for necessary functions is similar to how life emerged on earth more than 3 billion years ago (Schwille et al. 2018). The same process can be used in synthetic biology to encapsulate a cell-free system and build a synthetic minimal living system. Attempts toward creating synthetic cells using a minimal cell-free system or a lysate were able to demonstrate living cells-like behavior (Bhattacharya et al. 2019; Berhanu et al. 2019). Vogele et al. succeeded in encapsulating the TX-TL cell-free system with amphipathic peptides as the membrane (Vogele et al. 2018). They then used the gene that expresses the amphipathic peptide coding its membrane to extend the size of encapsulated system (Vogele et al. 2018).

16.3.5 Self-Assembly of Phages

Transcription-translation cell-free systems have been employed to assemble and amplify a number of phagemids (Pires et al. 2016). The phagemid assembly has been done in one-pot reaction from the genome of MS2, ΦX174, and T7 (Shin et al. 2012; Rustad et al. 2017). In a recent work, the complete T4 phage has been synthesized from its 169-kbp genome in single TX-TL reaction (Rustad et al. 2018). This achievement shows that genomes can be functionally expressed to build grand organized systems in vitro.

16.3.6 Medicine and Therapeutics

One of the earliest applications of the cell-free protein synthesis was the production of biologically active proteins (Katzen et al. 2005; Lee and Kim 2018). Key challenges are a correct protein folding and post-translational modifications as observed in native cells (Tokmakov et al. 2012). Since cell-free systems are open platforms, their components can be easily adjusted. For example, redox buffers can be used to control the disulfide bond formation (Matsuda et al. 2013). Post-translational modifications such as phosphorylation and glycosylation can also be

performed to produce functional proteins in vitro (Oza et al. 2015; Jenkins et al. 2008; Jaroentomeechai et al. 2018). The majority of commercially available technologies for these types of applications use mammalian cell-free systems. Mammalian cell-free systems are able to implement post-translational modifications which are necessary for many therapeutic proteins. However, a recent study described a method to implement glycosylation (a common post-translational modification of eukaryotic proteins) using bacterial transcription-translation system (Jaroentomeechai et al. 2018). This achievement brings a cheap and more efficient bacterial cell-free system for medical applications.

16.3.7 Proteomics and Protein Evolution

By decreasing the cost of cell-free systems, using automation and optimization, these systems allow high-throughput protein synthesis and characterizations (Rolf et al. 2019; Sawasaki et al. 2002). This trend brings the advantage of applying cell-free protein synthesis for proteomics analysis (Katzen et al. 2005). Moreover, cell-free systems enable directed evolution to generate proteins with desired phenotypes especially through applications such as ribosome display, in vitro compartmentalization, and in vitro virus (also known as RNA-peptide fusion or mRNA display) (Katzen et al. 2005).

16.3.8 Education Kit

The development of education kits in the field of biology is limited compared to other branches of science and engineering due to the obstacles of dealing with living organisms and the lack of portable and affordable devices. BioBits™ (Nguyen et al. 2018) is a collection of freeze-dried educational kit established recently by leveraging cell-free transcription-translation systems. This collection provides portable non-GMO kits for young students to learn and practice synthetic biology. BioBits™ kit consists of simple transcription and translation set of fluorescent proteins and more complex devices for enzymatic reactions and RNA responding circuits. The components of the kit are easily usable after adding water to freeze-dried cell-free systems. All the experiments are practiced by employing only the senses of sight, smell, and touch through outputs that produce fluorescence, fragrances, and hydrogels, respectively. The DIY collection provided by the kit makes it affordable and valuable for young students to get trained and learn molecular biology and synthetic biology.

16.4 Perspectives

Cell-free systems have an old history since the dawn of molecular biology and will have a future through their peculiar properties. Since high-throughput and engineering approaches are getting integrated into life science applications, cell-free systems play an important role in studying, prototyping, and engineering biological systems. In the coming years, the preparation of various cell-free systems should get more affordable and standardized. The limitations as the lack of post-translational modifications in prokaryotic systems or the lack of transcription machinery in eukaryotic systems can be compensated by adding specific components like kinase or T7 polymerase to the cell extracts (Gan and Jewett 2014). Moreover, new achievement such as glycosylation using bacterial extract is a cutting edge advancement in cell-free biology. Cell-free systems have extended their shadows in the whole field of biology from basic science to building sophisticated synthetic devices and synthetic cells. With the achievement gained in the twenty-first century, the cell-free synthetic biology has a bright future for medical, environmental, and industrial applications.

Acknowledgments AP is supported by INRA (French National Institute for Agricultural Research) and idEx interdisciplinary scholarship from Paris-Saclay. OB is supported by Genopole "Allocation Recherche 2017" and CRI Paris "Short-term Fellows." JLF acknowledges support from BBSRC/EPSRC (grant number BB/M017702/1) and ANR (grant number ANR-18-CE33-0015).

References

Alam KK, Jung JK, Verosloff MS, Clauer PR, Lee JW, Capdevila DA, et al. (2019). Rapid, Low-cost detection of water contaminants using regulated in vitro transcription. BioRxiv:619296. https://doi.org/10.1101/619296.

Berhanu S, Ueda T, Kuruma Y (2019) Artificial photosynthetic cell producing energy for protein synthesis. Nat Commun 10:1325

Bhattacharya A, Brea RJ, Niederholtmeyer H, Devaraj NK (2019) A minimal biochemical route towards de novo formation of synthetic phospholipid membranes. Nat Commun 10:300

Borkowski O, Bricio C, Murgiano M, Rothschild-Mancinelli B, Stan G-B, Ellis T (2018) Cell-free prediction of protein expression costs for growing cells. Nat Commun 9:1457

Carlson ED, Gan R, Hodgman CE, Jewett MC (2012) Cell-free protein synthesis: applications come of age. Biotechnol Adv 30:1185–1194

Caschera F, Noireaux V (2014) Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. Biochimie 99:162–168

Chang H-J, Voyvodic PL, Zúñiga A, Bonnet J (2017) Microbially derived biosensors for diagnosis, monitoring and epidemiology. Microb Biotechnol 10:1031–1035

Chong S (2014) Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. Curr Protoc Mol Biol 108:16.30.1–16.3011

Delépine B, Libis V, Carbonell P, Faulon J-L (2016) SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Res 44:W226–W231

delos Santos ELC, Meyerowitz JT, Mayo SL, Murray RM (2016) Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: developing a vanillin sensor. ACS Synth Biol 5:287–295

Des Soye BJ, Davidson SR, Weinstock MT, Gibson DG, Jewett MC (2018) Establishing a high-yielding cell-free protein synthesis platform derived from Vibrio natriegens. ACS Synth Biol 7:2245–2255

Didovyk A, Tonooka T, Tsimring L, Hasty J (2017) Rapid and scalable preparation of bacterial lysates for cell-free gene expression. ACS Synth Biol 6:2198–2208

Dudley QM, Karim AS, Jewett MC (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J 10:69–82

Dudley QM, Anderson KC, Jewett MC (2016) Cell-free mixing of Escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. ACS Synth Biol 5:1578–1588

Dudley QM, Nash CJ, Jewett MC. Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates. Synth Biol 2019;4, ysz003. https://doi.org/10.1093/synbio/ysz003.

Ezure T, Suzuki T, Ando E. A Cell-free protein synthesis system from insect cells. Methods Mol Biol 2014, 1118:285–296. https://doi.org/10.1007/978-1-62703-782-2_20.

Failmezger J, Scholz S, Blombach B, Siemann-Herzberg M. Cell-free protein synthesis from fast-growing Vibrio natriegens. Front Microbiol 2018;9, 1146. https://doi.org/10.3389/fmicb.2018.01146.

Fritz BR, Timmerman LE, Daringer NM, Leonard JN, Jewett MC (2010) Biology by design: from top to bottom and back. J Biomed Biotechnol 2010:232016

Fujiwara K, Doi N (2016) Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. PLoS One 11:e0154614. https://doi.org/10.1371/journal.pone.0154614

Gan R, Jewett MC (2014) A combined cell-free transcription-translation system from Saccharomyces cerevisiae for rapid and robust protein synthe. Biotechnol J 9:641–651

Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J et al (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356:438–442

Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 360:439–444

Göpfrich K, Platzman I, Spatz JP (2018) Mastering complexity: towards bottom-up construction of multifunctional eukaryotic synthetic cells. Trends Biotechnol 36:938–951

Gräwe A, Dreyer A, Vornholt T, Barteczko U, Buchholz L, Drews G et al (2019) A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. PLoS One 14:e0210940

Guo W, Sheng J, Feng X. Mini-review: in vitro metabolic engineering for biomanufacturing of high-value products. Comput Struct Biotechnol J 2017;15:161–167. https://doi.org/10.1016/j.csbj.2017.01.006.

Hodgman CE, Eric Hodgman C, Jewett MC. Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. Biotechnol Bioeng 2013;110:2643–2654. https://doi.org/10.1002/bit.24942.

Hutchison CA 3rd, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH et al (2016) Design and synthesis of a minimal bacterial genome. Science 351:aad6253

Jackson RJ, Hunt T (1983) Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. Methods Enzymol 96:50–74

Jaroentomeechai T, Stark JC, Natarajan A, Glasscock CJ, Yates LE, Hsu KJ et al (2018) Author correction: single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nat Commun 9:3396

Jenkins N, Murphy L, Tyther R. Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. Mol Biotechnol 2008;39:113–118. https://doi.org/10.1007/s12033-008-9049-4.

Jiang L, Zhao J, Lian J, Xu Z (2018) Cell-free protein synthesis enabled rapid prototyping for metabolic engineering and synthetic biology. Synth Syst Biotechnol 3:90–96

Johnston WA, Alexandrov K (2014) Production of eukaryotic cell-free lysate from Leishmania tarentolae. Methods Mol Biol 1118:1–15

Karig DK (2017) Cell-free synthetic biology for environmental sensing and remediation. Curr Opin Biotechnol 45:69–75
Karim AS, Jewett MC (2018) Cell-free synthetic biology for pathway prototyping. Methods Enzymol 608:31–57
Karim AS, Heggestad JT, Crowe SA, Jewett MC (2018) Controlling cell-free metabolism through physiochemical perturbations. Metab Eng 45:86–94
Katzen F, Chang G, Kudlicki W (2005) The past, present and future of cell-free protein synthesis. Trends Biotechnol 23:150–156
Kelwick R, Webb AJ, MacDonald JT, Freemont PS (2016) Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements. Metab Eng 38:370–381
Kelwick R, Ricci L, Chee SM, Bell D, Webb AJ, Freemont PS. Cell-free prototyping strategies for enhancing the sustainable production of polyhydroxyalkanoates bioplastics. Synth Biol, *3*(1), ysy016 2018. https://doi.org/10.1093/synbio/ysy016.
Kim T-W, Keum J-W, Oh I-S, Choi C-Y, Park C-G, Kim D-M (2006) Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. J Biotechnol 126:554–561
Koch M, Faulon J-L, Borkowski O (2018) Models for cell-free synthetic biology: make prototyping easier, better, and faster. Front Bioeng Biotechnol 6:182
Koch M, Pandi A, Borkowski O, Batista AC, Faulon J-L. Custom-made transcriptional biosensors for metabolic engineering. Curr Opin Biotechnol 2019;59:78–84. https://doi.org/10.1016/j.copbio.2019.02.016.
Kwon Y-C, Jewett MC (2015) High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Sci Rep 5:8663
Lavickova B, Maerkl SJ (2019) A simple, robust, and low-cost method to produce the PURE cell-free system. ACS Synth Biol 8:455–462
Lee K-H, Kim D-M (2018) Recent advances in development of cell-free protein synthesis systems for fast and efficient production of recombinant proteins. FEMS Microbiol Lett 365:fny174. https://doi.org/10.1093/femsle/fny174
Li J, Wang H, Kwon Y-C, Jewett MC (2017) Establishing a high yielding streptomyces-based cell-free protein synthesis system. Biotechnol Bioeng 114:1343–1353
Marshall R, Maxwell CS, Collins SP, Beisel CL, Noireaux V (2017) Short DNA containing χ sites enhances DNA stability and gene expression in E. coli cell-free transcription-translation systems. Biotechnol Bioeng 114:2137–2141. https://doi.org/10.1002/bit.26333
Marshall R, Maxwell CS, Collins SP, Luo ML, Jacobsen T, Beisel C, et al. Rapid and scalable characterization of CRISPR technologies using an E. coli cell-free transcription-translation system Mol Cell. 2018;69(1):146-157.e3 https://doi.org/10.1101/169441.
Matsuda T, Watanabe S, Kigawa T (2013) Cell-free synthesis system suitable for disulfide-containing proteins. Biochem Biophys Res Commun 431:296–301
Mikami S, Kobayashi T, Masutani M, Yokoyama S, Imataka H (2008) A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. Protein Expr Purif 62:190–198
Moore SJ, Tosi T, Hleba YB, Bell D, Polizzi K, Freemont P (2017) A cell-free synthetic biochemistry platform for raspberry ketone production. BioRxiv:202341. https://doi.org/10.1101/202341.
Moore SJ, MacDonald JT, Wienecke S, Ishwarbhai A, Tsipa A, Aw R et al (2018) Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. Proc Natl Acad Sci U S A 115:E4340–E4349
Nguyen PHB, Wu Y, Guo S, Murray RM (2016) Design Space Exploration of the Violacein Pathway in Escherichia coli Based Transcription Translation Cell-Free System (TX-TL). BioRxiv:027656. https://doi.org/10.1101/027656.
Huang A, Nguyen PQ, Stark JC, Takahashi MK, Donghia N, Ferrante T et al (2018) BioBits™ explorer: a modular synthetic biology education kit. Sci Adv 4:eaat5105

Nomoto M, Tada Y (2018) Cloning-free template DNA preparation for cell-free protein synthesis via two-step PCR using versatile primer designs with short 3'-UTR. Genes Cells 23:46–53

Oza JP, Aerni HR, Pirman NL, Barber KW, TerHaar CM, Rogulina S et al (2015) Robust production of recombinant phosphoproteins using cell-free protein synthesis. Nat Commun 6:8168

Pandi A, Koch M, Voyvodic PL, Soudier P, Bonnet J, Kushwaha M, et al (2019a) Metabolic perceptrons for neural computing in biological systems. Nature Commun 10(1): 3880 https://doi.org/10.1038/s41467-019-11889-0.

Pandi A, Grigoras I, Borkowski O, Faulon JL (2019b) Optimizing Cell-Free Biosensors to Monitor Enzymatic Production. ACS Synth Biol 8(8):1952–1957

Pardee K, Green AA, Ferrante T, Cameron DE, DaleyKeyser A, Yin P et al (2014) Paper-based synthetic gene networks. Cell 159:940–954

Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW et al (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165:1255–1266

Perez JG, Stark JC, Jewett MC. Cell-free synthetic biology: engineering beyond the cell. Cold Spring Harb Perspect Biol 2016;8: a023853. https://doi.org/10.1101/cshperspect.a023853.

Pires DP, Cleto S, Sillankorva S, Azeredo J, Lu TK (2016) Genetically engineered phages: a review of advances over the last decade. Microbiol Mol Biol Rev 80:523–543

Rolf J, Rosenthal K, Lütz S (2019) Application of cell-free protein synthesis for faster biocatalyst development. Catalysts 9:190. https://doi.org/10.3390/catal9020190

Rustad M, Eastlund A, Marshall R, Jardine P, Noireaux V. Synthesis of infectious bacteriophages in an E. coli-based cell-free expression system. J Vis Exp 2017, 126: e56144. https://doi.org/10.3791/56144.

Rustad M, Eastlund A, Jardine P, Noireaux V. Cell-free TXTL synthesis of infectious bacteriophage T4 in a single test tube reaction. Synth Biol 2018;3: ysy002. https://doi.org/10.1093/synbio/ysy002.

Sawasaki T, Ogasawara T, Morishita R, Endo Y. A cell-free protein synthesis system for high-throughput proteomics. Proc Natl Acad Sci 2002;99:14652–14657. https://doi.org/10.1073/pnas.232580399.

Schwille P, Spatz J, Landfester K, Bodenschatz E, Herminghaus S, Sourjik V et al (2018) MaxSynBio: avenues towards creating cells from the bottom up. Angew Chem Int Ed Engl 57:13382–13392

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K et al (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755. https://doi.org/10.1038/90802

Shin J, Jardine P, Noireaux V. Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth Biol 2012;1:408–413. https://doi.org/10.1021/sb300049p.

Shrestha P, Holland TM, Bundy BC (2012) Streamlined extract preparation for Escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. Bio Techniques 53:163–174

Soltani M, Davis BR, Ford H, Nelson JAD, Bundy BC. Reengineering cell-free protein synthesis as a biosensor: biosensing with transcription, translation, and protein-folding. Biochem Eng J 2018;138:165–171. https://doi.org/10.1016/j.bej.2018.06.014.

Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J Vis Exp 79:e50762

Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM (2014) Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. ACS Synth Biol 3:387–397. https://doi.org/10.1021/sb400131a

Takahashi MK, Tan X, Dy AJ, Braff D, Akana RT, Furuta Y et al (2018) A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. Nat Commun 9:3347

Takai K, Sawasaki T, Endo Y (2010) Practical cell-free protein synthesis system using purified wheat embryos. Nat Protoc 5:227–238

Tokmakov AA, Kurotani A, Takagi T, Toyama M, Shirouzu M, Fukami Y et al (2012) Multiple post-translational modifications affect heterologous protein synthesis. J Biol Chem 287:27106–27116

Villarreal F, Tan C (2017) Cell-free systems in the new age of synthetic biology. Front Chem Sci Eng 11:58–65. https://doi.org/10.1007/s11705-017-1610-x

Vogele K, Frank T, Gasser L, Goetzfried MA, Hackl MW, Sieber SA et al (2018) Towards synthetic cells using peptide-based reaction compartments. Nat Commun 9:3862

Voloshin AM, Swartz JR (2005) Efficient and scalable method for scaling up cell free protein synthesis in batch mode. Biotechnol Bioeng 91:516–521

Voyvodic PL, Pandi A, Koch M, Conejero I, Valjent E, Courtet P et al (2019) Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. Nat Commun 10:1697

Wang H, Li J, Jewett MC. Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. Synth Biol 2018;3, ysy003. https://doi.org/10.1093/synbio/ysy003.

Wen KY, Cameron L, Chappell J, Jensen K, Bell DJ, Kelwick R et al (2017) A cell-free biosensor for detecting quorum sensing molecules in P. aeruginosa-infected respiratory samples. ACS Synth Biol 6:2293–2301. https://doi.org/10.1021/acssynbio.7b00219

Wiegand DJ, Lee HH, Ostrov N, Church GM (2018) Establishing a cell-free Vibrio natriegens expression system. ACS Synth Biol 7:2475–2479

Wiegand DJ, Lee HH, Ostrov N, Church GM. Cell-free protein expression using the rapidly growing bacterium vibrio natriegens. J Vis Exp 2019, 145. e59495 https://doi.org/10.3791/59495.

Wu YY, Culler S, Khandurina J, Van Dien S, Murray RM (2015) Prototyping 1,4-butanediol (BDO) biosynthesis pathway in a cell-free transcription-translation (TX-TL) system. BioRxiv:017814. https://doi.org/10.1101/017814.

Xu C, Hu S, Chen X. Artificial cells: from basic science to applications. Mater Today 2016;19:516–532. https://doi.org/10.1016/j.mattod.2016.02.020.

Yim SS, Johns NI, Park J, Gomes ALC, McBee RM, Richardson M et al (2019) Multiplex transcriptional characterizations across diverse and hybrid bacterial cell-free expression systems. Mol Syst Biol 15:e8875. https://doi.org/10.1101/427559

Synthetic Biology for the Rapid, Precise and Compliant Detection of Microbes

17

Joseph P. Wheatley, Sahan B. W. Liyanagedera, Richard Amaee, Antonia P. Sagona, and Vishwesh Kulkarni

Abstract

Since the turn of the millennium, an extensive range of applications spanning across science and technology have been turning to synthetic biology for inspiration and innovation in order to enhance, accelerate and permeate their specific fields. The field of microbial detection is no different. Modern synthetic biology offers pioneering approaches to accelerate and fine-tune the detection process, subsequently enabling clinicians to offer fast, targeted treatments. This, essentially, will dramatically reduce medical burden across a wide array of diseases and will provide a crucial step towards limiting the use of antibiotic treatment to completely necessary cases only. In this chapter, some of the key synthetic biology-inspired approaches that are transforming the field of microbial detection are explored, with an emphasis on what the near future holds.

Keywords

Synthetic biology · Bacteriophage · Biosensor · Diagnostic · CRISPR

J. P. Wheatley · S. B. W. Liyanagedera
School of Engineering, University of Warwick, Coventry, UK

School of Life Sciences, University of Warwick, Coventry, UK
e-mail: joseph.wheatley.1@warwick.ac.uk

R. Amaee · A. P. Sagona
School of Life Sciences, University of Warwick, Coventry, UK

V. Kulkarni
School of Engineering, University of Warwick, Coventry, UK

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_17

17.1 Introduction

Synthetic biology (SynBio) provides a platform that enables the rewiring of natural biological parts, thereby giving rise to optimised and predictable entities capable of performing tasks to a degree that is unmatchable in the natural world. The field makes use of engineering-like principles in order to view biological systems as computable devices that can efficiently complete complex tasks whilst allowing a high level of control over all parameters. The comparison of biological systems to classical engineering circuitry was first made in the early 1960s, where a suggestion was made that 'teleonomic mechanisms' within cells operated in a more linear fashion than had first been perceived (Monod and Jacob 1961). This novel idea that biological organisms employed a command-and-control-like regulatory system, and therefore had the capacity to be externally controlled and modified much like a computational device, impelled the first philosophies that formed the foundations of SynBio.

Since the advent of SynBio, the methods used within the field have been significantly improved. Scientists now have access to high-throughput, state-of-the-art technology *and* expert knowledge to assist in tasks such as next-generation sequencing, genome engineering and the creation of synthetic gene networks. These available resources, coupled with the general SynBio principles, are well suited to the field of microbial detection. The enhancement of detection techniques will rely on increased speed, preciseness and accessibility; furthermore, these are all features that can be optimised and scaled-up via SynBio technologies. Whether it be through complex gene networks (Sedlmayer et al. 2018; Riglar and Silver 2018; Takahashi et al. 2018), engineered bacteriophages (Alcaine et al. 2016; Chen et al. 2017a, b; Wang et al. 2017a; Sharp et al. 2016) or biomaterials (Wang et al. 2017b; Li et al. 2017), microbial detection for applications including medical diagnosis and food safety is in high demand. With regard to microbial detection in the food industry, the inquiries undergone within this market constitute over 1B annual tests with a combined value of $3.5B and a compound annual growth rate (CAGR) of 7.7% (MDF 2018–2023). Six of the top seven tested pathogens are bacterial. This market is currently serviced by a mix of in situ ELISA and PCR devices which are slow, complex and inconvenient for end users. The current gold standard—cell culture—is also cumbersome and slow (Bhunia 2014), taking 48–72 h for preliminary results (Hagens and Loessner 2007).

Recent statistics state that antimicrobial-resistant 'superbugs' are the cause of approximately 700,000 deaths each year (Rappuoli et al. 2017). Furthermore, it is predicted that if the overwhelming surge of antimicrobial resistance is not seriously tackled over the next 30 years, then it will overtake cancer in the number of fatalities caused—with potential death tolls reaching ten million per annum (O'neill 2014). Antimicrobial resistance is a natural process; however, the misuse of antibiotics in humans and animals is dramatically accelerating the process at a dangerous rate. One approach for reducing antimicrobial resistance is to not disburse antibiotics in the first instance unless they are completely necessary. The problem with this

approach, however, is that it is impossible to know for certain whether or not antibiotics are necessary unless the microbes causing the infection are identified. Currently, there is no rapid diagnostic tool available for knowing precisely what microbe has caused an infection. Consequently, it is common that antibiotics are prescribed in situations where they will be completely ineffective. This perennial problem could be overcome if medical professionals would be able to rapidly diagnose which, if any, microbe has caused an infection. Armed with that information, a targeted, patient/microbe-specific—not symptom-specific—therapeutant could be administered, and this would be desirable not only for the individual patient but also, in a more general sense, in the fight against antimicrobial resistance. In some cases, this therapeutant may well be in the form of antibiotics; however, practising within these means is likely to lead to a reduction overall in antibiotic use by eradicating their misuse. Subsequently, this will have a considerably positive effect on reducing antimicrobial resistance worldwide.

If the detection process can be rapid and unambiguous, then the monetary and corporeal cost of disease can be significantly reduced worldwide. Correspondingly, since the vast majority of microbial tribulations are caused by either bacteria or viruses, the detection of these microbes is principally what this chapter will explore. SynBio offers an extensive toolbox that can, and will, be used to develop systems for the rapid, precise and compliant detection of microbes.

17.2 Utilising Cell-Free Expression Systems for Engineering Genetic Detection Programmes

17.2.1 Cell-Free: A Naturalistic Environment with No Restraints

Cell-free *transcription-translation* (TXTL) systems have been engineered for use in multiple applications across the field of SynBio. These applications include facilitating and accelerating the design-build-test cycle for genetic circuits, ranging from simple DNA circuits to complex gene networks (Siegal-Gaskins et al. 2014; Takahashi et al. 2015; Garamella et al. 2016), assembling whole bacteriophages in a cell-free environment (Rustad et al. 2017, 2018; Shin et al. 2012), metabolic engineering (Dudley et al. 2015), biomanufacturing (Bundy et al. 2018; You and Zhang 2013) and medical research (Ng et al. 2012; Ogonah et al. 2017). TXTL harnesses the endogenous transcriptional and translational machinery extracted from bacterial cells, commonly *E. coli*, and combines this cellular hardware with an energy solution and amino acid mix, allowing for the expression of DNA in a single cell-free reaction (Sun et al. 2013; Didovyk et al. 2017).

17.2.2 Genetic Circuits

The design and implementation of nucleic acid-based circuits is becoming one of the most rapidly growing areas within cell-free SynBio. Genetic circuits are

generally comprised of two domains: a sensor domain, which senses and responds to an input, and an actuator domain, which generates an output in the presence of a compatible input. The output of one circuit is often designed to become the input of the next; thus, complex circuitry networks can be engineered to have a vast array of outcomes. Although a cell-free system is not completely necessary for the mere implementation of gene circuits, by utilising TXTL, circuit cascades can be designed, tested, debugged and redesigned in a matter of days—a luxury that would not be easily apprehended in the absence of a cell-free environment. This approach has immensely accelerated both the magnitude *and* fidelity of research in this field. Furthermore, TXTL significantly expands and helps comprehend the real-world applications that are imminent for genetic circuits—with a particular relevance to microbial diagnostics.

Boolean logic-inspired genetic circuits are commonplace within SynBio. Exploiting Boolean theory in bioscience allows for biological processes to be analysed with an engineering perspective, which is essential when building a DNA computing system. Various digital logic gates have been experimentally achieved, including AND, NAND, OR and XOR (Zhang et al. 2019). In a biological context, a simple visualisation of the process by which these gates are achieved is through the expression of a plasmid. A small protein/molecule input (e.g. transcription factor, bacterial quorum sensing molecule) interacts with the sensor domain of the gate (e.g. promoter region), subsequently turning ON or OFF expression (n.b. for AND or NAND gate 2 inputs are required). The gene product, or lack of gene product, expressed from the plasmid can either elicit the overall output response or continue in the circuitry pathway and stimulate other gates to either switch ON or OFF. This ability to add layers of complexity to the programme allows for multiple factors to affect how the circuit will progress and makes it possible for multiple outcomes to be achieved.

It is also possible to design circuits where genetic material is the input or output. For example, in a toehold riboswitch, the signal-dependent conformational change in the RNA toehold structure is only activated when the aptamer (sensor) domain has detected the input, which could be a sequence-specific RNA strand or another small-molecule ligand depending on the switch design (Sherwood and Henkin 2016). DNA strand displacement (DSD) reactions are another assortment of gene circuits that use genetic material as their currency. DSD relies on the natural ability of ssDNA to displace a hybridised strand if it has a higher affinity to the parallel strand. These reactions arise through a toehold-mediated branch migration displacement mechanism fuelled by classic Watson-Crick base pairing (Sawlekar et al. 2016).

17.2.2.1 Bacterial Communication Utilised as a Circuitry Input

The key for designing a genetic bacterial detection programme is quorum sensing (QS). QS is a molecular language for cell-cell communication used by bacteria. The fundamental process of bacterial QS is the auto-production of chemical signals (e.g. N-acyl homoserine lactones—AHL) that, once accumulated to a certain threshold in the local environment, will induce gene expression via transcriptional regulator interactions (Whiteley et al. 2017). This aspect of the sociobiology of bacteria has

been widely unexplored until recent times, and it is only now that we are realising the full potential of this phenomenon. Succinctly, by engineering a QS-sensitive promoter to the sensor domain of a genetic detection device, the downstream expression of that cassette can be directly controlled by the presence of QS molecules, and therefore it is directly controlled by the presence of bacteria.

A well-characterised QS system used in gene circuit biology is the interaction between AHL and LuxR-type receptors. Bacterial expression of the LuxI gene leads to high levels of AHL in the local environment. AHL subsequently binds to its receptor, LuxR, and once the level of this interaction reaches a certain threshold, the expression of the gene/s under the control of the LuxR regulator is induced (Pedrolli et al. 2019). By using this genetic template, it is possible to design, for example, an AND gate plasmid whereby two inputs, one of which being bacterial-produced AHL, must be present in order for the induction of downstream gene expression.

A fundamental bottleneck that is restraining this research from being applied in real-world situations is the fact that the output of many of these circuits is a fluorescent protein such as GFP (Prindle et al. 2011; Garcia-Ojalvo et al. 2016). These fluorescent proteins aren't easily visible to the naked eye unless they are highly expressed, which often isn't the case when their expression is the result of a genetic circuit. If the output of the circuit could instead be a protein that is involved in a colourimetric enzymatic reaction, much like in previously published viral detection circuits (Pardee et al. 2016; Ma et al. 2018), then this would produce a fast, visible output that determines whether bacteria are present or not.

17.2.2.2 Circuit-Mediated Viral Detection

A virus' virulence rests on its ability to inject its genetic material into a host and then use the host's nucleic acid-expressing hardware to run its viral software in order to assemble its progeny, which proceeds to kill and escapes the host in search for a new victim to restart the cycle. The detectable constituent of a virion is its genetic material—DNA or RNA depending on the species.

One of the most impressive recent innovations to rise from the field of microbial detection is the lyophilised, paper-based, cell-free Zika virus RNA genome sensor (Pardee et al. 2016). This genetically programmed device, based on a toehold riboswitch, is instigated by the addition of isothermally amplified viral RNA. The riboswitch detects the presence of Zika virus RNA and, through a structural change, subsequently triggers the downstream transcription of LacZ—an enzyme that catalyses the colourimetric conversion of a yellow substrate (chlorophenol red-β-D-galactopyranoside) to a purple product (chlorophenol red), which indicates whether Zika virus is present or not. This technology has recently been rewired to detect norovirus (Ma et al. 2018) and also gamma-hydroxybutyrate—a substance used as date-rape drug (Grawe et al. 2019). These applications of state-of-the-art SynBio research display the true versatility of TXTL-based genetic detection programmes.

17.3 Smart Biomaterials: The Future of Microbial Detection?

The building of a SynBio-based smart material for the purpose of microbial detection requires an appreciation of the current trends and goals of the field itself. The ultimate aim of SynBio is to create and modify synthetic cells to understand the evolution and functionality of all life, as well as to engineer and exploit new life forms. Through the encapsulation and utilisation of TXTL systems and genetic circuits within biomaterials, it is conceivable to devise a smart biomaterial that emulates the functionality of a living cell. These synthetic entities can subsequently be programmed to perform tasks such as detecting microbes; and it is the physical materialisation of these processes that provides the potential to take them from the research laboratories to being applied in the real world. The current state of the art displays a twofold approach for these purposes. The top-down approach targets the synthesis of a minimal genetic network that is able to sustain life, as well as the modulation of metabolic pathways for biotechnological applications. The bottom-up approach is focused on the de novo design of cells, through creation of minimal structures that are capable of mimicking complex cellular functions. The application of SynBio to build smart materials falls into the latter bottom-up approach, as seen in the rapid advances in the field over the last few years (Adamala et al. 2017; Rampioni et al. 2018). The recent advances in the field and the emergence of hydrogels as a highly biocompatible and tunable substrate have paved the way for the creation of programmable soft materials. The rate of the continual expansion of the SynBio toolbox and innovations for the incorporation of such modules into materials of interest is what will determine the growth of this nascent field.

17.3.1 Biomimetic Design

The recent trends of smart materials based on biomimetic design are reflected in a general cultural shift towards the borrowing of engineering solutions that are in line with nature. In this respect, biomimetic smart material design is becoming ever more steadfast in their simulation of the biological environments of the cell. Initial attempts saw the creation of hydrogels that simply mimicked the structural integrity of naturally occurring biomaterials such as the extracellular matrix (ECM).

The ECM is a naturally occurring biomaterial that has been mimicked in the designs of synthetic hydrogels (Sackett et al. 2018; Kim et al. 2018). The ECM's role as the skeletal support that integrates nutrition, oxygen supply, signal transduction and waste removal for correct cellular function is well understood (Vogel 2018). Whilst current hydrogels are unable to facilitate such complex biological phenomena, the advent of DNA hydrogels has helped realise materials that have comparable physical properties to ECMs (Li et al. 2015; Zhou et al. 2016). Recent work builds on these advances by trying to emulate the ECM's ability to fine-tune its mechanical strength in a pre-programmed manner (Liu et al. 2018; Turturro et al. 2013).

In one such study, an ATP aptamer sequence was integrated into a DNA hydrogel to create a material with mechanical properties that could be tuned in three phases in situ (Liu et al. 2018). The single-stranded ATP aptamer sequence was inserted into one of the five structural DNA sequences that formed the hydrogel network. The high permeability of the hydrogel allowed ATP to diffuse easily through the system and bind to its aptamer, leading to a change in conformation of the aptamer from a single strand to a more rigid double-stranded structure. The conformational change, and the resulting shortening of the distance between the cross-linking points of the hydrogel, leads to an increased storage modulus (204–380 Pa). The third tuned mechanical phase can be achieved by the addition of a complementary sequence to the aptamer, resulting in a further increase in the storage modulus (to 570 Pa). The ability to tune the mechanical properties of the hydrogel at physiological conditions, and having the capacity to integrate any other aptamer sequence for the same effect (e.g. substituting the ATP input for a detectable, microbe-produced molecule), lends this technology well to the detection of microbial agents in complex serum samples.

Recently, a biomimetic hydrogel was described that had the ability to transform a small chemical signal input (adenosine) into a protein output (PDGF-BB) via a set of sequential displacement and hybridisation reactions under physiological conditions (Lai et al. 2017). The distinction between the approach used in this study and those published previously is that it does not rely on physical stimulations (i.e. light) or the use of exogenous cDNAs as molecular signal inputs. The polyethylene glycol-based hydrogel system consisted of two compartments, the core and the shell, for the displacement and hybridisation reactions, respectively. First, the rational design of an aptamer sequence (AS) capable of binding to the triggering molecule (TM) adenosine, and that was complementary to a triggering sequence (TS), was carried out. AS and TS were designed to form a stable duplex through hybridisation (within core compartment) that could be displaced by adenosine to result in the release of TS. The design of AS was based on a truncated adenosine aptamer, whilst TS was designed to hybridise to five nucleotides in the middle AS. Following displacement by adenosine, TS is able to move freely into the shell compartment. Here, a second aptamer sequence (AP) bound to the target protein is present. Following entry into the shell compartment, TS is able to hybridise completely with AP and thus release the bound target protein (PDGF-BB). To demonstrate the applicability of this elegant system, it was used to effect downstream signal transduction in smooth muscle cells (SMCs). Adenosine-triggered release of PDGF-BB was able to induce a calcium response in SMCs in a pattern that was consistent with the periodic stimulation of the hydrogel system by the signal input (adenosine). In principle, the modular design of this aptamer-functionalised hydrogel system enables the input to be changed to any metabolite (and its aptamer), for example, a molecule produced by a specific strain of bacteria, and the output to any signalling protein (and its aptamer), paving the way for its use in a microbial detection device, amongst other applications.

17.3.2 A Smart Outlook

In a biological sense, higher-order intelligence could be achieved through the incorporation of nucleic acid-based logic gates and sensors into hydrogels. These control systems are indeed the central tenet of SynBio, though their implementation in the building of smart hydrogels has yet to materialise. However, the recent and rapid progress that has been made in the techniques of cell-free TXTL could usher in a new era for smart biomaterials. As previously aforementioned, TXTL provides a means to unshackle the power of SynBio and enable its use in a wide range of settings without concern for the maintenance of 'live' GMOs and the associated risks of contamination. Furthermore, it aims to address another fundamental issue, that of stochastic variation of natural organisms, by cutting back the system to its most essential components. As such, this technology is the missing link that has held back the development of SynBio-based smart materials. It is with this in mind that the incorporation of TXTL genetic circuits capable of Boolean function into the highly versatile and useful hydrogel-based materials is anticipated. Conceptually, researchers can envision the ability to sense multiple microbe-borne inputs at differing localities and provide distinct and independent diagnostic responses, all within the same hydrogel. This orthogonal multi-sense and multi-response is not possible using conventional methods of functionalising hydrogels for smart capabilities. Moreover, recent advances in the field of artificial life, particularly the compartmentalised interactions of TXTL genetic circuit interactions within and between synthetic minimal cells (Adamala et al. 2017), are now readily available to be translated into material science. It is expected that, within the coming decade, the utilisation of smart biomaterials in the field of microbial detection will become ubiquitous and provide a platform for SynBio-based detection devices to be dispersed into the real world.

17.4 Synthetic Bacteriophage Detection

17.4.1 A Short History of Bacteriophages and Their Hosts

A bacteriophage, or phage, is a type of virus that infects bacteria. The relationship between a phage and its cognate bacterial prey constitutes the oldest predator-prey interaction on Earth, having existed for at least 1 billion years (Short et al. 2012). During this time, phages have evolved extreme specificity and sensitivity towards their hosts (Yosef et al. 2015). Today, phages are ubiquitous in all natural environments and represent the most abundant living entities on the Earth with an estimated 10^{32} bacteriophages in total (Kutter and Sulakvelidze 2005). In many cases it is the phage tail fibres, extending from their baseplate, that constitute the primary determinants of bacterial host specificity (Leiman and Shneider 2012). However, they can also recognise specific molecular motifs expressed on the bacterium surface, to which they bind prior to injecting their genetic material (Ahmed et al. 2014; Sorokulova et al. 2014).

Phages overlook all life except their specific bacterial hosts, upon which they are dependent for propagation on account of lacking any metabolic machinery (Chan and Abedon 2012). As a group, they can infect every known type of bacterium in every known environment on Earth (Santos et al. 2014; Abedon 2015; Pirnay et al. 2015; Catalao et al. 2013).

The bacteriophage offers a naturally occurring chassis that can, with the tools that SynBio offers, be modified and optimised for use in highly specific bacterial detection devices. Even prior to any modifications, they offer many unique benefits over traditional diagnostic approaches, including (a) high specificity and sensitivity towards their cognate host, (b) capacity to detect even traces of host presence, (c) capacity to function with impure samples under diverse or even harsh conditions, (d) discrimination between viable and incapacitated target pathogenic cells (i.e. removes false positive), (e) signal amplification capacity alongside signal transduction and (f) low-cost and easy propagation and purification (Schmelcher and Loessner 2014; Harada et al. 2018; Singh et al. 2013).

The ability to express and assemble phages from their isolated genome in a cell-free system (Rustad et al. 2017, 2018) allows for extensive control over when and how phages are deployed. This ground-breaking technique also gives researchers the ability to extensively analyse phage assembly and offers a unique premise for identifying genes completely necessary for assembly and other purposes through iterative gene knockouts. This knowledge is likely to supply ammunition for future phage-based detection systems and allows for an increased level of control over phages.

17.4.2 The Need for Rapid Detection of Specific Bacterial Species

Delays to the early, accurate diagnosis of many bacterial infections carry a significant personal and economic cost, both nosocomially and within the community. Hospital-acquired pathogenic bacterial infections are extremely common, with an annual occurrence of 4.1 M in Europe (Allegranzi 2011) and a mortality rate of 100,000 in the USA (Cimiotti et al. 2012). Aside from the emotional toll, these infections incur a huge financial burden estimated at €7.5B and $5B in Europe and the USA, respectively (Richter et al. 2018). The need for accurate diagnostics is huge and increasing, with the global in vitro diagnostics market estimated at $70B in 2018 and expected to reach $94B by 2023 with an enormous 5.6% CAGR (MDF 2018–2023). Furthermore, it is highly desirable within all affiliated fields to have the capability to differentiate between different bacterial infections so that specific and effective treatments can be administered.

17.4.3 Methods of Detection

Since phages are highly species-specific regarding their host, infection can elicit detection by lysis and/or expression of reporter genes. In either case, the positive

response is limited to the target host. This is a unique advantage. Detection via lysis exploits cytoplasmic enzymes that are present in the medium only upon lysis. Detection via reporter genes involves engineering exogenous reporter DNA into the phage genome and thus exclusively into the cells of a specified, target pathogen bacterial species. Nonlytic infection methods are limited to expression of phage-borne chromogenic, fluorescent or luminescent proteins.

Before considering these approaches in more detail, it should be noted that the unique ability of phage to discriminate live from dead cells can be exploited to reveal antibiotic resistance profiles of target host by comparing their detection responses with and without antibiotics (Rees and Barr 2017).

17.4.3.1 Phage-Borne Exogenous Reporter Genes

Phage genomes have been modified with several exogenous genes as a means of observing infection, thus confirming presence of the cognate host. Luciferase, through the engineering of *luxAB* genes, is perhaps the most common phage reporter system (Nguyen et al. 2017; Bhardwaj et al. 2017; He et al. 2017). Furthermore, a recent innovation is the development of the NanoLuc variant, which achieves far greater signal capacity from a smaller protein (England et al. 2016; Pulkkinen et al. 2019). Another similar system makes use of cytochrome c peroxidase (CCP). CCP catalyses the conversion of reduced cytochrome c, which is red, to oxidised cytochrome c, which is colourless. Integrating CCP into a phage genome therefore can confirm presence of cognate host through a loss of red colour (Hoang and Dien 2015). The methodology of this is that, once infection has begun, the recombinant phage will express CCP via its host's TXTL machinery and, upon the host being lysed by the phage progeny, the CCP will be exposed to its substrate in the extracellular medium.

Unsurprisingly, the ubiquitous reporter and marker protein GFP has found utility as a phage-based diagnostic tool for many bacterial species including *E. coli* K12 (Tanji et al. 2004), pathogenic O157:H7 (Oda et al. 2004) and *M. tuberculosis* (Piuri et al. 2009). Aside from its robustness and in-depth characterisation, the main advantage of GFP over other reporter gene systems is that it uses light instead of a substrate to generate signal, thereby making it a much simpler model. Moreover, it is also stable and well tolerated within living cells (Smartt et al. 2012).

In addition to the use of reporter genes carried by modified phages, it is possible to achieve detection by following increases in the concentration of phage genome via real-time quantitative PCR as recently shown for *Brucella abortus* samples in mixed culture and spiked blood (Sergueev et al. 2017).

17.4.3.2 Detecting Host Enzymes in Phage-Induced Lysate

Another diagnostic strategy relies upon detection of enzymes that are present in the extracellular medium only after lysis. The presence of such enzymes therefore can act as indication of phage activity which in turn indicates presence of the cognate host. Several enzymatic phage-based systems have been previously exhibited, including alkaline phosphatase (Alcaine et al. 2015), ATP + luciferase (Schmelcher and Loessner 2014; He et al. 2017) and β-glycosidase (Hagens et al. 2011).

Either the enzyme *or* substrate is present only within bacterial cells; therefore, its extracellular presence indicates cell lysis due to phage activity, allowing these engineered phages to act as a bacterial diagnostic.

Lastly, although not evident in the literature, it is entirely conceivable that one of the proteins released into the extracellular medium as a result of phage-induced lysis could be utilised as an input for a genetic circuit (also present in the extracellular medium), which subsequently produces a detectable output, thereby combining the bacteriophage and circuitry approaches for a novel microbial detection system.

17.4.4 The Role of CRISPR in the Next Generation of Phage-Mediated Bacterial Detection

The CRISPR/Cas system is based on an RNA-directed DNA endonuclease mechanism (Sampson and Weiss 2014) that allows bacteria to adaptively immunise themselves against harmful, invading nucleic acids, for example, from phages or plasmids (Sorek et al. 2008). This is achieved by storing—on the host chromosome—sequences corresponding to past encounters with harmful foreign DNA (Shmakov et al. 2017). These 'signatures', called spacers, are then used by Cas proteins to detect and destroy any DNA or RNA in the cell that harbours the same sequences (Bolotin et al. 2005).

In this way, spacers act as memory—that can be updated—hence the immunity is acquired and not merely innate (Barrangou et al. 2007; Horvath et al. 2008). Spacers can target any nucleic acid sequence and thus can even cause bacterial autoimmunity (Stern et al. 2010)—although in some cases mechanisms exist to prevent self-digestion (Goldberg et al. 2014).

17.4.4.1 Mechanism of Action

Functionally, CRISPR/Cas systems are composed of chromosomal CRISPR DNA sequences and CRISPR-associated (Cas) proteins (Karimi et al. 2018). CRISPR DNA sequences carry distinctive, repeating segments 23–55 base pairs long that are regularly positioned, interspersed with non-repeating sequences called spacers that are 26–72 base pairs long (Pourcel et al. 2005). Spacers are of non-host origin and carry the memory of what sequences to destroy (Barrangou et al. 2007). The repeating parts often contain palindromic sequences and are postulated to form stem and loop structures when transcribed (Deveu et al. 2010). The number of spacers in a host can be extremely variable, but no host carries more than one copy of a given spacer (Lillestol et al. 2006). Flanking these interspersed sequences, there is also always a leader sequence at the 5′ end and sometimes at the 3′ too. Leader sequences carry the promoters for transcription of the whole interspersed sequence as a single block (Pul et al. 2010) that is later spliced to remove most of each repeated motif, leaving small RNAs corresponding to the spacer sequences flanked by part of the repeat sequence (Lillestol et al. 2009).

Cas genes occur only in CRISPR-competent hosts and always lie adjacent to CRISPR sequences (Horvath and Barrangou 2010). They serve various functions,

including adding new spacer sequences (Barrangou 2013), cutting DNA matching spacer RNA fragments (Barrangou and Marraffini 2014) and even editing target DNA (Xiao-Jie et al. 2015).

17.4.4.2 Phage Editing by CRISPR

The use of CRISPR for phage engineering has been vastly developed over the past few years, and various combinations of point mutation, gene insertion, deletion and sway have been demonstrated for phages targeting various species including *E. coli* (Tao et al. 2017), *K. pneumoniae* (Shen et al. 2018) and *S. thermophilus* (Martel and Moineau 2014). Tao et al. further confirmed that glycosylation and hydroxylmethylation of cytosine, as preferred by phage T4, does not prevent the action of the CRISPR/Cas9 system. Their approach allowed for the efficient and rapid sequential inactivation of individual genes—a very useful capability given that approximately half of T4's genes have yet to be characterised (Miller et al. 2003; Mohanraju et al. 2016).

Weak guide RNAs can be used for precise genome engineering that—crucially—do not act on random sites (Shen et al. 2018). This greatly facilitates the rapid selection of desired edits and, coupled with the observation that CRISPR-based selection of targeted changes was successful in all tested plaques (Martel and Moineau 2014), demonstrates the shear potency of CRISPR. Martel and Moineau further demonstrated that plasmid-borne spacers function with the same efficiency as their natural chromosomal counterparts. Thus, by placing spacers onto phage genomes under late promoters, the possibility of viability being restored to engineered diagnostic phages by recombination is reduced since the phage DNA will be digested.

The use of plasmid-borne spacers to target phage-borne sequences was applied in an alternative approach that used CRISPR in conjunction with homologous recombination (Kiro et al. 2014). In this approach, CRISPR is directed to degrade non-edited genomes, thereby hugely expediting retrieval of successful recombinants. Further to that, in a recent paper, CRISPR was used to select genetically modified K1F phages, which had inserted a sfGFP gene on their minor protein, over wild-type phages (Moller-Olsen et al. 2018).

17.4.4.3 CRISPR's Potential to Enhance Phage-Mediated Bacterial Detection

Indirect phage-mediated detection of bacterial species has been partially demonstrated. Phage-acquired bacterial spacer sequences can aid in the diagnosis of potentially dangerous strains (Ranjbar et al. 2014) by amplifying conserved regions via a technique later called 'spoligotyping' (Mokrousov et al. 2007). However, this approach is limited by the observation that strains can rapidly obtain new spacers (Mohammed 2017) and thus cannot always be correctly detected.

Still, the utility of CRISPR for direct phage-based detection of bacteria is still in its infancy. Since any sequence can be introduced into a bacterium as a spacer sequence, it is possible to target the host's own chromosome (Stern et al. 2010). By engineering phages to coincide with or avoid spacer sequences, it may be possible

to type strains based on their response. This method could allow for the rapid, phage-mediated lysis and thus downstream detection as aforementioned. It should also be possible to overcome current host immunity to phage by modifying phage protospacer sequences. In this way, this immunity can be reversed and these strains can be rendered detectable by phages. Furthermore, engineering phages to be both replication-defective and self-digesting can preclude the release of viable modified progeny.

As can be seen, the necessary tools exist to allow for the rapid detection of bacterial species using CRISPR-modified phages. Such endeavours are all but certain to yield fruit in the coming decade.

17.5 Conclusions

Accessible SynBio-enhanced diagnostic systems within the field of microbiology are imminent; and as this chapter has revealed, there are multiple avenues to explore and exploit in order to achieve rapid and accurate microbial detection. There are still significant milestones to be met before these types of technologies will be routinely used in general practice. However, with the extensive development and use of versatile tools such as cell-free TXTL—which considerably accelerate the design-build-test cycle for the efficient execution of gene circuit and bacteriophage detection systems—the applications this field can offer are forthcoming.

Automation will be a key process within the field in the coming years. Computational design of microbial detection systems will allow for the seamless screening of large, complex data sets to identify the most accurate and efficient methods of detection. Furthermore, two recent software developments, Cello and SensiPath, have offered tools that aid in automating the genetic circuit design process (Nielsen et al. 2016; Delepine et al. 2016). These ground-breaking technologies, and many like them, are helping to evolve conventional biologists into a new breed of hybrid biological engineers. With the fundamental engineering principles underlying an ever-increasing proportion of biological research, the immaculate level of control seen in computational science can be emulated in synthetic biological systems. These advancements lend well to the field of microbial detection; moreover, the progression and/or collaboration between the fields of gene circuits, synthetic phages and smart biomaterials will undisputedly yield high-performance microbial detection systems in the coming years.

References

Abedon ST (2015) Phage therapy of pulmonary infections. Bacteriophage 5:e1020260

Adamala KP, Martin-Alarcon DA, Guthrie-Honea KR, Boyden ES (2017) Engineering genetic circuit interactions within and between synthetic minimal cells. Nat Chem 9:431–439

Ahmed A, Rushworth JV, Hirst NA, Millner PA (2014) Biosensors for whole-cell bacterial detection. Clin Microbiol Rev 27:631–646

Alcaine SD, Pacitto D, Sela DA, Nugen SR (2015) Phage & phosphatase: a novel phage-based probe for rapid, multi-platform detection of bacteria. Analyst 140:7629–7636

Alcaine SD, Law K, Ho S, Kinchla AJ, Sela DA, Nugen SR (2016) Bioengineering bacteriophages to enhance the sensitivity of phage amplification-based paper fluidic detection of bacteria. Biosens Bioelectron 82:14–19

Allegranzi B (2011) Report on the burden of endemic health care-associated infection worldwide. World Health Organisation, Geneva

Barrangou R (2013) CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA 4:267–278

Barrangou R, Marraffini L (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 54:234–244

Barrangou R, Fremaux C, Deveu H, Romero D, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712

Bhardwaj N, Bhardwaj SK, Mehta J, Kim KH, Deep A (2017) MOF-bacteriophage biosensor for highly sensitive and specific detection of *Staphylococcus aureus*. ACS Appl Mater Interfaces 9:33589–33598

Bhunia A (2014) One day to one hour- how quickly can foodborne pathogens be detected? Future Microbiol 9(8):935–946

Bolotin A, Quinquis B, Sorokin A, Ehrlich S (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151:2551–2561

Bundy BC, Hunt JP, Jewett MC, Swartz JR, Wood DW, Frey DD, Rao G (2018) Cell-free biomanufacturing. Curr Opin Chem Eng 22:177–183

Catalao MJ, Gil F, Moniz-Pereira J, Sao-Jose C, Pimentel M (2013) Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev 37:554–571

Chan B, Abedon S (2012) Phage therapy pharmacology: phage cocktails. Adv Appl Microbiol 78:1–23

Chen J, Alcaine SD, Jackson AA, Rotello VM, Nugen SR (2017a) Development of engineered bacteriophages for Escherichia coli detection and high-throughput antibiotic resistance determination. ACS Sens 2:484–489

Chen J, Picard RA, Wang D, Nugen SR (2017b) Lyophilized engineered phages for Escherichia colidetection in food matrices. ACS Sens 2:1573–1577

Cimiotti JP, Aiken LH, Sloane DM, Wu ES (2012) Nurse staffing, burnout, and health care-associated infection. Am J Infect Control 40:486–490

Delepine B, Libis V, Carbonell P, Faulon JL (2016) SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Res 44:W226–W231

Deveu H, Garneau J, Moineau S (2010) CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64:475–493

Didovyk A, Tonooka T, Tsimring L, Hasty J (2017) Rapid and scalable preparation of bacterial lysates for cell-free gene expression. ACS Synth Biol 6:2198–2208

Dudley QM, Karim AS, Jewett MC (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J 10:69–82

England CG, Ehlerding EB, Cai W (2016) NanoLuc: a small luciferase is brightening up the field of bioluminescence. Bioconjug Chem 27:1175–1187

Garamella J, Marshall R, Rustad M, Noireaux V (2016) The all E. coli TX-TL toolbox 2.0: a platform for cell-free synthetic biology. ACS Synth Biol 5:344–355

Garcia-Ojalvo J, Borek B, Hasty J, Tsimring L (2016) Turing patterning using genecircuits with gas-induced degradation of quorum sensing molecules. PLoS One 11:e0153679

Goldberg G, Jiang W, Bikard D, Marraffini L (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature 514:633–637

Grawe A, Dreyer A, Vornholt T, Barteczko U, Buchholz L, Drews G, Ho UL, Jackowski ME, Kracht M, Luders J, Bleckwehl T, Rositzka L, Ruwe M, Wittchen M, Lutter P, Muller K, Kalinowski J (2019) A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. PLoS One 14:e0210940

Hagens S, Loessner MJ (2007) Application of bacteriophages for detection and control of foodborne pathogens. Appl Microbiol Biotechnol 76:513–519

Hagens S, De Wouters T, Vollenweider P, Loessner MJ (2011) Reporter bacteriophage A511:celB transduces a hyperthermostable glycosidase from Pyrococcus furiosus for rapid and simple detection of viable Listeria cells. Bacteriophage 1:143–151

Harada LK, Silva EC, Campos WF, Del Fiol FS, Vila M, Dąbrowska K, Krylov VN, Balcão VM (2018) Biotechnological applications of bacteriophages: state of the art. Microbiol Res 212:38–58

He Y, Wang M, Fan E, Ouyang H, Yue H, Su X, Liao G, Wang L, Lu S, Fu Z (2017) Highly specific bacteriophage-affinity strategy for rapid separation and sensitive detection of viable Pseudomonas aeruginosa. Anal Chem 89:1916–1921

Hoang H, Dien T (2015) Rapid and simple colorimetric detection of Escherichia coli O157-H7 in apple juice using a novel recombinant bacteriophage-based method. Biocontrol Sci 20:99–103

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167–170

Horvath P, Romero D, Richards M, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 190(4):1401–1412

Karimi Z, Ahmadi A, Najafi A, Ranjbar R (2018) Bacterial CRISPR regions: general features and their potential for epidemiological molecular typing studies. Open Microbiol J 12:59–70

Kim SH, Lee SH, Lee JE, Park SJ, Kim K, Kim IS, Lee YS, Hwang NS, Kim BG (2018) Tissue adhesive, rapid forming, and sprayable ECM hydrogel via recombinant tyrosinase crosslinking. Biomaterials 178:401–412

Kiro R, Shitrit D, Qimron U (2014) Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biol 11:42–44

Kutter E, Sulakvelidze A (2005) Bacteriophages - biology and applications. CRC Press, Boca Raton

Lai J, Li S, Shi X, Coyne J, Zhao N, Dong F, Mao Y, Wang Y (2017) Displacement and hybridization reactions in aptamer-functionalized hydrogels for biomimetic protein release and signal transduction. Chem Sci 8:7306–7311

Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. Adv Exp Med Biol 726:93–114

Li C, Chen P, Shao Y, Zhou X, Wu Y, Yang Z, Li Z, Weil T, Liu D (2015) A writable polypeptide-DNA hydrogel with rationally designed multi-modification sites. Small 11:1138–1143

Li Q, Wu Y, Lu H, Wu X, Chen S, Song N, Yang YW, Gao H (2017) Construction of supramolecularnanoassembly for responsive bacterial elimination and effective bacterial detection. ACS Appl Mater Interfaces 9:10180–10189

Lillestol R, Redder P, Garret R, Brugger K (2006) A putative viral defence mechanism in archaeal cells. Archaea 2:59–72

Lillestol R, Shah S, Brugger K, Redder P, Phan H, Christiansen J, Garret R (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol 72:259–272

Liu H, Cao T, Xu Y, Dong Y, Liu D (2018) Tuning the mechanical properties of a DNA hydrogel in three phases based on ATP aptamer. Int J Mol Sci 19:E1633

Ma D, Shen L, Wu K, Diehnelt CW, Green AA (2018) Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. Synth Biol (Oxf) 3:ysy018

Martel B, Moineau S (2014) CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res 42:9504–9513

MDF(2018–2023). In-vitro diagnostics market by product (instruments, reagents and kits, data management system/software and services), by technology (clinical chemistry, hematology, immunoassays, coagulation and hemostasis, molecular diagnostics, microbiology and other IVD technologies), by application (diabetes, infectious diseases, oncology/cancer, cardiology, nephrology, autoimmune diseases, drug testing and HIV) and by region - global industry analysis, size, share, growth, trends, and forecasts (2018-2023). Market Data Forecast

Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W (2003) Bacteriophage T4 genome. Microbiol Mol Biol Rev 67:86–156

Mohammed M (2017) Phage typing or CRISPR typing for epidemiological surveillance of Salmonella typhimurium? BMC Res Notes 10:578

Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, Van Der Oost J (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353:aad5147

Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O (2007) Corynebacterium diphtheriae spoligotyping based on combined use of two CRISPR loci. Biotechnol J 2:901–906

Moller-Olsen C, Ho SFS, Shukla RD, Feher T, Sagona AP (2018) Engineered K1F bacteriophages kill intracellular Escherichia coli K1 in human epithelial cells. Sci Rep 8:17559

Monod J, Jacob F (1961) General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol 26:389–401

Ng PP, Jia M, Patel KG, Brody JD, Swartz JR, Levy S, Levy R (2012) A vaccine directed to B cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. Proc Natl Acad Sci U S A 109:14526–14531

Nguyen C, Makkar R, Sharp NJ, Page MA, Molineux IJ, Schofield DA (2017) Detection of Bacillus anthracis spores from environmental water using bioluminescent reporter phage. J Appl Microbiol 123:1184–1193

Nielsen AA, Der BS SJ, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA (2016) Genetic circuit design automation. Science 352:aac7341

O'neill J (2014) Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Rev Antimicrob Resist 20:1–6

Oda M, Morita M, Unno H, Tanji Y (2004) Rapid detection of Escherichia coli O157:H7 by using green fluorescent protein-labeled PP01 bacteriophage. Appl Environ Microbiol 70:527–534

Ogonah OW, Polizzi KM, Bracewell DG (2017) Cell free protein synthesis: a viable option for stratified medicines manufacturing? Curr Opin Chem Eng 18:77–83

Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, Ferrante T, Ma D, Donghia N, Fan M, Daringer NM, Bosch I, Dudley DM, O'connor DH, Gehrke L, Collins JJ (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165:1255–1266

Pedrolli DB, Ribeiro NV, Squizato PN, De Jesus VN, Cozetto DA, Team AQAUAI (2019) Engineering microbial living therapeutics: the synthetic biology toolbox. Trends Biotechnol 37:100–115

Pirnay JP, Blasdel BG, Bretaudeau L, Buckling A, Chanishvili N, Clark JR, Corte-Real S, Debarbieux L, Dublanchet A, De Vos D, Gabard J, Garcia M, Goderdzishvili M, Gorski A, Hardcastle J, Huys I, Kutter E, Lavigne R, Merabishvili M, Olchawa E, Parikka KJ, Patey O, Pouilot F, Resch G, Rohde C, Scheres J, Skurnik M, Vaneechoutte M, Van Parys L, Verbeken G, Zizi M, Van Den Eede G (2015) Quality and safety requirements for sustainable phage therapy products. Pharm Res 32:2173–2179

Piuri M, Jacobs WR Jr, Hatfull GF (2009) Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of Mycobacterium tuberculosis. PLoS One 4:e4870

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151:653–663

Prindle A, Samayoa P, Razinkov I, Danino T, Tsimring LS, Hasty J (2011) A sensing array of radically coupled genetic 'biopixels'. Nature 481:39–44

Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol 75:1495–1512

Pulkkinen EM, Hinkley TC, Nugen SR (2019) Utilizing in vitro DNA assembly to engineer a synthetic T7 Nanoluc reporter phage for Escherichia coli detection. Integr Biol (Camb) 11:63

Rampioni G, D'angelo F, Messina M, Zennaro A, Kuruma Y, Tofani D, Leoni L, Stano P (2018) Synthetic cells produce a quorum sensing chemical signal perceived by Pseudomonas aeruginosa. Chem Commun (Camb) 54:2090–2093

Ranjbar R, Karami A, Farshad S, Giammanco GM, Mammina C (2014) Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. New Microbiol 37:1–15
Rappuoli R, Bloom DE, Black S (2017) Deploy vaccines to fight superbugs. Nature 552:165–167
Rees J, Barr J (2017) Detection of methicillin-resistant Staphylococcus aureus using phage amplification combined with matrix-assisted laser desorption/ionization mass spectrometry. Anal Bioanal Chem 409:1379–1386
Richter L, Janczuk-Richter M, Niedziolka-Jonsson J, Paczesny J, Holyst R (2018) Recent advances in bacteriophage-based methods for bacteria detection. Drug Discov Today 23:448–455
Riglar DT, Silver PA (2018) Engineering bacteria for diagnostic and therapeutic applications. Nat Rev Microbiol 16:214–225
Rustad M, Eastlund A, Marshall R, Jardine P, Noireaux V (2017) Synthesis of Infectious Bacteriophages in an E. coli-based Cell-free Expression System. J Vis Exp 126:e56144
Rustad M, Eastlund A, Jardine P, Noireaux V (2018) Cell-free TXTL synthesis of infectious bacteriophage T4 in a single test tube reaction. Synth Biol 3:ysy002
Sackett SD, Tremmel DM, Ma F, Feeney AK, Maguire RM, Brown ME, Zhou Y, Li X, O'brien C, Li L, Burlingham WJ, Odorico JS (2018) Extracellular matrix scaffold and hydrogel derived from decellularized and delipidized human pancreas. Sci Rep 8:10452
Sampson T, Weiss D (2014) Exploiting CRISPR/Cas systems for biotechnology. BioEssays 36:34
Santos SB, Carvalho C, Azeredo J, Ferreira EC (2014) Population dynamics of a Salmonella lytic phage and its host: implications of the host bacterial growth rate in modelling. PLoS One 9:e102507
Sawlekar R, Montefusco F, Kulkarni VV, Bates DG (2016) Implementing nonlinear feedback controllers using DNA strand displacement reactions. IEEE Trans Nanobioscience 15:443–454
Schmelcher M, Loessner MJ (2014) Application of bacteriophages for detection of foodborne pathogens. Bacteriophage 4:e28137
Sedlmayer F, Hell D, Muller M, Auslander D, Fussenegger M (2018) Designer cells programming quorum-sensing interference with microbes. Nat Commun 9:1822
Sergueev KV, Filippov AA, Nikolich MP (2017) Highly sensitive bacteriophage-based detection of Brucella abortus in mixed culture and spiked blood. Viruses 9:E144
Sharp NJ, Molineux IJ, Page MA, Schofield DA (2016) Rapid detection of viable *Bacillus anthracis* spores in environmental samples by using engineered reporter phages. Appl Environ Microbiol 82:2380–2387
Shen J, Zhou J, Chen G, Xiu Z (2018) Efficient genome engineering of a virulent Klebsiella bacteriophage using CRISPR-Cas9. J Virol 92:e00534-18
Sherwood AV, Henkin TM (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. Annu Rev Microbiol 70:361–374
Shin J, Jardine P, Noireaux V (2012) Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth Biol 1:408–413
Shmakov S, Sitnik V, Makarova K, Wolf Y, Severinov K, Koonin E (2017) The CRISPR spacer space is dominated by sequences from species-specific mobilomes. MBio 8:e01397-17
Short FL, Blower TR, Salmond GP (2012) A promiscuous antitoxin of bacteriophage T4 ensures successful viral replication. Mol Microbiol 83:665–668
Siegal-Gaskins D, Tuza ZA, Kim J, Noireaux V, Murray RM (2014) Gene circuit performance characterization and resource usage in a cell-free "breadboard". ACS Synth Biol 3:416–425
Singh A, Poshtiban S, Evoy S (2013) Recent advances in bacteriophage based biosensors for food-borne pathogen detection. Sensors (Basel) 13:1763–1786
Smartt A, Xu T, Jegier P, Carswell J, Blount S, Sayler G, Ripp S (2012) Pathogen detection using engineered bacteriophages. Anal Bioanal Chem 402:3127–3146
Sorek R, Kunin V, Hugenholtz P (2008) CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol 6:181–186
Sorokulova I, Olsen E, Vodyanoy V (2014) Bacteriophage biosensors for antibiotic-resistant bacteria. Expert Rev Med Devices 11:175–186
Stern A, Keren L, Wurtzel O, Amitai G, Sorek R (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet 26:335–340

Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V (2013) Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J Vis Exp 79:e50762

Takahashi MK, Chappell J, Hayes CA, Sun ZZ, Kim J, Singhal V, Spring KJ, Al-Khabouri S, Fall CP, Noireaux V, Murray RM, Lucks JB (2015) Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. ACS Synth Biol 4:503–515

Takahashi MK, Tan X, Dy AJ, Braff D, Akana RT, Furuta Y, Donghia N, Ananthakrishnan A, Collins JJ (2018) A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. Nat Commun 9:3347

Tanji Y, Furukawa C, Na S, Hijikata T, Miyanaga K, Unno H (2004) Escherichia coli detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. J Biotechnol 114:11–20

Tao P, Wu X, Tang WC, Zhu J, Rao V (2017) Engineering of bacteriophage T4 genome using CRISPR-Cas9. ACS Synth Biol 6:1952–1961

Turturro MV, Sokic S, Larson JC, Papavasiliou G (2013) Effective tuning of ligand incorporation and mechanical properties in visible light photopolymerized poly(ethylene glycol) diacrylate hydrogels dictates cell adhesion and proliferation. Biomed Mater 8:025001

Vogel V (2018) Unraveling the mechanobiology of extracellular matrix. Annu Rev Physiol 80:353–387

Wang D, Chen J, Nugen SR (2017a) Electrochemical detection of Escherichia coli from aqueous samples using engineered phages. Anal Chem 89:1650–1657

Wang J, Leong MC, Leong EZW, Kuan WS, Leong DT (2017b) Clinically relevant detection of Streptococcus pneumoniae with DNA-antibody nanostructures. Anal Chem 89:6900–6906

Whiteley M, Diggle SP, Greenberg EP (2017) Progress in and promise of bacterial quorum sensing research. Nature 551:313–320

Xiao-Jie L, Hui-Ying X, Zun-Ping K, Jin-Lian C, Li-Juan J (2015) CRISPR-Cas9: a new and promising player in gene therapy. J Med Genet 52:289–296

Yosef I, Manor M, Kiro R, Qimron U (2015) Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. ProcNatlAcadSci U S A 112:7267–7272

You C, Zhang YH (2013) Cell-free biosystems for biomanufacturing. Adv Biochem Eng Biotechnol 131:89–119

Zhang C, Ge L, Zhuang Y, Shen Z, Zhong Z, Zhang Z, You X (2019) DNA computing for combinational logic. Sci China Inf Sci 62(6):61301

Zhou X, Li C, Shao Y, Chen C, Yang Z, Liu D (2016) Reversibly tuning the mechanical properties of a DNA hydrogel by a DNA nanomotor. ChemCommun (Camb) 52:10668–10671

Application and Challenges of Synthetic Biology

18

Zandile Nxumalo and Deepak B. Thimiri Govinda Raj

Abstract

Synthetic biology is an emerging technology that uses the principles of physics, chemistry, mathematics, and biology. Synthetic biology is the technology enabler that could be applied across all fields such as agricultural, food, health, chemical, and industrial manufacturing sectors. In this chapter, several applications of synthetic biology across all disciplines are delineated, and challenges associated with the implementation of synthetic biology technologies are addressed in detail.

Keywords

Bioengineered bugs · UN sustainable developmental goals · Synthetic biology · CRISPR and genetically modified organisms (GMO)

18.1 Introduction

Synthetic biology is an emerging field that combines principles of engineering, mathematics, physics, chemistry, and biology (Cheng and Lu 2012). As it is an enabler and interdisciplinary, synthetic biology has been applied in several fields with a diverse range of applications. Some of the applications include use of synthetic microbes for probiotics, gene therapy, and vaccine production (Claesen and Fischbach 2015; Álvarez and Fernández 2017; Jain et al. 2012). Several countries currently implement synthetic biology research in their research centers and universities in order to harness synthetic biology use in various fields (Si and Zhao 2016). In addition, the European Union (EU) uses synthetic biology

Z. Nxumalo · D. B. Thimiri Govinda Raj (✉)
Synthetic Biology Center, NextGeneration Health, CSIR, Pretoria, South Africa
e-mail: dgovindaraj@csir.co.za

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_18

technologies to achieve the 2030 UN sustainable development goals (UN SDGs) such as UN SDG 1, no poverty; UN SDG 3, good health and well-being; UN SDG 8, decent work and economic growth; and UN SDG 9, industry, innovation, and infrastructure (https://sustainabledevelopment.un.org/partnership/register/?source=90). For example, synthetic biology could potentially be applied for biomanufacturing, bioremediation, and pollution control. To achieve the UN sustainable development goals, it would be essential to use all the applications of synthetic biology and biotechnology across all continents. Another critical factor that governs the use of synthetic biology for UN SDGs requires global acceptance for this enabling technology (https://www.un.org/sustainabledevelopment/).

In this review, we detailed the list of current and future applications of synthetic biology across all the fields. We also elaborated various synthetic biology technologies used in universities, hospitals, industry, agriculture, and chemical industries. In addition, in this article, we decipher the list of current and potential future challenges for synthetic biology. We particularly identified the potential challenges for synthetic biology to be implemented in order to achieve UN sustainable development goals. We also addressed the importance of political, geographical, and economic aspects that also govern applications and challenges of synthetic biology. In this article, we also identified potential approaches to tackle current and future challenges of synthetic biology.

18.2 State-of-the-Art and Recent Developments in Synthetic Biology

A most common definition for synthetic biology is that synthetic biology is an emerging research field that involves design and construction of new bio-parts, bio-devices, and bio-systems along with redesigning of existing bio-parts for superior applications (EU report on synthetic biology; see https://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_044.pdf). Although this definition is broad enough to include biotechnology and use of living systems in all fields, synthetic biology can be differentiated from genetic engineering, metabolic engineering, and classical biotechnology. Synthetic biology is built on the capacities of in silico design-based genome engineering and metabolic engineering that involve concepts of bio-parts standardization and synthetic biology abstraction (Kelwick et al. 2014). Most of the important technologies that governed recent advancements in synthetic biology and its capacity to be applied in several fields have the ability to artificially synthesize biological parts such as DNA, RNA, protein, and other bio-parts. Some of the important technological advancements in genetic engineering-driven synthetic biology include synthesis and assembly of virus (Thimiri Govinda Raj et al. 2014), bacteria (Gibson et al. 2008), and yeast genome (Synthetic Yeast Genome Project (Sc2.0) syntheticyeast.org/). However, synthetic biology by itself is diverse and is applied in generating both novel bio-

logical parts and a design that can perform customized application. Thus, synthetic biology is divided into two main classes such as top-down synthetic biology and bottom-up synthetic biology. Bottom-up approaches include de novo synthesis of minimal and artificial life, while top-down approaches include biological design systems generated based on existing knowledge in order to address a specific task (Roberts et al. 2013). Based on its applications, synthetic biology is also classified as medical and industrial synthetic biology.

If we could map the history of synthetic biology, we could perhaps better understand the fundamental principles governing the synthetic biology field. Development of synthetic biology started as early as 1953, when DNA double helix followed by the genetic code was elucidated (Liang et al. 2011). Advancement of whole-genome sequencing with the applications of next-generation sequencing methods initiated large-scale sequencing projects at the J. Craig Venter Institute (JCVI). By combining the molecular biology, cell biology, and chemistry techniques, JCVI scientists recreated synthetic genome of the *Mycoplasma genitalium* bacteria as a template (Gibson et al. 2008). This work confirmed the capacity of synthetic biology to design and experimentally produce synthetic cells. This work propelled synthetic biology field to the next level, where JCVI scientists could work at whole-genome scale and build a living *Mycoplasma mycoides* JCVI-syn1.0 cell harboring a completely synthetic chromosome (Gibson et al. 2010). JCVI scientists successfully demonstrated that the synthetic genome replicated like the native genome found in the wild-type *M. genitalium* or *M. mycoides*.

In 2016, JCVI scientists further produced synthetic genome with only essential genes ($n = 91$) of *M. mycoides* JCVI-syn3.0, which harbors the smallest genome of any free-living organism (Hutchison et al. 2016), (Danchin and Fang 2016). The most recent advancement in synthetic biology field is the use of artificial intelligence and machine learning for both genome-scale modeling and metabolic modeling of synthetic bio-systems (Breuer et al. 2019). In this recent work, the JCVI team reconstructed the first-of-its-kind computational or "in silico" model of a comprehensive set of metabolic reactions that could take place in the minimal cell. Another recent development is the application of synthetic biology in minimal biology. Minimal biology is a new and nascent field that uses the technologies in systems and synthetic biology toward finding the least number of components needed to create living minimal cells (Xavier et al. 2014). Some of the concepts related to minimal biology include minimal genome, LUCA, chassis cell, artificial or semi-artificial cell, and minimal cell models (protocells, in silico cell, and minimal cells). In this article, we also review these concepts and synthetic biology applications (Figs. 18.1 and 18.2).

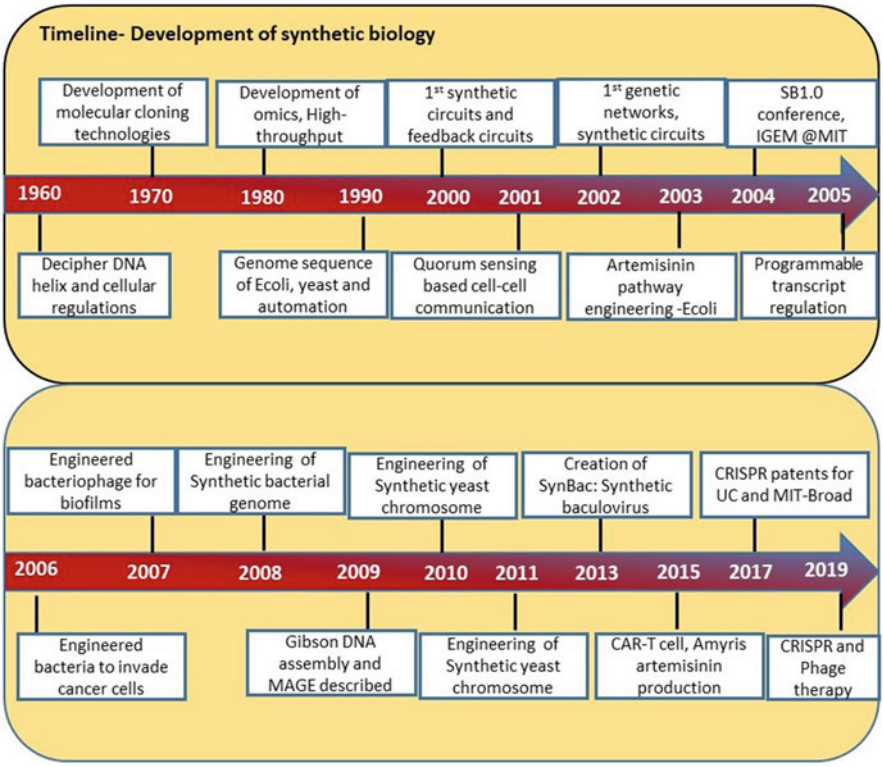


Fig. 18.1 History of synthetic biology

18.3 Applications of Synthetic Biology in Medicine

In this review, we detail three applications of synthetic biology in medicine that includes (a) phage-based therapy, (b) CAR T-cell therapy, and (c) CRISPR/Cas9 therapy (Table 18.1) (Hong 2018).

Phage-based therapy: Bacteriophage or viral phage therapy is the treatment used to treat pathogenic bacterial infections.

CAR T-cell therapy: Chimeric antigen receptor T-cell therapy involves the genetic engineering of patient's immune T cells to express a gene that codes for a specific receptor that binds to a protein of cancerous cells (Zhang et al. 2017).

CRISPR/Cas9 therapy: CRISPR is short for clusters of regularly interspaced short palindromic repeats, and CRISPR-associated protein 9 encodes for two key molecules that allow for a change/mutation into the strands of DNA, at a specific location in the genome so that bases of DNA can be added or removed (Jiang and Doudna 2017).

Although research is ongoing in the application of these synthetic biology techniques, researchers involved in clinical studies are collecting clinical samples

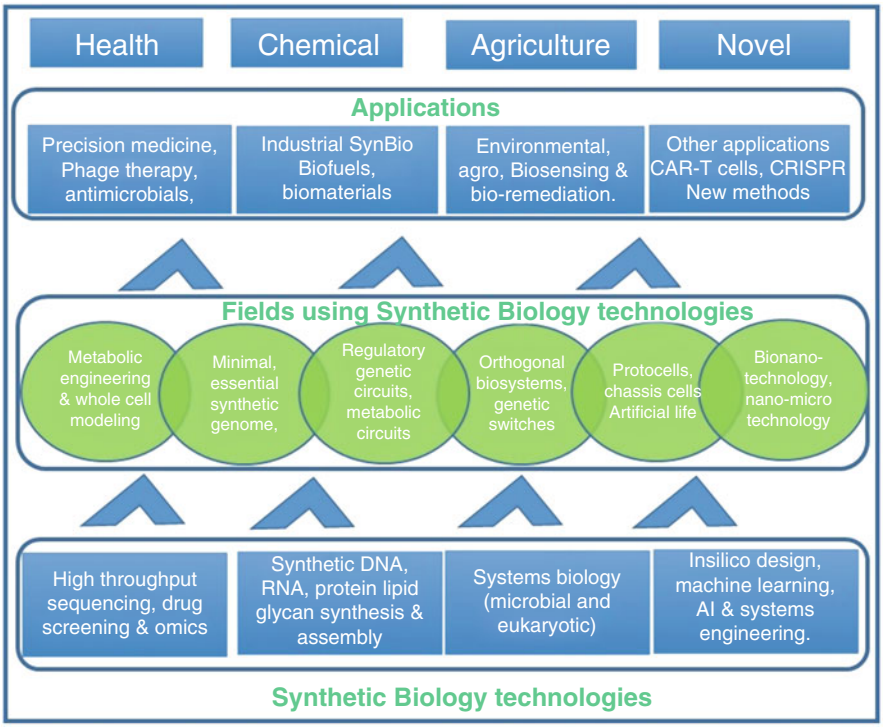


Fig. 18.2 Applications of synthetic biology across all disciplines

and information from patients to genetically modify T cells using CRISPR so that they can engineer cells that when transplanted back into host are unable to cause graft versus host complications. Unlike CAR T-cell therapy, which has a number of FDA-approved therapeutics for cancer, CRISPR/Cas9 therapy promises to have different targets and indications with a number of approvals upon successful evaluation of the data generated from the clinical trial (Zhen and Li 2019).

18.4 Applications of Synthetic Biology in Agriculture

The biggest problem that the agricultural sector is facing in the twenty-first century is the unsustainable need due to rapid population growth and which ultimately has a negative impact on global society. Recent advancements in the synthetic biology field, particularly, new techniques in genome design, such as CRISPR/Cas9, harbor opportunities to engineer microbial metabolism; novel anti-microbials and more recently the Yeast 2.0 project have a significant potential to deliver short- and long-term transformative changes to the agricultural sector (Goold et al. 2018). Synthetic biology can also help in the preservation of our natural resources, improvement

Table 18.1 Applications of synthetic biology in medicine

Synthetic biology	Mechanism	Application
Phage therapy	Unlike antibiotics, phages are specifically engineered in the lab to target and consume bacteria through an antagonistic arm mechanism that counter-adapts against phage-resistant bacteria, a process known coevolution (Azam and Tanji 2019)	Phage therapy in the past was used to treat wound infections and other associated wound infection complications after surgery (Morozova et al. 2018). A classical example of phage therapy in post-operative wounds treatment is observed in cancer patients. More recently with advances in synthetic biology methods, the application for phage therapy has been developed as an alternative therapy to antibiotics for the treatment of multiresistant bacterial infections (Furfaro et al. 2018). However, it is recommended to include antibiotics together with bacteriophage therapy. The proposed phase 2 clinical trial is evaluating the safety, tolerability, and efficacy of a combination of antibiotics with AB-SA01 bacteriophage therapy in participants with a *Staphylococcus aureus* (*S. aureus*) ventricular assist device (VAD) infection (Azam and Tanji 2019)
CAR T-cell therapy	In the human body, T cells normally function by detecting and destroying viruses and other pathogens (Zhang et al. 2017). Similarly, CAR T-cell therapy functions by selectively targeting malignant cells because of genetic engineering of autologous immune T cells (Ding 2018)	Although CAR T-cell therapy is experimental, the Food and Drug Administration (FDA) has approved tisagenlecleucel and axicabtagene for the treatment of high-grade B-cell lymphoma and diffuse B-cell leukemia, respectively (National Cancer Institute 2019). CAR T-cell therapy's success in enhancing the immune system has resulted in significant result, such that other applications are currently in development for the treatment of viral infections, like HIV (Patel et al. 2016)
CRISPR/Cas9	CRISPR/Cas9 system functions as a pair of "molecular scissors" capable of base editing the two strands of DNA or RNA at a precise location to modulate gene expression through protein attachments (Qi et al. 2013)	Although research is still ongoing, the applications of CRISPR methods are currently being investigated in clinical trials to understand methods of administration and efficacy of CRISPR-based therapeutics in patients

of food quality, and elimination of hypo-allergens associated with some of the food we are currently consuming. In Table 18.1, we list a few of the synthetic biology tools that have been developed, which can tap into agriculture and its production process. These advances have allowed synthetic biologists to develop novel solutions that are already into relevant markets or are moving from lab to the market. In agriculture, the CRISPR/Cas9 technique can allow for the modification of plant genome resulting in desirable traits, thus enhancing the performance of seeds beyond conventional strain development. Cellular agriculture allows for the production of food comprising of engineered microbes and biosynthetic pathways that have medicinal value and for the removal of harmful ingredients containing allergens (Table 18.2).

18.5 Applications of Synthetic Biology in Industrial and Chemical Sector

This section details the applications of synthetic biology in industrial bioproduction platforms and in the chemical sector. Here, we summarize the current contribution of synthetic biology in industrial biotechnology and also focus on its applications in bioeconomy. Industrial synthetic biology is considered as a key enabler for bioeconomy, in promoting a wider application of biological processes in the manufacturing of bio-based products. Here, we summarize the current contribution of synthetic biology in industrial biotechnology and also focus on its applications in bioeconomy. In addition, we focused on TerraVia (microalgae), Evolva (yeast), and Amyris (yeast) as three industrial platform examples.

TerraVia (www.terravia.com) was a biotechnology company that was sold to Corbion N.V., and it has microalgae-based *proprietary technology to convert low-cost plant-based sugars into high-value oils. TerraVia (formally Solazyme)'s raw material* is Brazilian sugarcane that is supplemented with US corn. However, TerraVia's microalgae platform is more flexible in its use of different feedstocks, including cellulosic feedstocks. The microalgae from TerraVia are heterotrophic, are able to ferment in the dark, and are genetically engineered with higher production capacity of over 80%. TerraVia's microalgae are genetically engineered by introducing foreign plant genes and by shutting off certain microalgae's genes involved in undesired oil component production. TerraVia's product portfolio includes industrial products (e.g., surfactants, detergents, biofuels, and hydraulic fluids), cosmetics, detergents for home care, and food products (e.g., omega oils, vegan food products).

Evolva's (www.evolva.com) platform is based on fermentation platform by bioengineered baker's yeast *Saccharomyces cerevisiae*. Here, the plant genes are inserted into the yeast's genome in order to produce plant end products in yeast. In addition, yeast's genes were modified to achieve metabolic pathway optimization that would result in increased yield. In addition, Evolva has proprietary technologies for optimizing transporters, protein glycosylation, and cytochrome 450 enzymes in order to improve the efficiency of production pathways. In addition, Evolva's

Table 18.2 List of applications of synthetic biology in the food and agricultural sector

Synthetic biology tools	Function	Application
Artificial photosynthesis	This process mimics the natural process of converting natural energy by rearranging water molecules and carbon dioxide to create liquid fuel without the need of chlorophyll (Yu and Jain 2019)	It enables the design and engineering of artificial plants using DNA recombination techniques of plants. However, this process is currently evaluated in the production of clean, self-sustaining renewable source of energy and in transportation applications. It is aimed at reducing the release of greenhouse gases into the atmosphere (Yu and Jain 2019)
Engineered bacteria/cellular agriculture	Act as biosensors for the detection of soil contaminants or pathogens that are resistant to disease agents	A classical commercial example of the use of engineered bacteria is the production of sweeteners/sugar substitutes derived from plants, produced in commercial scales using engineered microbes
CRISPR/Cas9	Genome editing in plants, microbes, and algae	The CRISPR technology is already showing promise in the deletion of peanut genes responsible for allergic reactions and ensuring that these genes are switched off
Yeast	Provides a classic platform for metabolic engineering and also the production of medical and food products	A classic example is the biosynthesis of cannabinoids from engineered yeast pathways (Szymanski and Calvert 2018)

terpene platform is currently used for producing a wide variety of medical products and high-quality biofuels. Evolva's current product portfolio focuses on alternatives for high-value plant-based products with limited production capacities such as vanillin.

Amyris (www.amyris.com) is also like Evolva's platform which uses bioengineered yeast *Saccharomyces cerevisiae for fermentation.* Amyris used synthetic biology for bioengineering of their yeast strains and for screening desirable yeast cells which can convert plant-based sugars into high-value hydrocarbon molecules. Currently, Amyris' technology development focuses on a pipeline for selecting a new target molecule that involves producing and creating bioengineered yeast that can produce the target molecule and bioengineering the yeast strains for higher yield and for improved fermentation process (scale-up). Currently, Amyris' product portfolio is based on chemical production platforms such as acetyl-Co-A, farnesene, isoprene, muconic acid platform, and biofuels. Amyris also collaborated with Sanofi to produce artemisinic acid using yeast. Artemisinic acid is a precursor of artemisinin (anti-malarial drug) that is produced from plant-based system and is limited by the production capacity. Similarly, several industries are focusing on engineering microbes as "microbial cell factories" as the superior production biosystems that are highly capable of producing a specific product. This also includes strain development by inserting foreign genes required for host systems to produce the desired product.

18.6 Other Applications of Synthetic Biology

Synthetic biology has been applied for several other fields by acting as an enabler for biotechnology (Vijayachandran et al. 2013), nanotechnology (Thimiri Govinda Raj et al. 2019a), and systems biology applications (Thimiri Govinda Raj et al. 2011). Some of the examples of synthetic biology for other applications include minimal biology, bionanoscience (Thimiri Govinda Raj and Khan 2019), novel research technologies, and new biological applications. To understand the applications of synthetic biology in minimal biology, it is also essential to review the minimal biology concepts (shown in Table 18.3).

18.7 Potential Opportunities and Challenges in Harnessing Synthetic Biology for 2030 UN Sustainable Development Goals

Recently French KL from the University of California Berkeley reviewed the capacities of synthetic biology in addressing the UN SDGs by developing novel products, materials, and service platforms (French 2019). In 2015, the United Nations (UN) listed 17 sustainable development goals that would focus on the development of health care and education, conservation of natural resources, and development of new technologies toward a highly sustainable global society within

Table 18.3 Minimal biology concepts, definition, and synthetic biology application

Concepts	Definition	Ex. of SynBio application
Minimal genome	Bare minimum genome with only essential genes for cellular activity	SynBac: Synthetic baculovirus (Thimiri Govinda Raj et al. 2014)
LUCA	Living systems that existed before the existence of bacteria, archaea, etc.	Directed evolution (Mirkin et al. 2003)
Chassis cell	High-performance cell that is used for laboratory- or industry-scale applications	*M. genitalium* synthetic genome (Gibson et al. 2008)
Artificial/semi-artificial cell	Liposomes or lipid vesicles that encapsulate minimal cells with functional essential biological materials	(Chen et al. 2004)
Minimal cell models (protocells, in silico minimal cells)	Computational cell model with potentially possible constructs required for a minimal cell with high-performance functional capabilities	(Shuler et al. 2012)

2030. Being an enabler, synthetic biology can contribute to sustainability goals by (a) enabling better understanding of living systems' response and its adaption between single-cell and complex communities' environment and (b) creating a generic technology pipeline wherein synthetic biology techniques could be used to slow down global problems. Current understanding shows that synthetic biology could be applied for more than seven sustainable development goals. Some of the contributions include improving the use of biomaterials such as proteins and lipids for new product development and establishing novel technology that can restore the natural ecosystems. In this review, we detailed four areas where synthetic biology contributions are most useful and that coincide with UN SDGs (Tables 18.4 and 18.5).

Addressing key challenges in applying synthetic biology for achieving UN SDGs is a current unmet need to support and promote synthetic biology technologies

Table 18.4 List of synthetic biology applications in bionanoscience and other techniques

Bionanoscience and novel tools	Synthetic biology application
Designer nanoparticles for cell biology and precision medicine	Use of synthetic nanoparticles for subcellular analysis and targeted drug delivery (Thimiri Govinda Raj et al. 2019b,c,d; Son et al. 2019; Chu et al. 2019)
Superparamagnetic virus-like particles (VLPs)	Assembly and characterization of VLPs with magnetic core to study the antigenic potential using ELISA
Single-molecule biophysics	Magnetic tweezers and neuron-on-a-chip (NERF.be)
Quantitative cell biology	Use of synthetic biology to study dynamic cell cycle
Biomaterials and microfluidics	Hydrogels, drug screening, point-of-care diagnostics
Skin microbiome	Acne treatment and cosmetics (S-Biomedic)
Gut microbiome	Human microbiome in cancer (Eligo Bioscience)
Animal microbiome	Applied in the veterinary sector for effective production
Genetic code expansion	Antibody-drug conjugates (ADCs), immunotherapies

Table 18.5 List of synthetic biology applications for UN sustainable development goals

UN sustainable development goals	Synthetic biology applications
SDG 1: No poverty	Implementation of synthetic biology tools in cities promotes economic progress
SDG 3: Good health and well-being	Use of CRISPR and CAR T-cell therapy in medicine
SDG 6: Clean water and sanitation	Engineered microbes used for bioremediation during oil spills
SDG 8: Decent work and economic growth	Establishing start-ups and incubators such as shared lab; Norway promotes synthetic biology firms
SDG 9: Industry, innovation, and infrastructure	Use of skin microbiome in cosmetics. Use of engineered microbes for high production capacities
SDG 14: Life below water	Development of biodegradable plastics to limit pollution and preserve aquatic ecosystems
SDG 15: Life on land	Engineered microbes and plants to remove non-degradable compounds from polluted land

in developing countries. This is particularly critical to achieving UN SDG targets as major progress and developments must occur in developing countries. There is also a need for more progressive policy to moderate the risk factors due to bioengineered living systems on social and environmental sectors. In order to benefit from synthetic biology, more openness and inclusiveness are essential for achieving highly sustainable economies across the world. Currently, synthetic biology innovation is dominated by developed countries, and there is a lack of diversity in global contributions. A key factor in this issue is the current approach with synthetic biology funding.

18.8 Conclusion

The emerging field of synthetic biology has the potential to unravel some of the most unrelenting issues. However, in order to realize the potential benefits of synthetic biology, we will need to address some of the following key issues: optimization for cell-based therapies in vivo, integration of synthetic biology in the fourth industrial revolution (4IR) and society, evaluation of public perception, government regulation, and ethics approval. Although success has been achieved at increasing control and optimization in order to improve bioengineered systems, there are challenges in harnessing synthetic biology, such as the quality control for bioengineered therapies and complexity in using synthetic bio-systems. In order to address these challenges, we need to have an open discussion across all fields including all stakeholders.

Acknowledgments This work was supported by the Synthetic Biology Centre, NextGen Health Cluster, CSIR, Pretoria, South Africa. Authors thank all team members in the CSIR Synthetic Biology Centre.

Competing Interests There is no competing interest.

References

Álvarez B, Fernández LÁ (2017) Sustainable therapies by engineered bacteria. Microb Biotechnol 10(5):1057–1061

Azam A, Tanji Y (2019) Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. Appl Microbiol Biotechnol 103(5):2121–2131

Breuer M et al (2019) Essential metabolism for a minimal cell. Elife 8:e36842

Chen IA, Roberts RW, Szostak JW (2004) The emergence of competition between model protocells. Science 305:1474–1476

Cheng AA, Lu TK (2012) Synthetic biology: an emerging engineering discipline. Annu Rev Biomed Eng 14(1):155–178

Chu DT, Nga VT, Nguyen KH, Tao Y, To TV, Ngoc VTN, Ba NV, Dao DTA, Hoan LN, Hung NP, Thu NTT, Pham V, Nghia LL, Vu LN, Pham TAV, Thimiri Govinda Raj DB (2019) An update on anti-CD137 antibody in immunotherapies for cancer. Int J Mol Sci 20(8):E1822. https://doi.org/10.3390/ijms20081822

Claesen J, Fischbach MA (2015) Synthetic microbes as drug delivery systems. ACS Synth Biol 4(4):358–364
Danchin A, Fang G (2016) Unknown unknowns: essential genes in quest for function. Microb-Biotechnol 9:530–540
Ding Z (2018) The promise and challenges of chimeric antigen receptor T cells in relapsed B-cell acute lymphoblastic leukemia. Ann Transl Med 6(11):235–235
French KE (2019) Harnessing synthetic biology for sustainable development. Nat Sustain 2:250–252
Furfaro L, Payne M, Chang B (2018) Bacteriophage therapy: clinical trials and regulatory hurdles. Front Cell Infect Microbiol 8:376
Gibson DG et al (2008) Completechemical synthesis, assembly, and cloning of a mycoplasma genitaliumgenome. Science 319:1215–1220
Gibson DG et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56
Goold H, Wright P, Hailstones D (2018) Emerging opportunities for synthetic biology in agriculture. Genes 9(7):341
Hong A (2018) CRISPR in personalized medicine: industry perspectives in gene editing. Semin Perinatol 42(8):501–507
Hutchison CA et al (2016) Design and synthesis of a minimal bacterial genome. Science 351(6280):aad6253
Jain A, Bhatia P, Chugh A (2012) Microbial synthetic biology for human therapeutics. Syst Synth Biol 6(1–2):9–22
Jiang F, Doudna J (2017) CRISPR–Cas9 structures and mechanisms. Annu Rev Biophys 46(1):505–529
Kelwick R, MacDonald JT, Webb AJ, Freemont P (2014) Developments in the tools and methodologies of synthetic biology. Front Bioeng Biotechnol 26(2):60
Liang J, Luo Y, Zhao H (2011) Synthetic biology: putting synthesis into biology. Wiley Interdiscip Rev Syst Biol Med 3(1):7–20
Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 3:2
Morozova V, Vlassov V, Tikunova N (2018) Applications of bacteriophages in the treatment of localized infections in humans. Front Microbiol 9:1696
National Cancer Institute (2019) FDA approves second CAR t-cell therapy. https://www.cancer.gov/news-events/cancer-currents-blog/2017/yescarta-fda-lymphoma. Accessed 18 Jul 2019
Patel S, Jones R, Nixon D, Bollard C (2016) T-cell therapies for HIV: preclinical successes and current clinical strategies. Cytotherapy 18(8):931–942
Qi L, Larson M, Gilbert L, Doudna J, Weissman J, Arkin A, Lim W (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of geneexpression. Cell 152(5):1173–1183
Roberts MA, Cranenburgh RM, Stevens MP, Oyston PC (2013) Synthetic biology: biology by design. Microbiology 159(Pt 7):1219–1220
Shuler ML, Foley P, Atlas J (2012) Modeling a minimal cell. Methods Mol Biol 881:573–610
Si T, Zhao H (2016) A brief overview of synthetic biology research programs and roadmap studies in the United States. Synth Syst Biotechnol 1(4):258–264
Son TM, Ngoc VTN, Tran PT, Nguyen NP, Luong HM, Nguyen HT, Sharma K, Tu PV, Ha LS, Ha VN, Huy PV, Thimiri Govinda Raj DB, Chu DT (2019) Prevalence of dental fear and its relationship with primary dental caries in 7-year-old-chidren. Pediatr Dent J 29(2):84–89. https://doi.org/10.1016/j.pdj.2019.04.002
Szymanski E, Calvert J (2018) Designing with living systems in the synthetic yeast project. Nat Commun 9(1):2950
Thimiri Govinda Raj DB, Khan NA (2019) Protocol for eukaryotic plasma membrane isolation using superparamagnetic nanoparticles. J Magn Magn Mater 476:628–631

Thimiri Govinda Raj DB, Ghesquiere B, Tharkeshwar AK, Coen K, Derua R, Vanderschaeghe D, Rysman E, Bagadi M, Baatsen P, De Strooper B, Waelkens E, Borghs G, Gevaert K, Annaert W (2011) A novel strategy for the comprehensive analysis of the biomolecular composition of isolated plasma membranes. Mol Syst Biol 7:541

Thimiri Govinda Raj DB, Garzoni F, Gavin AC, Gibson T, Berger I (2014) SynBac: designer minimal baculovirus genome for drug discovery. FEBS J 281(s1):9–55

Thimiri Govinda Raj DB, Khan NA, Venkatachalam S, Chu DT (2019a) Step by step protocol for superparamagneticnanoparticle based plasma membrane isolation from eukaryotic cell. Methods Mol Biol. https://doi.org/10.1007/7651_2019_211

Thimiri Govinda Raj DB, Khan NA, Venkatachalam S, Arumugam S, Palaniswamy S (2019b) Synthesis of hybrid gold nanoparticle (AuNP) functionalized superparamagnetic nanoparticles (SPMNP) for efficient coupling of biomolecules. Methods Mol Biol. https://doi.org/10.1007/7651_2019_254

Thimiri Govinda Raj DB, Khan NA, Venkatachalam S, Arumugam S (2019c) BacMam system for rapid recombinant protein expression in mammalian cells. Methods Mol Biol. https://doi.org/10.1007/7651_2019_249

Thimiri Govinda Raj DB, Khan NA, Venkatachalam S, Chu DT, Arumugam S (2019d) Step by step protocol for superparamagnetic nanoparticle based endosome and lysosome isolation from eukaryotic cell. Methods Mol Biol. https://doi.org/10.1007/7651_2019_212

Vijayachandran LS, Thimiri Govinda Raj DB, Edelweiss E, Gupta K, Maier J, Gordeliy V, Fitzgerald DJ, Berger I (2013) Gene gymnastics: synthetic biology for baculovirus expression vector system engineering. Bioengineered 4(5):279

Xavier JC et al (2014) Systems biology perspectives on minimal and simpler cells. Microbiol Mol Biol Rev 78(3):487–509

Yu S, Jain P (2019) Plasmonic photosynthesis of C1–C3 hydrocarbons from carbon dioxide assisted by an ionic liquid. Nat Commun 10(1):2022

Zhang M, Zhang D, Shi H (2017) Application of chimeric antigen receptor-engineered T cells in ovarian cancer therapy. Immunotherapy 9(10):851–861

Zhen S, Li X (2019) Application of CRISPR-Cas9 for long noncoding RNA genes in cancer research. Hum Gene Ther 30(1):3–9

Development and Application of Microfluidics in Synthetic Biology

19

Boris Kirov

Abstract

The developments of synthetic biology and microfluidics have been intertwined ever since the introduction of the first synthetic genetic parts. In this chapter the main physical and technological features that rendered microfluidics technology indispensable for synthetic biology research are discussed, and some preliminary guidelines for microfluidics devices design are shared by the author. Advances in the field regarding the utilization of the discussed technique at many levels of biological systems organization are reviewed. Finally, some considerations of the author regarding the future development and applicability of microfluidics in synthetic biology and synthetic biology per se are presented.

Keywords

Synthetic biology · Microfluidics · On-chip technology · Artificial cells · Artificial organs · Cell-free systems

19.1 Introduction

The developments of synthetic biology and microfluidics have been intertwined ever since the introduction of the first synthetic genetic parts. This relationship, however, is not only based on love at first sight, but mostly on mutual interest. In this evergreen duet, microfluidics provided the means and synthetic biology generated the ends. Indeed, microfluidics rendered possible the explosive development of synthetic biology in the twenty-first century, and synthetic biology justified by far

B. Kirov (✉)
Faculty of Automatics, TU – Sofia, Sofia, Bulgaria
e-mail: boris.kirov@tu-sofia.bg

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_19

the existence and evolution of microfluidics in the same time period (Supramaniam et al. 2019).

The specifics of microfluidics, which made the latter invaluable for synthetic biology, have been thoroughly reviewed elsewhere (Gulati et al. 2009). In summary, it is mainly about physics. The strict definition of microfluidics device is a fluidics device, which comprises compartments, which have at least one dimension within the micron scale (i.e., between 10^{-6} and 10^{-3} m). The resulting compact space for fluidic flow squeezes tightly the fluidic particles converting them into a single stable package, in which the separate particles move very slowly (if at all) with respect to each other. In physical terms this is a strictly laminar flow characterized by an extremely low ratio of inertial to viscous forces (i.e., Reynolds number) (Bruus 2008).

The direct consequence of such type of organized flow is that the cultivation of cells within such type of devices differs harshly from the typical microbiological or biotechnological processes of the same type. In the latter two cases, the molecules from the medium, which are in direct contact with each individual cell, are a random combination of a whole plethora of compounds—the constituents of the fresh medium, the products of the metabolic activity exported by the cells, cytosolic constituents of lysed cells, etc. Conversely, the strictly laminar flow in microfluidics devices guarantees that there is no mixing between the incoming fresh medium and the culture milieu and that the latter is always completely removed by the former. Consequently, the substances contacting the individual cells are very well defined and completely known. In short, cell cultivation in microfluidics devices is conducted in *chemically defined and known environment*.

Secondly, fabrication of microfluidics devices was based directly on the accomplishments of large-scale microfabrication for microelectronics. Thus, on one hand, it was a very well-studied, straightforward, and relatively cheap process, and on the other, the integration of microsensors in the microfluidics devices was easy and affordable. The latter means that in addition to the known chemical composition of the cellular environment, one could *control precisely also the physical properties of the growth environment* such as temperature, impedance, different types of absorbance and emission spectra, etc. Furthermore, microfluidics devices could be produced in direct contact with microscope cover slips, thus also allowing for direct optical observation of the processes within. From an engineering point of view, this type of observability provides the opportunity for precise mathematical modeling and *controllability of the processes*.

Last, but not least, the size of the microfluidics devices allows for the fabrication of cultivation chambers of dimension similar to the cellular ones. Thus, one could easily produce chambers that fit a limited number of cells (even only one) and observe directly the *single-cell physiology*, the latter being only dreamt of by generations of biologists. In addition, other types of structures, both intra- and extracellular, could be fabricated within microfluidics chips, thus allowing for detailed studies of biological processes at different scales.

For all the abovementioned, microfluidics was lavishly endorsed by synthetic biologists in our quest to study and engineer life and its systems at any possible

level. In this chapter we are going to review the plethora of applications that this technology has received for the study of biological systems at molecular, subcellular, cellular, tissue, organ, and even ecosystem level. Of course, the suitability of microfluidics for biological studies has led to the engineering also of many such devices for purely fundamental research purposes, which we are going to omit in the following discussion.

19.2 Microfluidics Devices and Synthetic Biology Research

19.2.1 Molecular Level

The task to pinpoint precisely the first synthetic biology device is a difficult one, since the field in itself is not yet thoroughly defined. However, it is fair to assume that what we understand nowadays by the term began with the engineering of the first genetic devices of logical type, namely, logic gates, switches, and oscillators (Ausländer et al. 2017). All of those are examples of synthetic parts engineered at the molecular level. In some of those cases, the activity of the genetic part (or circuit of parts) has stationary expression in time, and hence, this activity is similar throughout the whole population of cellular chassis. Typical examples of such devices are logic gates and toggle switches, and more of them are being engineered with different applications in many biological and biomedical fields. For example, a tunable dual promoter system was used to create a logical "AND" gate for the specific detection of cancer cells (Nissim and Bar-Ziv 2010). Other researchers have gone even further with the complexity of their devices and have introduced a quadruple-input and gate functioning with more than ten protein species and organized in multiple layers (Moon et al. 2012). Consequently, characterization of such parts could be performed under normal cultivation conditions, which allow for population-level observations without explicit need for microfluidics devices.

Other types of synthetic genetic devices like genetic oscillators, however, exhibit dynamically changing behavior, which is not necessarily synchronized between the individual cells. Consequently, population-level observations could not be utilized for the precise characterization of such parts. It was exactly for the first tunable stably oscillating genetic device developed by the group of Jeff Hasty (Stricker et al. 2008) that microfluidics cultivation device for *Escherichia coli* was utilized for single-cell-level characterization in a thoroughly elaborate manner.

This work was pioneering in the field in many senses, since it was the first to introduce the basic rules for the design of microfluidics devices for single-layer cultivation of bacteria and for the operation of such devices (Ferry et al. 2011). In short, in order to obtain single layer, the designer needs to define growth-chamber depth no bigger than 1.5 times the small diameter of the cells to be cultivated, so that the latter are forced to grow in a single layer. Of course, one also needs to observe the elastic properties of the building material when designing the horizontal dimensions of the device if ceiling collapse is to be avoided. In the case of PDMS, this means that the ratio between the long horizontal side of a given hollow structure

and the height (or depth) of the same structure should not exceed 50 (better still 30). For single-layer cultivation of *E. coli*, e.g., the depth of the growth chamber should be between 1 and 1.5 μm (given the small diameter of those bacteria of about 1 μm), and the length of the walls of such chamber should not exceed 50 μm.

Regarding the operation of microfluidics devices for living cell cultivation, the group of Hasty has introduced the practice of gravitational method for fluid flow generation. This method has a number of advantages with respect to typical syringe pumps such as soft fluid motion without mechanical stress over the cells, reduced hydraulic pressure, and no specific equipment requirement. However, lower pressure typical of this method renders fluid priming of the microfluidics chips (including air bubbles removal) rather difficult and dynamic programming of input medium practically impossible without additional equipment.

Since this work, similar methods have been used for the characterization of a number of genetic parts and devices (Potvin-Trottier et al. 2018).

19.2.2 Tissue Level

It is interesting to note that within the context of utilization of microfluidics chips for single-cell characterization, those devices have a somewhat double function. On one hand, they represent simply equipment for observation and study of the given object. However, in this case microfluidics devices are also utilized as means to transport chemical compounds and maintain proper temperature, in short to generate some sort of homeostasis for the studied cells and to coordinate the physiological activity of the latter. Thus, microfluidics devices also become some sort of artificial circulatory system for an engineered synthetic bacterial "tissue."

Structural parts of microfluidics devices have been successfully used to mimic parts of similar types of artificial tissues of higher organisms in a number of studies, most of which are aimed at the study of the interactions of their natural counterparts with different agents (pharmaceutics, foreign cells, other cell types, etc.).

19.2.2.1 Immune System

In most of those cases microfluidics were used to mimic the natural physical and chemical environment (including circulatory system) for the engineering of artificial tissues. One particular example is the simulation of different aspects of immune cells' reaction to specific stimuli, which were mimicked by the shape of microfluidics chips (i.e., lesions), patterning of chambers with specific antibodies aimed for the selection of active forms of neutrophils (Bose et al. 2013), and fabrication of microfluidics devices capable of generating stable gradients of chemoattractants (Li Jeon et al. 2002), which led to interesting discoveries regarding the human neutrophils' chemotaxis and revealed that gradient steepness and absolute concentration have complex interactions, which lead to specific behavior of the immune cells. Furthermore, microfluidics devices allowed for the study of leukocyte migration in confined spaces (Irimia et al. 2007), which showed that narrow passages lead to constant progression speed in contrast to the behavior

observed on flat surfaces. The cited are just few of the research efforts toward imitation and control of some of the characteristic stimuli controlling the immune cells. More studies were aimed at cellular interaction, competing chemical signals resolution, hereditary differences between lineages' behavior, etc.

19.2.2.2 Neural Tissue

Maybe the most important target tissue for environmental physical and chemical mimicking remains the neural one. Neural cells have been successfully maintained in artificial 2D and 3D scaffolds; however, without proper nutrients and oxygen delivery, their behavior remained somewhat unpredictable. Microfluidics allowed for the tissue-engineered muscle cells to be grown directly on chip with space for contraction left (Taylor et al. 2003). After the development of neural cells, activation of the muscle cells was successfully monitored. Furthermore, a microfluidics chip was fabricated with dedicated 500-μm-long channel separating compartments where neurons and muscle fibers were separately grown (Park et al. 2013). The developed motoneurons could successfully bridge those channels, and this system was successfully utilized to study the effect of different growth factors on the neural development. Another research aimed at the reconstruction of the succession of axon-glia connection formation by separating the neural cell bodies from the glial cells (Park et al. 2009) and connecting them only through microchannels. Consequently, axons were successfully formed and the system was used for the study of myelination processes. Many other studies employing microfluidics worked on neural cell-cell interaction, neural tissue inflammation, peripheral nerve regeneration, blood-brain barrier, brain cancer, and many more.

19.2.3 Organ Level

Same methods for utilization of microfluidics devices could, of course, be used also at the next biological organization level within a certain organism, namely, organs. There are also a number of studies involving organ-on-chip.

19.2.3.1 Skin

The skin being the largest organ, obviously, is of main interest. There are some skin-on-chip models, which differ in their level of complexity and vascularization. For example, there is a simplistic model developed by a simple cultivation of human epidermal keratinocytes grown on collagen pieces in microfluidics device (O'Neill et al. 2008). The next level of complexity of skin-on-chip consists of both dermis and epidermis layers (Sriram et al. 2018). The dermal part is engineered fibroblasts seeded in fibrin-rich matrix, whereas the epidermal part is generated by keratinocytes cultured directly on the dermal analogue. This model allowed for the testing of the effect different surface interfaces have on the development of the organ and for the proof that air interface is benevolent for the formation of the final model epidermis layer (Abaci et al. 2015). As stated above, the next step toward even higher level of complexity is the vascularization of the skin model. A

successful model comprising endothelial cell penetrating through the separate skin layers has been developed (Wufuer et al. 2016). In addition, this model has been further rendered more engineer-able through the separation of the individual skin layers by polymer porous membrane, thus laying the bases for later exchange of individual layers with artificial substitutes avoiding the utilization of living cells and the generation of a completely artificial organ.

19.2.3.2 Liver

The next organ that has provoked large interest in the field of microfluidics organ-on-chip development is the liver. The latter combines a number of functions of utmost importance for the organism such as metabolism, detoxification, pharmaceutics removal, immunity, etc. Being relatively simply organized, the liver has been the subject of the attempts for a lot of simplistic artificial analogues. Some studies are based on a model comprising solely two channels separated by a porous membrane being able to support living hepatocytes for as many as 28 days (Prodanov et al. 2016). A simple addition to this model of endothelial cells in another work (Kang et al. 2015) was enough for the successful study of hepatitis B virus replication for more than 30 days. Furthermore, this model continued to produce urea for the whole experiment duration, showing that such simplistic organization is a good basis for the development of a functional artificial organ.

19.2.3.3 Spleen

Although the spleen is considered to be an expendable organ, there were also successful attempts to reproduce its function in vitro within microfluidics devices. In one example, the authors engineered a device, which utilized antibody-coated magnetic nanobeads capable of binding almost 100 types of pathogenic surface antigens (Kang et al. 2014). The pathogens bound to the nanoparticles were magnetically removed from the blood stream, whereas the latter was returned to the patient. The described device was developed as sepsis treatment option and is another example for a completely artificial instrument, which mimics just the function of the organ, without the utilization of any living cells or other biological systems.

19.2.3.4 Lung

The artificial lung-on-chip introduced one more level of depth in the natural organ mimicking through the addition of physiology-like tissue motions in a daring research (Huh 2015). In short, a relatively simple microfluidics device was used to grow alveolar epithelial cells and pulmonary microvascular endothelial cells on the opposite sides of a porous 10-μm-thick membrane made from the typical material for microfluidics devices poly(dimethylsiloxane) (PDMS). The membrane was used both to provide interface surface for the two tissue types and for nurturing them through the micropores. Additionally, laterally to the double membrane-separated chamber were fabricated hollow elastic chapters, which could be easily bent. This way, upon vacuum generation within the hollow chambers, the latter squeeze and stretch laterally the perfused membrane, which is directly contacting them through

an elastic wall. Hence, simulation of physiological breathing motion was obtained. In this organ model the effects of alveolar exposure to nanoparticles, as well as pulmonary edema, were successfully studied.

19.2.3.5 Kidney

On the other hand, the kidney is the organ of utmost importance for drug discovery, since a lot of proposed pharmaceutics get discarded through trials at the animal kidney toxicity testing. The latter, unfortunately, is not a reliable method, since it does not take into account metabolic differences between humans and animals. Consequently, non-toxic-for-humans drugs get rejected and toxic-for-humans drugs are allowed to continue toward the next evaluation phase. The simplest kidney-on-chip instrument comprises human kidney cells attached to the bottom of a microfluidics channel, the latter consequently being subjected to shear stress (Huang et al. 2013). The latter increased cellular thickness and proved to be beneficial for the artificial organ function. The next development of that organ-on-chip included a two-layered device, which successfully imitated reabsorption function (Jang and Suh 2010).

19.2.3.6 Gut

The microfluidics device consisting of two compartments separated by a porous membrane was used successfully also for the engineering of a gut-on-chip. Initially, this type of device did not show proper physiological function (Kimura et al. 2008) unless subjected to shear stress and peristalsis-like movements (Kim and Ingber 2013). In the latter case spontaneous 3D villi morphogenesis was observed, and the device was even utilized for successful co-culturing of typical commensal microbes.

19.2.4 Bioprocess Level

Another obvious application of microfluidics technology within the realm of synthetic biology is for the study of synthetic bioprocesses. The already discussed capacity of those devices to provide good control and unmatched process observability represents an extremely interesting opportunity for biotechnology research. Apart from those features, there is also the opportunity for study of novel strains at extremely small scales within fully automated devices which renders the process of introduction of novel improved strains much cheaper, faster, and a lot less laborious. Microfluidics devices have already been used for the engineering of bioprocesses in a number of cases. One such example is the direct selection of improved strains of the lipid-producing yeast *Yarrowia lipolytica*, which was conducted directly in droplets formed in microfluidics device (Huang et al. 2015). This approach allowed for the measurement of the extracellular lipid concentration, which remained inside the droplet, thus overcoming one of the major obstacles for selection of strain productivity in microfluidics chips. Still, the capacity of the latter technology to provide opportunity for parallel cell sorting at significantly lower costs with respect to macro-technologies remained unmatchable. A similar approach, but based on

dielectrophoretic selection, was also utilized for the characterization of *Escherichia coli* strains (Wang et al. 2014) and mammalian cells (Mazutis et al. 2013).

Additionally, analogous microfluidics setups have been used for the characterization of total bioprocess parameters such as oxygen uptake (Super et al. 2016), growth kinetics (Oliveira et al. 2016), and ethanol production (Churski et al. 2015). Droplet-based microfluidics reactors have also been utilized for top-down approach or miniaturization of existing bioprocesses (Lin et al. 2015).

However, such miniaturized cell factories have limited supply of nutrients and other compounds (e.g., oxygen) and cannot be used for the scaling-up of continuous biotechnological production processes. Therefore, a number of microbioreactors of chemostat type were developed proving that the scaling-up process is efficient even at that micro-level, providing technology characterized by immense cost reduction for bioprocess characterization. Such devices were developed for continuous bacterial (Kim et al. 2017) and microalgae (Paik et al. 2017) culture growth, for aerated yeast growth (Funke et al. 2010), etc.

19.2.5 Supracellular Level

Supracellular organization exists naturally in bacteria, and one typical example of such behavior is the formation of biofilms. As already discussed microfluidics channels and chambers provide artificial method to control and synchronize the intercellular environment and, hence, cellular physiology. This capacity, of course, could be used to incite biofilm formation. Actually, it is an automatic consequence of scarce nutrient delivery within the microfluidics chips, which is usually provoked by channel clogging. However, there are a number of studies aimed precisely at biofilm formation and physiology experiments within microfluidics devices. Studies related to biofilm formation and physiology include research over the effect antibiotics have on individual bacterial cells (Boedicker et al. 2008), over the emergence and propagation of antibiotic resistance in interconnected microcolonies (Zhang et al. 2011), on bacterial persistence (Balaban et al. 2004), in the field of quorum sensing (Boedicker et al. 2009), and finally, on bacterial growth parameters in microfluidics environment (Balagaddé et al. 2005).

Furthermore, mature biofilms in microfluidics devices were studied for their reaction to different compounds (Kim et al. 2012) and oxygen concentrations (Skolimowski et al. 2010). Additionally, opportunities for extensive fluid flow control within microfluidics devices were utilized for the study of flow velocity influence over biofilm formation and the proof that slower flow rates are connected to thicker biofilms (Lee et al. 2008) and that curved fluid paths provoke the formation of streamers (Rusconi et al. 2010), the latter being a large component of the biofilms formed within man-made technological environment. Additionally, a similar approach was utilized for the characterization of adhesion forces of pili produced by biofilms (De La Fuente et al. 2007).

Another interesting aspect of biofilms is their species heterogeneity. Microfluidics devices were successfully utilized for precise control of quorum sensing

parameters with a dual-bacteria mixed population and thus to force different ratios between the individual species within the formed biofilm (Hong et al. 2012).

19.2.6 Ecosystem Level

19.2.6.1 Human Microbiome

Microfluidics devices are utilized to mimic the extracellular environment at one final biological organizational level, i.e., the ecosystem. Furthermore, this technology has already been utilized extensively for the study of microbiomes in a number of articles. Microfluidics devices have been used for the study of the increased antimicrobial tolerance within microbial consortia (Mohan et al. 2015). Furthermore, the gut-on-chip microfluidics technology has been used for the study of gut bacterial overgrowth and anti-inflammatory probiotics (Kim et al. 2016) and intestinal barrier integrity through the direct electrode measurement of transepithelial electrical resistance (Henry et al. 2017). A promising approach toward the study of microbiome-host interactions is the co-culture systems (Shah et al. 2016). There are even reports for devices for the cultivation of personal microbiomes from samples from different organs for the purpose of pharmaceutics and probiotics reaction studies. Microfluidics-grown bacterial consortia have also been used for the study of bacterial vaginosis and specifically antimicrobial properties of some bacterial species (Matu et al. 2010) and metabolic interspecies dependence (Pybus andOnderdonk 1997).

19.2.6.2 Aquatic and Soil Microbiome

Microscopic ecosystems, however, exist not only within the organs of higher organisms but also in natural aquatic and soil habitats. For example, microfluidics-based device was utilized to study the chemotactic properties of marine microorganisms in Australian oceans allowing for capturing of microbes with a nutrient "bait" and later individualize their species through 16S RNA methods (Lambert et al. 2017).

Finally, microfluidics technology provides a unique opportunity to study soil microorganisms in physical environment closely mimicking their original habitat through microfabrication of specific structures resembling soil particles (Deng et al. 2015) or even by directly integrating such particles within the microfluidics chip (Pronk et al. 2017).

19.2.7 Cellular Level

Up till now we have discussed solely the application of microfluidics devices and their comprising structures as means to reproduce or mimic the natural surroundings of cells, tissues, organs, organisms, or even ecosystems. Important as they are, those types of studies utilized microfluidics technology simply as a passive vessel in which laboratory engineered biological systems are tested, observed, and characterized.

However, this technology is utilized even further for the purposes of synthetic biology as an indispensable part of the engineering of synthetic cells. The latter activity is the first attempt of humanity to reconstruct de novo the building block of all living creatures that populate our planet. Indeed, one of the major impediments of biology has always been the obligatory reproduction of living cells from living cells. We have succeeded in manipulating, destroying, engineering, selecting, evolving, etc. living cells in whatever way we thought of; however, we could never generate a single cell from scratch. Microfluidics finally seems to have provided us with the means to attempt even that puzzle in a number of ways.

As stated above, we know how to manipulate and engineer cellular subparts. We studied quite well the underlying mechanisms of cellular metabolism, information propagation, growth, and division. We know how to use this knowledge for our purposes in order to generate and control processes within already existing living cells or in vitro. However, the passage from in vitro experiment to a living entity seemed impenetrable until recently.

We could even strip a living cell from all of its external barriers that separate it from its environment (i.e., membrane and wall) and utilize the remaining cytoplasmic broth to reproduce cellular metabolism (Shin and Noireaux 2012). We could further exchange the genes of this mixture and it would still function. If we supplied it with additional nutrients and energy sources, it would remain functional for quite some time. Similar systems have been used within microfluidics devices for the studies of the precise kinetics of synthetic genetic parts and devices such as genetic logic gates (Shin and Noireaux 2012), cascades (Noireaux et al. 2003), and oscillators (Niederholtmeyer et al. 2013). Cell-free systems were capable even of reconstructing peptide biosynthesis (Goering et al. 2016).

What was missing in such systems to complete a living cell was the membrane. Hence, there appeared a number of studies where the generation of artificial membrane for cell-free systems was studied employing different methods all relying on the microfluidics devices' capacity to control precisely timing and positioning of different types of fluids—lipids, solvents, water-based solutions, etc. Indeed, microfluidics devices' capacity to be utilized for droplet formation became indispensable for membrane vesicle construction. Through such approaches lipid bilayers were successfully constructed around vesicles, thus allowing for the successful capturing of large charged molecules (van Swaay 2013). Giant unilamellar vesicles are nowadays usually produced within microfluidics devices in a number of methods: electroformation, hydration, fluidic jetting, etc.

Interesting as they were, the above experiments overlooked one important common feature shared between cellular membranes and microfluidics. Namely, it is the selective control over what types of substances that gain access to the different cellular machineries. The cellular membrane separates the cytoplasm and everything it contains from the cellular environment, and only certain small molecules could passively penetrate inside it. The only other remaining option is the mediated transport, which is under strict control through the types of transporting molecular machines that could be found at certain moment within the cellular boundaries.

Therefore, two of the important functions of the artificial membrane would be also to protect the cell and to control the access to its interior.

However, precise fluidic control and extremely difficult diffusion in microfluidics devices also allow for the strict control of access of different substances within confined microfluidics chambers. Consequently, the generation of second layer of control through artificial membrane within an already controlled microfluidics environment represents an expensive and unnecessary functional redundancy. A different and more engineering approach toward the generation of artificial cell would be to utilize the properties that are already provided by the physics of microfluidics and to work toward overcoming the remaining challenge, i.e., active transport through the physical barrier surrounding the cell. A number of studies have already aimed at the generation of functional cell-free systems capable of genetic expression and metabolic activity utilizing microfluidics chambers instead of artificial membrane-separated vesicles. One such system was utilized to characterize zinc-finger-based promoters and their capacity to regulate genetic expression, including for the formation of logic gates and cooperative genetic parts (Swanka et al. 2019). In another example, studies for protein-protein interactions were conducted within such cell-free microfluidics system (Gerber et al. 2009). Finally, one cell-on-chip system was constructed utilizing DNA brushes to encode the device behavior, which demonstrated metabolic, synthetic, and even inter-compartment communication activity (Karzbrun et al. 2014).

19.3 Microfluidics and Laboratory Methods

Finally, one could not underestimate the importance of microfluidics technology for the automation of traditional and novel laboratory tasks, especially in the field of molecular genetics. Given that the most typical characteristic of microfluidics devices is the capacity to precisely control fluid flow and that most (if not all) molecular biology laboratory methods consist of fluid mixing and separation, it is of no surprise that automated microfluidics-based procedures were developed for many of the steps comprising a genetics experiment. On top of the simplest elementary processes, there were also developed devices for automated DNA assembly methods based on restriction enzymes (Ellis et al. 2011), multiplex automated genome engineering (Wang et al. 2009), CRISPR/Cas9 (Han et al. 2015), and direct combinatorial DNA synthesis (Ochs and Abate 2015).

19.4 Conclusion

In conclusion, we have shown that microfluidics has penetrated modern biology research at any possible organizational level—molecular, subcellular, cellular, supracellular, tissue, organ, system, and ecosystem. This technology has provided a number of experimental advantages allowing for experiments that would be impossible in any other environment and at the same time reducing significantly research

costs. Furthermore, utilization of microfluidics devices provided the opportunity to better control the research object, i.e., rendered all of the abovementioned types of biological systems engineer-able. At some organizational levels, actual artificial biological systems have been created, namely, cells and spleen, and many more seem to be on the way. As a consequence, all biology seems synthetic nowadays.

Seemingly, it would be hard to think of modern synthetic biology experiment outside of a chip, and it becomes unavoidable for scientists from our field to develop without knowledge of this technology, which involves physics, chemistry, and microfabrication the least.

However, this heavy burden is alleviated by the natural course of events, because like all technologies that reach mass utilization and production, microfluidics has reached the point at which it is not possible to produce research prototype devices at the academic foundries cheaper than specialized producers. Recently, the price for direct order for PDMS prototype of a microfluidics device, based solely on design CAD file provided by the client, has fallen to 20 euros per chip. The latter is a price much lower than the cost of the consumables and labor, which need to be invested in the fabrication of a prototype chip, without calculating the price of the equipment required for such type of fabrication.

Therefore, instead of yoking oneself with heavy studies within many different knowledge fields, the modern synthetic biology researcher needs to learn just the design principles of microfluidics and be brave. As for the rest, imagination seems really to be the only limit.

References

Abaci HE, Gledhill K, Guo Z, Christiano AM, Shuler ML (2015) Pumpless microfluidic platform for drug testing on human skin equivalents. Lab Chip 15:882–888

Ausländer S, Ausländer D, Fussenegger M (2017) Synthetic biology-the synthesis of biology. AngewChemInt Ed Engl 56(23):6396–6419. https://doi.org/10.1002/anie.201609229

Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. Science 305:1622–1625

Balagaddé FK, You L, Hansen CL, Arnold FH, Quake SR (2005) Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. Science 309:137–140

Boedicker JQ, Li L, Kline TR, Ismagilov RF (2008) Detecting bacteria and determining their susceptibility to antibiotics by stochastic confinement in nanoliter droplets using plug-based microfluidics. Lab Chip 8:1265–1272

Boedicker JQ, Vincent ME, Ismagilov RF (2009) Microfluidic confinement of single cells of bacteria in small volumes initiates high-density behavior of quorum sensing and growth and reveals its variability. AngewChemInt Ed Engl 48:5908–5911

Bose S, Singh R, Hanewich-Hollatz M, Shen C, Lee CH, Dorfman DM, Karp JM, Karnik R (2013) Affinity flow fractionation of cells via transient interactions with asymmetric molecular patterns. Sci Rep 3:2329

Bruus H (2008) Theoretical microfluidics. Oxford University Press, Oxford

Churski K, Ruszczak A, Jakiela S, Garstecki P (2015) Droplet microfluidic technique for the study of fermentation. Micromachines 6(10):1514–1525. https://doi.org/10.3390/mi6101435

De La Fuente L, Montanes E, Meng Y, Li Y, Burr TJ, Hoch H, Wu M (2007) Assessing adhesion forces of type I and type IV pili of Xylella fastidiosa bacteria by use of a microfluidic flow chamber. Appl Environ Microbiol 73:2690–2696

Deng J, Orner EP, Chau JF, Anderson EM, Kadilak AL, Rubinstein RL et al (2015) Synergistic effects of soil microstructure and bacterial EPS on drying rate in emulated soil micromodels. Soil BiolBiochem 83:116–124

Ellis T, Adie T, Baldwin GS (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. IntegrBiol 3(2):109–118. https://doi.org/10.1039/c0ib00070a

Ferry MS, Razinkov I, Hasty J (2011) Microfluidics for synthetic biology: from design to execution. Methods Enzymol 497:295–372. https://doi.org/10.1016/B978-0-12385075-1.00014-7

Funke M, Buchenauer A, Schnakenberg U, Mokwa W, Diederichs S, Mertens A, Müller C, Kensy F, Büchs J (2010) Microfluidic biolector-microfluidic bioprocess control in microtiter plates. BiotechnolBioeng 107(3):497–505. https://doi.org/10.1002/bit.22825

Gerber D, Maerkl SJ, Quake SR (2009) An in vitro microfluidic approach to generating protein-interaction networks. Nat Methods 6:71–74

Goering AW et al (2016) In vitro reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. ACS Synth Biol 6:39–44

Gulati S, Rouilly V, Niu X, Chappell J, Kitney RI, Edel JB, Freemont PS, deMello AJ (2009) Opportunities for microfluidic technologies in synthetic biology. J R Soc Interface 6(Suppl 4):S493–S506. https://doi.org/10.1098/rsif.2009.0083.focus

Han X, Liu Z, Jo MC, Zhang K, Li Y, Zeng Z, Li N, Zu Y, Qin L (2015) CRISPR-Cas 9 delivery to hard-to-transfect cells via membrane deformation. SciAdv 1(7):e1500454. https://doi.org/10.1126/sciadv.1500454

Henry OYF, Villenave R, Cronce MJ, Leineweber WD, Benz MA, Ingber DE (2017) Organs-on-chips with integrated electrodes for trans-epithelial electrical resistance (TEER) measurements of human epithelial barrier function. Lab Chip 17:2264–2271. https://doi.org/10.1039/C7LC00155J

Hong SH, Hegde M, Kim J, Wang X, Jayaraman A, Wood TK (2012) Synthetic quorum-sensing circuit to control consortial biofilm formation and dispersal in a microfluidic device. Nat Commun 3:613

Huang H, Chang Y, Chen W, Harn H, Tang M, Wu C (2013) Enhancement of renal epithelial cell functions through microfluidic-based coculture with adipose-derived stem cells. Tissue Eng Part A 19(17–18):2024–2034

Huang M, Bai Y, Sjostrom SL, Hallström BM, Liu Z, Petranovic D, Uhlén M, Joensson HN, Andersson-Svahn H, Nielsen J (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. PNAS 112(34):E4689–E4696. https://doi.org/10.1073/pnas.1506460112

Huh DD (2015) A human breathing lung-on-a-chip. Ann Am ThoracSoc 12(Suppl 1):S42–S44. https://doi.org/10.1513/AnnalsATS.201410-442MG

Irimia D, Charras G, Agrawal N, Mitchison T, Toner M (2007) Polar stimulation and constrained cell migration in microfluidic channels. Lab Chip 7:1783–1790

Jang KJ, Suh KY (2010) A multi-layer microfluidic device for efficient culture and analysis of renal tubular cells. Lab Chip 10(1):36–42. https://doi.org/10.1039/b907515a

Kang JH, Super M, Yung CW, Cooper RM, Domansky K, Graveline AR, Mammoto T, Berthet JB, Tobin H, Cartwright MJ et al (2014) An extracorporeal blood-cleansing device for sepsis therapy. Nat Med 20:1211–1216

Kang YBA, Sodunke TR, Lamontagne J, Cirillo J, Rajiv C, Bouchard MJ, Noh M (2015) Liver sinusoid on a chip: long-term layered co-culture of primary rat hepatocytes and endothelial cells in microfluidic platforms. BiotechnolBioeng 112:2571–2582

Karzbrun E, Tayar AM, Noireaux V, Bar-Ziv RH (2014) Programmable onchip DNA compartments as artificial cells. Science 345:829–832

Kim HJ, Ingber DE (2013) Gut-on-a-Chip microenvironment induces human intestinal cells to undergo villus differentiation. IntegrBiol 5(9):1130–1140. https://doi.org/10.1039/c3ib40126j

Kim J, Hegde M, Kim SH, Wood TK, Jayaraman A (2012) A microfluidic device for high throughput bacterial biofilm studies. Lab Chip 12:1157–1163
Kim HJ, Li H, Collins JJ, Ingber DE (2016) Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. ProcNatlAcadSci U S A 113:E7–E15. https://doi.org/10.1073/pnas.1522193112
Kim M, Bae J, Kim T (2017) Long-term and programmable bacterial subculture in completely automated microchemostats. Anal Chem 89(18):9676–9684
Kimura H, Yamamoto T, Sakai H, Sakai Y, Fujii T (2008) An integrated microfluidic system for long-term perfusion culture and on-line monitoring of intestinal tissue models. Lab Chip 8(5):741–746. https://doi.org/10.1039/b717091b
Lambert BS, Raina JB, Fernandez VI, Rinke C, Siboni N, Rubino F, Hugenholtz P, Tyson GW, Seymour JR, Stocker R (2017) A microfluidics-based in situ chemotaxis assay to study the behaviour of aquatic microbial communities. Nat Microbiol 2(10):1344–1349. https://doi.org/10.1038/s41564-017-0010-9
Lee J-H, Kaplan J, Lee W (2008) Microfluidic devices for studying growth and detachment of Staphylococcus epidermidis biofilms. Biomed Microdevices 10:489–498
Li Jeon N, Baskaran H, Dertinger SK, Whitesides GM, Van de Water L, Toner M (2002) Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in a microfabricated device. Nat Biotechnol 20:826–830
Lin CH, Hsiao YH, Chang HC, Yeh CF, He CK, Salm EM, Chen C, Chiu IM, Hsu CH (2015) A microfluidic dual-well device for high-throughput single-cell capture and culture. Lab Chip 15(14):2928–2938. https://doi.org/10.1039/c5lc00541h
Matu MN, Orinda GO, Njagi EN, Cohen CR, Bukusi EA (2010) In vitro inhibitory activity of human vaginal lactobacilli against pathogenic bacteria associated with bacterial vaginosis in Kenyan women. Anaerobe 16:210–215
Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA (2013) Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc 8:870–891
Mohan R, Sanpitakseree C, Desai AV et al (2015) A microfluidic approach to study the effect of bacterial interactions on antimicrobial susceptibility in polymicrobial cultures. RSC Adv 5:35211–35223
Moon TS, Lou C, Tamsir A, Stanton BCand Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. Nature 491:249–253
Niederholtmeyer H, Stepanova V, Maerkl SJ (2013) Implementation of cell-free biological networks at steady state. ProcNatlAcadSci U S A 110:15985–15990
Nissim L, Bar-Ziv RH (2010) A tunable dual-promoter integrator for targeting of cancer cells. Mol Syst Biol 6:444
Noireaux V, Bar-Ziv R, Libchaber A (2003) Principles of cell-free genetic circuit assembly. ProcNatlAcadSci U S A 100:12672–12677
O'Neill AT, Monteiro-Riviere NA, Walker GM (2008) Characterization of microfluidic human epidermal keratinocyte culture. Cytotechnology 56:197–207
Ochs CJ, Abate AR (2015) Rapid modulation of droplet composition with pincer microvalves. Lab Chip 15(1):52–56. https://doi.org/10.1039/c4lc01040j
Oliveira AF, Pelegati VB, Carvalho HF, Cesar CL, Bastos RG, de la Torre LG (2016) Cultivation of yeast in diffusion-based microfluidic device. BiochemEng J 105(Part A):288–295
Paik SM, Sim SJ, Jeon NL (2017) Microfluidic perfusion bioreactor for optimization of microalgal lipid productivity. BioresourTechnol 233:433–437. https://doi.org/10.1016/j.biortech.2017.02.050
Park J, Koito H, Li J, Han A (2009) A multi-compartment CNS neuron-glia Co-culture microfluidic platform. J Vis Exp 31:e1399
Park HS, Liu S, McDonald J, Thakor N, Yang IH (2013) Neuromuscular junction in a microfluidic device. ConfProc IEEE Eng Med BiolSoc 2013:2833–2835
Potvin-Trottier L, Luro S, Paulsson J (2018) Microfluidics and single-cell microscopy to study stochastic processes in bacteria. CurrOpin Microbiol 43:186–192. https://doi.org/10.1016/j.mib.2017.12.004

Prodanov L, Jindal R, Bale SS, Hegde M, McCarty WJ, Golberg I, Bhushan A, Yarmush ML, Usta OB (2016) Long-term maintenance of a microfluidic 3D human liver sinusoid. BiotechnolBioeng 113:241–246

Pronk GJ, Heister K, Vogel C, Babin D, Bachmann J, Ding GC et al (2017) Interaction of minerals, organic matter, and microorganisms during biogeochemical interface formation as shown by a series of artificial soil experiments. BiolFertil Soils 53:9–22

Pybus V, Onderdonk AB (1997) Evidence for a commensal, symbiotic relationship between *Gardnerella vaginalis* and *Prevotella bivia* involving ammonia: potential significance for bacterial vaginosis. J Infect Dis 175:406–413

Rusconi R, Lecuyer S, Guglielmini L, Stone HA (2010) Laminar flow around corners triggers the formation of biofilm streamers. J R Soc Interface 7:1293–1299

Shah P, Fritz JV, Glaab E, Desai MS, Greenhalgh K, Frachet A et al (2016) A microfluidics-based in vitro model of the gastrointestinal human microbe interface. Nat Commun 7:11535. https://doi.org/10.1038/ncomms11535

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. https://doi.org/10.1021/sb200016s

Skolimowski M, Nielsen MW, Emnéus J, Molin S, Taboryski R, Sternberg C, Dufva M, Geschke O (2010) Microfluidic dissolved oxygen gradient generator biochip as a useful tool in bacterial biofilm studies. Lab Chip 10:2162–2169

Sriram G, Alberti M, Dancik Y, Wu B, Wu R, Feng Z, Ramasamy S, Bigliardi PL, Bigliardi-Qi M, Wang Z (2018) Full-thickness human skin-on-chip with enhanced epidermal morphogenesis and barrier function. Mater Today 21:326–340

Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456(7221):516–519. https://doi.org/10.1038/nature07389

Super A, Jaccard N, Cardoso Marques MP, Macown RJ, Griffin LD, Veraitch FS, Szita N (2016) Real-time monitoring of specific oxygen uptake rates of embryonic stem cells in a microfluidic cell culture device. Biotechnol J 11(9):1179–1189. https://doi.org/10.1002/biot.201500479

Supramaniam P, Ces O, Salehi-Reyhani A (2019) Microfluidics for artificial life: techniques for bottom-up synthetic biology. Micromachines (Basel) 10(5):E299. https://doi.org/10.3390/mi10050299

Swanka Z, Laohakunakorna N, Maerkla SJ (2019) Cell-free gene-regulatory network engineering with synthetic transcription factors. Proc Natl Acad Sci U S A 116(13):5892–5901. https://doi.org/10.1073/pnas.1816591116

Taylor AM, Rhee SW, Tu CH, Cribbs DH, Cotman CW, Jeon NL (2003) Microfluidic multicompartment device for neuroscience research. Langmuir 19:1551–1556

van Swaay D (2013) Microfluidic methods for forming liposomes. Lab Chip 13:752–767

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

Wang BL, Ghaderi A, Zhou H, Agresti J, Weitz DA, Fink GR, Stephanopoulos G (2014) Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. Nat Biotechnol 32:473–478

Wufuer M, Lee G, Hur W, Jeon B, Kim BJ, Choi TH, Lee S (2016) Skin-on-a-chip model simulating inflammation, edema and drug-based treatment. Sci Rep 6:37471

Zhang Q, Lambert G, Liao D, Kim H, Robin K, Tung C-K, Pourmand N, Austin RH (2011) Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. Science 333:1764–1767

The Ethics of Synthetic Biology Research and Development: A Principlist Approach

20

Kevin Smith

Abstract

A principlist approach is adopted to analyse the ethical status of synthetic biology (synbio) research and development. The principle of nonmaleficence generates precaution-driven conclusions that are excessively restrictive to the field of synbio. The principle of beneficence is best served by permitting synbio research to flourish and not have it treated as a special case warranting the imposition of a high degree of external and self-regulation. Synbio may offend the principle of justice in certain circumstances; however, such issues are largely restricted to the initial stages of synbio innovation, whilst in the longer term the development of the field can be expected to promote just ends. The principle of respect for autonomy entails that scientists ought to be afforded a broad scope to freely pursue synbio research and development in a curiosity-driven fashion. In balancing the various conclusions under the four principles, the author concludes that society has an ethical obligation to support the development of synbio research and development and not restrict this important nascent field by the imposition of stern regulation.

Keywords

Autonomy · Beneficence · Bioethics · Justice · Nonmaleficence · Principlism · Synthetic biology

K. Smith (✉)
School of Applied Sciences, Abertay University, Dundee, UK
e-mail: k.smith@abertay.ac.uk

V. Singh (ed.), *Advances in Synthetic Biology*,
https://doi.org/10.1007/978-981-15-0081-7_20

20.1 Introduction

This chapter explores the question: what is the ethical status of research into synthetic biology (synbio) and its attendant development? Ethics is a domain where it is essentially impossible to achieve absolute truth. Ethical laws and practices are inevitably open to questioning, and the aphorism 'there are as many ethical positions as there are individuals' contains some truth. However, to hold that subjectivity reigns supreme in the ethics domain and that there are 'no correct nor wrong responses' would be a depressing position. Fortunately, whilst ethics can never yield the same levels of (provisional) near-certainty as can scientific research, rational ethical analysis can yield important conclusions. This is just as well, because ethics is crucial: there would inevitably be much societal, individual and environmental damage and exploitation in the absence of ethics-driven restraints; and, in the absence of ethical imperatives to act, many negative consequences would flow from omission.

To facilitate a rational exploration of the ethics of synbio research and development, this chapter will utilise a well-established approach to ethical analysis known as *principlism* (Beauchamp and Childress 1994). The principlist approach is based on four core ethical principles:

- Nonmaleficence
- Beneficence
- Justice
- Respect for autonomy

As a tool of ethical analysis, principlism entails that each of the four principles (in turn) is applied to the issues raised by an ethical conundrum—namely, synbio. Accordingly, this chapter will proceed by reviewing the ethical claims that have been made by various ethicists, with the material organised into each of the four categories listed above. Finally, an overall conclusion will be presented, in which the present author will seek to integrate and balance the individual conclusions under each of the four principles.

20.2 Nonmaleficence

The principle of nonmaleficence can be summarised as a normative rule: 'do no harm'. This notion is somewhat problematic on strictly logical grounds, as it would be virtually impossible to do anything in biomedical research that could be entirely guaranteed free from any conceivable risk of causing (even very slight) harm. A more common-sense interpretation of the concept is as follows: it should be considered unethical to do something that has a high risk of causing serious harm. Whilst this latter interpretation lacks precision (when is a risk 'high' enough to count; when is a harm 'serious' enough to count?), it has resonance with the kind

of everyday ethical outlook commonly employed by individuals—including many scientists and laypeople.

Whilst promising major benefits, it has been asserted that synbio has the potential for catastrophic consequences, including some which are far more severe—and unpredictable—than most associated with most forms of bioscience and some which may be unstoppable once underway. These disastrous potential outcomes include runaway environmental degradation and abuse of the products of synbio for malign ends. For example, the prominent bioethicist Savulescu, Director of the Oxford Uehiro Centre for Practical Ethics, has claimed that synbio 'introduces new jokers into the pack' so that amongst other dystopian scenarios, 'one mistake or abuse could be catastrophic', as Savulescu alleges that synbio carries 'unprecedented risk to natural ecosystems and human health' (Savulescu 2012). On a similar vein, Stabinski of Greenpeace International has claimed that '. . . synthetic biology is like genetic engineering on steroids'; 'Tinkering with living organisms that could be released in the environment poses a grave biosafety threat to people and the planet' (Stabinski 2006). Such siren calls have been heeded and, dating from the earliest days of synbio research, these concerns have led to various national regulatory responses to control synbio research and development (Rager-Zisman 2012).

Does synbio research contain intrinsic features that render its proscription or heavy regulation ethically desirable? The notion synbio represents a special case in bioethics is highly contentious (Newson 2015). Synthetic organisms may be designed to reproduce autonomously, and major concerns have been raised as to the biosafety aspects of such organisms. What if they would escape from the laboratory? The environmental and health-related negative consequences of such escape could be considerable. However, numerous biomedical laboratories (most of which are not specifically involved in synbio research) routinely deal with highly pathogenic organisms, including potentially lethal viruses and bacteria, as well as genetically modified organisms. In such work, standard safety procedures are successfully employed to prevent accidental release of the organisms. Thus, it would appear to be a fundamental error to place synbio in a special category in which the normal safety approaches of biomedical laboratories are deemed insufficient.

The concept of 'dual use' (or 'dual-use complexity') has been particularly prominent in the ethical discourse surrounding synbio. The concept refers to research aimed at improving human wellbeing having the potential for misappropriation by bioterrorists. A fairly substantive body of published work exists that explores this notion (Drew and Mueller-Doblies 2017; Evans and Selgelid 2015; Rager-Zisman 2012; Wimmer 2018). The published work on dual use, taken in toto, has been broadly negative in its conclusions (though a minority of authors take a contrary view) and has in general tended to suggest that the risk of harm or disaster warrants strong regulation or proscription of much synbio research, or at least placing obligations (sometimes of an onerous nature) on synbio researchers to become aware of possible harmful uses of their work and thence take responsibility for averting the anticipated harmful misuses.

However, a problem with the concept of dual use is that, taken literally, it appears to be so widely applicable that it offers little of value. Almost all technologies

could in some way be misused such as to cause harm; and all forms of synbio research have at least some potential for dual use. Thus, the concept of dual use can logically only apply in cases where the risks are high and the dangers are of a very serious sort. Such experiments, where the scope for malign uses of a given technology is clear, have been termed 'experiments of concern' (Miller and Selgelid 2007). In such cases, the dual-use concept appears undisputable and the ethical disapprobation of pursuing such research without due safeguards inevitable and irrefutable. Nevertheless, in such clear-cut cases, the researchers involved are likely to recognise the potential for misuse and act accordingly—with no need for complex ethical theorising around concepts of dual use. (This assumes they are not corrupt: but ethical injunctions will not prevent corrupt individuals from carrying out their immoral activities.) In terms of regulatory control or proscription, it is not clear that this is necessary at the research phase of development of (say) a synthetic microorganism for medical use that could also be put to use as a weapon by bioterrorists. External regulation would be appropriate at a future developmental stage—i.e. when a clearly dangerous product (such as the synthetic microorganism in the above example) is ready for use, as opposed to when the initial research is being conducted. Thus, in cases where the dangers of misuse are obvious and realistic, the concept of dual-use ethics as a normative tool would appear to be somewhat redundant.

Yet it is clear that many advocates of dual-use ethics would not restrict the approach to scenarios in which the risks of misapplication of synbio were credible and severe, as in the foregoing example. The literature on dual-use ethics tends to suggest that a dual-use ethics approach should apply universally to the field. Thus, synbio researchers would be obliged to be permanently vigilant in respect of potential misuses of all aspects of their work. In general, proponents of dual-use ethics suggest that *education* should be the tool for engagement with dual-use concerns. Dual-use proponents have argued that research communities ought to provide substantive educational provision for researchers—preferably incorporating the involvement of laypeople—aimed at raising awareness of the possibilities for technological misuse. Sadly, professional ethics education more often than not falls far short of its well-meant goals for enhancing ethical behaviour (Bazeman and Tenbrunsel 2011), and it seems almost inevitable that efforts to inculcate dual-use ethics amongst researchers will be similarly futile. One might argue that such education, although of minimal benefit, is at least harmless. But this assumption is questionable: the costs of dual-use education could be substantial. An obligation for such education would be time consuming and would tend to distract researchers from their primary work, thus impacting negatively upon scientific progress. Moreover, because synbio is multidisciplinary and dual-use ethics entails that any activity of potential relevance to synbio ought to be considered for its abuse potential, these costs would potentially be shouldered by an extensive range of bioscientific disciplines and a very large number of researchers.

It has been claimed that ethical attention should be paid not only to the research processes underlying synbio but also to *knowledge* itself: namely, what knowledge should be pursued and how much of it ought to be disseminated (Douglas and

Savulescu 2010). The underlying contention is that it is desirable to control and limit synbio knowledge in order to avert bad outcomes accruing from the misuse of such knowledge by bioterrorists or other malign actors.

The argument for widening ethical consideration to include knowledge has prima facie appeal. However, it carries substantial danger of unintended consequences. It is indisputable that the application of bioscientific knowledge to various domains, such as agriculture and medicine, has led to major gains for humanity. Hence, it is dangerous—and ethically fraught—to restrict bioscientific enquiry, lest the flow of improvements to human wellbeing be impeded. Whilst authors favouring such restrictions on knowledge want to limit such control exclusively to synbio-associated knowledge with the potential for serious misuse, in practice this position is highly problematic. It is generally impossible to discriminate a priori those forms of research and knowledge so produced that will lead to catastrophe versus those that are either neutral or will lead to improvements in human wellbeing; additionally, it is necessary to ask the question: *who decides* which types of knowledge (and underlying research) ought to be prohibited? To which there are no uncontentious or unproblematic answers (Pierce 2012). Bioscientific research may reasonably be viewed as the process by which reliable knowledge about biology is accumulated and deposited in the public domain. Some of this knowledge will then be used for technological ends, thus improving human wellbeing. On this basis, it appears ethically correct to proceed on the default assumption that any given piece of synbio knowledge, and the research entailed in its discovery, ought not to be restricted. Of course, it would be ethically required to prohibit dissemination of truly dangerous knowledge (for instance, instructions for the construction of a synthetic microorganism capable of being unleashed as a potent bioweapon). However, this type of knowledge is an end product: by contrast, most research knowledge is not a product; its potential future uses—good or ill—cannot be known with any accuracy, and attempts to restrict such knowledge would be impractical and damaging to the bioscientific enterprise. To employ a non-biological analogy, from physics: whilst it would be ethically necessary to prevent public dissemination of a detailed set of instructions for building a 'dirty' nuclear bomb, it would be ethically problematic to prevent public access to the fundamental physics knowledge underpinning such a weapon. Attempts to restrict generation of (through research) or public access to such basic physics knowledge would not only be manifestly impractical and highly intrusive for the research communities involved, but also such restrictions would impede development of valuable future technologies (such as medical devices based on material physics knowledge, etc.).

Aside from the spectre of bioterrorism, it is a truism that accidents happen, and fears about extremely undesirable or catastrophic consequences from mishaps involving synbio have led to some opponents of technological progress to posit the theoretical risks from synbio—particularly unidentified risks—as grounds for a complete ban on synbio research and development. More subtly, there have been declarations from many in society calling for the use of a *precautionary principle* to regulate the field. This principle holds that until all possible risk has been identified and eliminated, one ought not to proceed. Originally devised in

the context of environmentalism, the precautionary principle has been applied to many types of scientific endeavour including synbio. A precautionary approach is intuitively appealing to many, as it chimes with the tendency towards caution innate in the human psyche. The precautionary principle, however, means unacceptably adverse or ridiculous implications, at least in the case of synbio. Taken literally, it maintains *any* degree of danger unacceptable *regardless* of the future advantages. Unfortunately, without some danger, it is impossible to do anything innovative in biomedicine, or to conduct any research that could ultimately produce some undesirable results, however remote or minor. For instance, owing to unexpected impacts of engineered virus-based agents used in clinical studies for somatic gene therapy, a (small) number of patients have died or become seriously ill. If the strong precautionary principle had been implemented in the field of somatic gene therapy, accidental deaths and morbidity would have been prevented, but this would have been to the detriment of the increasing number of patients who have so far benefited from somatic gene therapy and would have excluded benefits for potential patients who stand to be successfully treated in increasing numbers as the technology matures. Therefore, if implemented, the precautionary principle in its strong form would allow nothing to be accomplished at all!

In response to this problem, alternative or attenuated variants of the precautionary principle have frequently been posited and are of relevance to the ethics of synbio. The ways in which the principle has been interpreted are many and include a means to recalibrate the onus of proof, a means of focusing on certain types of risks and a rule-of-thumb reminder to 'err on the side of caution' (Smith et al. 2012). For example, in the specific context of synbio, Wareham and colleagues have outlined a systematic approach for arriving at a threshold below which potential dangers can be disregarded (Wareham and Nardini 2015). Their approach includes a Bayesian tool to assign probabilities to deleterious outcomes. These authors argue that their proposed approach effectively rehabilitates the precautionary principle as a feasible policy tool to deal with the unknowns of future synbio research.

This approach may be of practical worth; however, it is intellectually difficult to defend the concept of a threshold. The problem lies in the essentially arbitrary nature of a danger threshold. More broadly, the use of any attenuated form of precautionary principle will tend to contribute irrational considerations—such as an obligation to give arbitrary weight to certain kinds of danger (such as unknown dangers or environmental dangers)—that will tend to skew an otherwise rational analysis of risk.

In conclusion, applying the principle of nonmaleficence to synbio research and development is problematic if the principle is interpreted in a literal manner. Whilst harm is certainly possible from the abuse of synbio knowledge or products, an ethic that literally demanded 'do no harm' would entail undue restriction to the field and would thus slow down, limit or prevent the developments of applications of synbio that would otherwise deliver major benefits.

20.3 Beneficence

The principle of beneficence can be summarised as an ethical onus to create as much good as possible. This naturally begs the question, what is 'good'? Whilst there are of course a wide range of claims of what is good, made by philosophers over the centuries, in the context of principlism, the good refers to *utility*—namely, the existence of positive mental states (or, more simply, 'happiness'), in aggregate across all affected individuals. Thus, beneficence is a *consequentialist* ethical principle, in that it requires evaluation of the consequences of actions. Actions having the best consequences—i.e. those producing the greatest quantity of utility—are ethically praiseworthy. This approach is also known as *utilitarianism*.

Claims of good consequences do not need to be explicitly linked to the maximisation of utility in order to fall within the scope of the principle of beneficence. For example, suppose a plausible claim is made that a particular synbio development is likely to enable pharmaceutical scientists to produce a new, highly effective family of antibiotic drug. Whilst it would be possible and valid to morally praise the associated synbio work on the explicit grounds that the outcome would likely be an increase in utility, through boosting happiness and reducing suffering as a result of saving lives and minimising the duration of infectious illness, such description is somewhat superfluous. Rather, it would be sufficient simply to establish that the synbio work would lead to the ability to create such drugs in order for the activity to count as ethically desirable under the beneficence principle.

The possible practical applications of synbio, and their effectiveness, cannot be foreseen with any accuracy. This is a general issue for any nascent technology: there are plentiful instances of this in the history of science and technology. For example, at the time that the ability to artificially transfer genes from one organism to another became an experimental possibility (in the 1970s), it could not be accurately predicted that this basic work would lead to the major applications that have—and continue—to revolutionise key aspects of medicine. For example, we have seen a follow-on revolution in the production of vital therapeutic agents such as insulin, together with successful gene therapy attempts and, most recently, gene-edited human cells for successful anticancer therapy. To take another example, early experiments with bacteria eventually led to the discovery of penicillin and thus to the development of the powerful antibiotics that have become invaluable to modern medicine. It is clear that such benefits could not have been anticipated by the pioneers of bacteriology working on basic research questions back in the nineteenth century.

Synbio is clearly at a relatively early stage of development as a field; for example, only fairly recently (2010) was the production of the first semi-synthetic organism achieved (Gibson et al. 2010). It is entirely possible that in the future synbio could allow the generation of artificial cells of great medical importance. Such potential synbio products might allow the production of synthetic microorganisms designed to exterminate cancer cells within the patient's body, or artificial stem cells able to replace cells lost to disease in neurodegenerative conditions. However, at

present there is no way of predicting with any accuracy when or whether which of these or similar valuable products may be obtained from synbio (Heavey 2017). Nevertheless, simply because it is not possible to predict the exact benefits of synbio certainly does not imply that major benefits—and the beneficence that will flow from them—ought not to be factored into the evaluation of the value of synbio. Innumerable examples of the accretion of fundamental knowledge leading to technologies that substantially benefit humankind can be discerned in the history of science. Thus, fundamental scientific research tends to lead to major gains in utility; and this is a feature of the scientific enterprise that can be applied across all major fields. Accordingly, to the extent that synbio research is permitted to develop in an unrestricted manner, ultimately much utility will flow from such endeavours, as the products of synbio emerge and help inter alia to combat disease and solve other human problems. Elsewhere in this book, other authors describe current research in the field of synbio, as well as allude to nascent products, the development of which is of course based on basic research knowledge. All this counts as positive according to the principle of beneficence and as such lends strong ethical support to research scientists working in the field of synbio.

In conclusion, the principle of beneficence places an ethical duty on society to support synbio research and to refrain from unduly impeding its progress through excessive caution and unduly onerous regulation. A rational approach would seek to balance the possible harms of synbio (as described in Sect. 20.2) with the potential benefits.

20.4 Justice

The principle of beneficence can be summarised as the ethical duty to fairly distribute benefits, risks and costs. For example, if a form of synbio were to benefit only one part of society, particularly one comprising privileged individuals, this would be considered unjust and therefore ethically problematic.

Justice issues are frequently raised as objections to scientific and technological developments, including synbio. An important issue in terms of the principle of justice arises where governance of synbio imposes onerous and protracted regulatory requirements on synbio: this leads to a situation in which only rich countries or multinational corporations can participate (Schmidt 2016). This threatens to limit the potential benefits from synbio for poorer nations. A parallel concern exists in terms of harms: if a synbio product were to harm a specific societal group and not other sectors of society, this would be unjust (ETC Group 2007). Indeed, popular opinion holds that social justice concerns provide reason to proscribe the investigation and development of radical new technologies such as (at least some forms of) synbio, a societal perspective that is to a large extent mirrored by government agencies and policy forums (Kaebnick et al. 2014).

However, such concerns tend to focus on the initial stages of the introduction of the technology and neglect the more important long-term patterns of distribution (Hunter 2013). An analogy with mobile phones is apposite to illustrate this

tendency: initially, research into mobile telephony was restricted to laboratories in rich countries and was frequently conducted by moneyed Western corporations; and when the first products were marketed, only rich individuals and prosperous companies could afford them. Now, we see mobile phones permeating all countries, rich and poor, and in the latter nations they are a huge force for good, permitting impoverished peoples to communicate and access information to their marked benefit. Yet a strong justice-based ethical objection to mobile telephony research could have been mounted, and had it been heeded two negative consequences would have arisen: (a) the technology would have been stunted in its development (or not developed at all); and therefore (b) impoverished people would not have had the benefits of mobile phones—which would have impacted their lives to a much greater degree than wealthy people.

From this analogy, a general ethical rule may be discerned: justice-based concerns imply that restrictions ought to be placed on the investigation of new technologies, yet these restrictions lead to reduced justice in the long run. This paradoxical situation clearly applies to synbio research, as much as it does to the investigation of any radical new technologies.

In conclusion, whilst the application of future synbio technologies needs to take social justice into account, the principle of justice is best served by permitting synbio research and development to proceed without stern regulatory impositions and restrictions.

20.5 Respect for Autonomy

The principle of respect for autonomy holds that a competent, informed person ought to have the freedom to choose whether or not to participate in, or consent to be part of, a proposed course of action. Any violation of such freedom is, under the respect for autonomy principle, unethical.

The autonomy of synbio researchers counts in this regard. Those who wish to exercise their freedom to explore this field of science are prima facie morally entitled to do so, according to the autonomy principle. Of course, any demonstrably negative impacts of their work, of the sort described previously in this chapter, would serve as good reason to restrict the autonomy of synbio researchers. However, as argued above, whilst applications of synbio may be misused by malign actors and accidents may occur, stern restrictions to synbio at the level of research are generally not warranted on ethical grounds. Similarly, concerns over autonomy do not morally justify proscription of synbio research and development. Thus, insofar as synbio does not amount to a net negative in the foregoing terms, synbio researchers ought to be free to continue their endeavours.

Social science research has indicated that where patients are presented with the idea of being treated with synbio-based medical technologies, they frequently display excessive fear concerning the technology, for instance worrying unnecessarily that 'something uncontrollable' could be created. It is ethically necessary that such patients be adequately informed about the actual risks and benefits of the synbio

application in question if they are to be able to exercise their true autonomy. Under the principle of autonomy, it is ethically important that patients be accorded the information to allow them not only to decline treatment if they so desire but also the 'right to try' (Rakic et al. 2017; Heidari et al. 2017).

Another autonomy-related concern arising from synbio is the notion that the technology could lead to fundamental changes in our understanding of life and what it means to be human—particularly if developments in synbio proceed in parallel with developments in robotics (Nasuto and Hayashi 2016; Murillo-Sanchez and Ruiz-Mirazo 2016). However, such concerns are clearly of a very distant-future nature; to some extent they are at the cusp of what can reasonably be evaluated within the methodology of principlism and begin to enter into the realms of metaphysics. As such, these claims and concepts lie beyond the scope of the present chapter.

20.6 Conclusions

There are many ways to look at the ethical issues raised by radical new technologies. This chapter has considered the issues raised by synbio through the normative lens of principlism. In terms of the first principle addressed above, nonmaleficence, many ethicists and other critics have raised fears about harms—some of them catastrophic—arising from the misuse of synbio, or accidents associated with it. The common (although by no means universal) view amongst such commentators has been that the risks require a precautionary approach, entailing at least heavy regulation, if not proscription, of much synbio research and development. However, adherence to the implications of the nonmaleficence principle is generative of problematic outcomes: primarily a hindering of the development of synbio products. Because such products will include tools that have the potential to deliver substantial human improvement in the future, the principle of nonmaleficence is fraught in the context of synbio, since its application would lead to less utility (human happiness) than had it not been invoked.

In this regard, the principle is in direct conflict with the second principle addressed above, beneficence. Whilst the commentariat, including ethicists (with some notable exceptions), have generally been unwilling to emphasise the upsides of synbio, preferring instead to concentrate on the potential risks or other negatives, this chapter has argued that these upsides are very substantial for the future. This implies that a resolution to the tension between these two principles, of nonmaleficence versus beneficence, must depend upon a rational cost-versus-benefit evaluation. Such an analysis is outside the scope of the present chapter: it would be appropriate only for specific aspects or instances of synbio research and development, as opposed to having applicability to the overall question of the ethical status of synbio. Nevertheless, it is clear that the potential benefits of synbio are very large, albeit that their exact shape cannot be discerned with much clarity at the present point in time. This means that a maleficence-driven focus on more clearly identifiable dangers can easily, but irrationally, serve to downplay the beneficence-

promoting aspects of synbio. Readers must make up their own minds as to which direction the balance lies—this is, after all, an ethical question, requiring judgement of a sort different to scientific analysis. However, the present author suggests that the balance is in favour of synbio being permitted to be investigated and developed without special regulation or proscription beyond the normal standards of safety that apply to bioscientific research in general.

The third principle addressed above, justice, is intrinsically harder to evaluate than either of the two preceding principles, in any given context such as synbio research and development. The notion of justice is commonly held to be axiomatically valid and important as an ethical principle, and all else being equal, it is difficult to argue against this principle. However, in certain contexts, including synbio, justice can be in tension with the first two principles. This is so because a strict application of the principle would, as discussed above, frequently mean that scientific progress and the attendant delivery of products that boost human welfare will be stymied. How is this tension resolved? To a large extent this depends on the basic ethical outlook of the individual who is judging: someone who holds justice to be the most important principle will not be swayed by considerations arising from other principles, such as greater utility. By contrast, someone who holds beneficence to be the most important principle will only support appeals to justice insofar as maximising justice also yields greater human happiness; where the two principles conflict, this individual will prefer an unjust situation with greater utility to the converse. Philosophers continue to argue about the relative merits and demerits of particular ethical principles (an academic field known as meta-ethics), but this debate lies beyond the scope of the present chapter, and the reality is that most individual's come to the ethics table with pre-established ideas which principle(s) they hold dearest. The present author considers the principle of justice to be subordinate to that of beneficence and accordingly is not swayed by appeals to justice where the outcome would be a reduction in utility.

Aside from far-future possibilities that might arise from synbio, involving synthetically generated people, the principle of respect for autonomy is taxed less by questions around synbio than the other three principles. The greatest issue that synbio (presently) raises in respect of autonomy resides in the domain of the freedom of scientists to conduct their research in an open-ended, curiosity-driven manner. Those who hold autonomy to be the most important principle will agree that it is intrinsically right to accord such autonomy to individual researchers. Others will be happy to accede to the notion that such freedom be accorded, to the extent that it does not impact negatively in respect of the other principles. So, the present author supports synbio researcher freedom, not primarily because it is an intrinsic good, but rather because it aids the synbio enterprise and thus will help deliver benefits in terms of enhanced human happiness.

In overall conclusion: whilst synbio research and development may lead to knowledge and products that could cause harm accidentally or by deliberate misuse and may impact negatively in terms of societal justice in the short term, the promise of this emergent arena of bioscientific progress offers so much potential

that researchers in the field ought to be supported by society in their endeavours. To do otherwise would be ethically unacceptable.

References

Bazeman H, Tenbrunsel AE (2011) Blind spots: why we fail to do what's right and what to do about it. Princeton University Press, Princeton

Beauchamp TL, Childress JF (1994) Principles of biomedical ethics. Oxford University Press, New York

Douglas T, Savulescu J (2010) Synthetic biology and the ethics of knowledge. J Med Ethics 36(11):687–693

Drew TW, Mueller-Doblies UU (2017) Dual use issues in research - a subject of increasing concern? Vaccine 35(44):5990–5994

ETC Group (2007) New directions: the ethics of synthetic biology and emerging technologies report. Presidential Commission for the Study of Bioethical Issues, Washington

Evans NG, Selgelid MJ (2015) Biosecurity and open-source biology: the promise and peril of distributed synthetic biological technologies. Sci Eng Ethics 21(4):1065–1083

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R, 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 Z, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56

Heavey P (2017) Consequentialism and the synthetic biology problem. Camb Q Healthc Ethics 26(2):206–229

Heidari R, Shaw DM, Elger BS (2017) CRISPR and the rebirth of synthetic biology. Sci Eng Ethics 23(2):351–363

Hunter D (2013) How to object to radically new technologies on the basis of justice: the case of synthetic biology. Bioethics 27(8):426–434

Kaebnick GE, Gusmano MK, Murray TH (2014) The ethics of synthetic biology: next steps and prior questions. Hastings Cent Rep 44:S4–S26

Miller S, Selgelid MJ (2007) Ethical and philosophical consideration of the dual-use dilemma in the biological sciences. Sci Eng Ethics 13(4):523–580

Murillo-Sanchez S, Ruiz-Mirazo K (2016) Synthetic biology as a challenge to understand the autonomy of the living. Isegoria 55:551–575

Nasuto SJ, Hayashi Y (2016) Anticipation: beyond synthetic biology and cognitive robotics. Biosystems 148:22–31

Newson AJ (2015) Synthetic biology: ethics, exceptionalism and expectations. Macquarie Law J 15:45–58

Pierce RL (2012) Whose ethics of knowledge? Taking the next step in evaluating knowledge in synthetic biology: a response to Douglas and Savulescu. J Med Ethics 38(10):636–638

Rager-Zisman B (2012) Ethical and regulatory challenges posed by synthetic biology. Perspect Biol Med 55(4):590–606

Rakic M, Wienand I, Shaw D, Nast R, Elger BS (2017) Autonomy and fear of synthetic biology: how can patients' autonomy be enhanced in the field of synthetic biology? A qualitative study with stable patients. Sci Eng Ethics 23(2):375–388

Savulescu J (2012) Master the new loom before life's tapestry unravels at our hands. Times High Educ 2046:32–33

Schmidt JC (2016) Prospective technology assessment of synthetic biology: fundamental and propaedeutic reflections in order to enable an early assessment. Sci Eng Ethics 22(4):1151–1170

Smith KR, Chan S, Harris J (2012) Human germline genetic modification: scientific and bioethical perspectives. Arch Med Res 43(7):491–513

Stabinski D (2006) Synthetic biology - global societal review urgent! http://www.etcgroup.org/node/8

Wareham C, Nardini C (2015) Policy on synthetic biology: deliberation, probability, and the precautionary paradox. Bioethics 29(2):118–125

Wimmer E (2018) Synthetic biology, dual use research, and possibilities for control. In: Defence against bioterrorism: methods for prevention and control. Springer, Dordrecht, pp 7–11