The University Of Sheffield.

EXPANDING MOLECULAR TOOLBOX FOR SYNTHETIC BIOLOGY

Thesis submitted for the degree of Doctor of Philosophy

Pawel Jajesniak

Department of Chemical and Biological Engineering

University of Sheffield

September 2017

DECLARATION

This thesis and the work presented in it are my own and has been generated by me as the result of my own original research. Where other sources of information have been used, they have been duly acknowledged.

No part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution. Parts of this work have been, or will be, presented elsewhere - e.g., in the form of scientific publications.

This dissertation conforms to the rules of Alternative Format Thesis, as described by the Code Of Practice For Research Degree Programmes 2017–18 of the University of Sheffield. In accordance with the underlined guidelines, this thesis contains sections that are written in the form of scientific publications, some of which have already been published in peer-reviewed journals and other outlets for academic research. Whenever such materials are used, their sources are clearly acknowledged.

ASSOCIATED PUBLICATIONS

(1) **Jajesniak P**, Wong TS. (2015) Tapping into biodiversity: From metagenomics to industrial enzymes. In: Biodiversity and Conservation (Vol. 2). USA: Studium Press LLC.

(2) **Jajesniak P**, Wong TS. (2015) From genetic circuits to industrial-scale biomanufacturing: bacterial promoters as a cornerstone of biotechnology. AIMS Bioengineering, 2(3), 277-296.

(3) **Jajesniak P**, Wong TS. (2015) QuickStep-Cloning: a sequence-independent, ligation-free method for rapid construction of recombinant plasmids. Journal of Biological Engineering, 9, 15.

(4) **Jajesniak P**, Wong TS. (2017) Rapid Construction of Recombinant Plasmids by QuickStep-Cloning. In: Synthetic DNA. Methods in Molecular Biology (Vol. 1472). USA: Humana Press, New York, NY; 205-214.

ACKNOWLEDGEMENTS

A pursuit of PhD degree, while seemingly a solitary endeavour, is never conducted in isolation. Its successful completion depends heavily upon the goodwill and all-encompassing support – academic, technical, financial, emotional - of many people whose role and generosity simply cannot be overstated.

I would like to sincerely thank my supervisor, Dr Tuck Seng Wong, for his invaluable guidance during the course of my degree, in particular, for always keeping his door open and being willing to help.

I am grateful to the Department of Chemical and Biological Engineering, the funder of my project, for providing me with an opportunity of pursuing the PhD degree at the University of Sheffield.

I would like to express my gratitude to Dr Kang Lan Tee for her technical support, fruitful scientific discussions and invaluable help in refining my experimental techniques.

I wish to give special thanks to my friends, and fellow PhD students - Hossam, Yomi, Miriam, Zaki, Abdul, Inas and José - for making the past four years, including even the most daunting laboratory sessions, truly enjoyable.

Last but certainly not least, I am extremely grateful to my whole family for their unceasing and unconditional support.

SUMMARY

Synthetic biology can broadly be defined as a construction of new biological systems and modification of the existing ones. In recent years, synthetic biology has presented itself as a potential solution to many global challenges - e.g., climate change and scarcity of natural resources. The progress of synthetic biology is, however, largely reliant on the availability of a large repertoire of functional genetic parts. As a result, there exists an urgent need for enhancing and expediting the process of tapping into genetic diversity - both natural and laboratory-induced.

Analysing the current landscape of synthetic biology and various approaches of accessing biological diversity, two high-impact areas, namely the basic molecular cloning and identification of genes encoding for complex microbial phenotypes, have been identified as particularly lacking and, as such, constitute the focal point of my PhD project. The culmination of the work presented in this dissertation are two molecular methods, QuickStep-Cloning and Multi-Genius, that aim to accelerate the development of synthetic biology.

QuickStep-Cloning is a new molecular cloning technique that builds upon recent advances in megaprimer-based cloning to allow for seamless integration of a DNA fragment of interest into a plasmid in less than 6 hours – the result that could not be reproduced using state-of-the-art methods. The new improved version of the method, QuickStepS-Cloning, utilises phosphorothioate oligonucleotides to not only simplify the overall procedure but also significantly increase its cloning efficiency. It also shown that

incorporating random mutagenesis into the method allows for streamlining directed evolution experiments.

Whereas the potential applications of QuickStep-Cloning revolve around artificially-induced diversity, Multi-Genius builds upon the concept of genomic libraries to tap into naturally-existing diversity and expedite identification of genes encoding for useful phenotypes. The usefulness of the method has been proven by isolating thermotolerant variants of *Escherichia coli* DH5α and indentifying the gene responsible for the observed phenotype.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ABBREVIATIONS

ABI-REC	asymmetric bridge PCR with intramolecular homologous recombination
Amp	ampicillin
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
CS1	cloning site 1
CS2	cloning site 2
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EGFP	enhanced green fluorescent protein
EMP	exponential megaprimer PCR
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
IAA	3-β-indoleacrylic acid
IFPC	inverse fusion PCR cloning
LB	lysogeny broth
LIC	Ligation Independent Cloning
iGEM	International Genetically Engineered Machine
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
MSC	multiple cloning site
PCR	polymerase chain reaction

RAM	recombination-assisted megaprimer
rDNA	ribosomal DNA
RNA	ribonucleic acid
RPU	Relative Promoter Unit
RF	restriction-free
RFP	red fluorescent protein
SDS	sodium dodecyl sulfate
ssDNA	Single-stranded DNA
UV	ultraviolet

SYMBOLS

Amp^r	containing ampicillin resistance gene
bp	base pairs
°C	degree Celsius
$	dollar
g	gram
Gb	gigabase
h	hour
Kan^r	containing kanamycin resistance gene
kg	kilogram
kb	kilobase
µl	microlitre
Mb	megabase
µM	micromolar
ml	millilitre

mM	milimolar
min	minute
nm	nonometer
nt	nucleotide
%	percent
s	second
U	enzyme unit
V	volt

Chapter 1
General introduction

1.1 Brief introduction to synthetic biology

Synthetic biology is an emerging area of research that encompasses many traditional branches of molecular biology and biotechnology, and strives to develop novel biological systems and devices. There exist several working definitions of synthetic biology. According to the Royal Society, synthetic biology can be described as *the design and construction of novel artificial biological pathways, organisms or devices, or the redesign of existing natural biological systems* (The Royal Society, 2007). At its inception, the emphasis of synthetic biology centred around the concepts of *de novo* design and construction of biological systems from standardised biological components, analogous to how computers and other sophisticated electronic systems are built from transistors and other basic electronic components. In order to achieve this aim, synthetic biologists try to apply engineering principles – e.g., abstraction and standardisation - to biological systems.

A primary example of the underlined approach is Registry of Standard Biological Parts, a large repository of a wide range of biological components, such as promoters, ribosome binding sites and reporter proteins. The available genetic parts conform to BioBrick assembly standard (Knight, 2003), allowing to combine them in a relatively simple and time-efficient manner. Furthermore, the functionality of many genetic parts is well characterised to make sure that selected components are fit for purpose - e.g., selected promoter allows for a desired expression level. As such, the Registry of Standard Biological Parts constitutes one of the first attempts of applying the concepts of modularity and standardisation, fundamental engineering principles, to biological

systems. It is worth noting that the repository is used every year by undergraduate students from around the world to design and build functional biological devices, as part of the International Genetically Engineered Machine competition, iGEM.

As synthetic biology is still in a phase of rapid development, many of the reported breakthroughs revolve around proof-of-concept research or development of much-needed enabling tools. Well-known examples of this approach include genetic circuits (Nandagopal and Elowitz, 2011; Siuti et al., 2013) and *Synthia* (Gibson et al., 2010), the first living organism based on purely synthetic chromosomal DNA, developed by researchers at the J. Craig Venter Institute.

At the same time, an increasing number of synthetic biologists demonstrate the usefulness of the available molecular tools by attempting to tackle various global challenges. One of such issues has been the production of artemisinin, a common antimalarial drug - recommended for the treatment of malaria caused by a protozoan parasite, *Plasmodium falciparum*. Artemisinin originates from the plant *Artemisia annua* (known also as sweet wormwood) and its discovery was a focal point of Nobel Prize in Physiology or Medicine in 2015 won by Tu Youyou (Guo, 2016). As artemisinin needs to be isolated from *A. annua*, its availability and cost is heavily dependent on the erratic supply of the plant (Van Noorden, 2010). To address this problem, Jay Keasling and his team from the University of California, Berkley, utilised a synthetic biology approach to incorporate the metabolic pathway responsible for synthesis of an artemisinin precursor into yeast (Ro et al., 2006), providing an alternative way of acquiring the drug. The method has been commercialised and used for industrial production of semi-synthetic

artemisinin by the French pharmaceutical company Sanofi (Paddon and Keasling, 2014). Despite the fact that, due to complex socio-economic factors, artemisinin remains to be produced primarily from the plant-based material (Peplow, 2016), this story clearly demonstrates the immense potential of synthetic biology in tackling global challenges.

If achieved, the ability to freely design and engineer new, functional biological systems and devices would, undoubtedly, bring countless benefits. Production of pharmaceuticals is only one of the many examples of potential applications of synthetic biology, such as biosensor-based devices, productions of biofuels and removal of toxic contaminants from the environment (Choffnes et al., 2011; Konig et al., 2013). Many of the most pressing global challenges – e.g., scarcity of natural resources, environmental degradation, climate change and infectious diseases – could potentially be solved by judicious application of such technology. Unfortunately, experience shows that there exist detrimental differences between living organisms and electronic systems that make the concept of *de novo* design difficult to achieve in practice. Intrinsic biological complexity, a sheer number of functional components, complex, nonlinear interactions, constant evolution of biological organisms and spontaneous variations in the behaviour of biological systems – these are only some of the challenges that synthetic biologists have to cope with, in order to construct fit-for-purpose biological devices from standardised biological components. These challenges have been succinctly described in 2010 article written by Roberta Kwok titled 'Five hard truths for synthetic biology' (Kwok, 2010). The titular 'hard truths' are the main challenges that, according to the

author, constitute the biggest hindrance to the development of synthetic biology – poorly defined biological parts, the unpredictability of biological circuitry, the intrinsic complexity, incompatibility of biological components and ubiquitous variability of biological systems. As such, the progress of synthetic biology is largely reliant on the constant supply of well-defined biological parts that can be combined together, in a predictable fashion, to create functional devices. It is also important to note that the main factor behind the discussed problem of biological complexity is the need of constructing and testing a large amount of biological circuits in order to find one that fulfils all the necessary design criteria. As such, any technique that expedite such process could be of great benefit to the field of synthetic biology.

The distinctive nature of biological systems, while being a major hindrance to applying engineering strategies to biology, provides unique opportunities if properly utilised. Natural biological diversity constitutes a nearly unlimited source of biological components. Methods such as metagenomics attempt to tap into this largely unexploited potential. Whereas constant evolution of biological systems interferes with the construction of well-defined biological devices, strategies such as directed evolution make use of evolution to improve, or even develop, useful properties of proteins by further expanding the existing biological diversity. In the same way, adaptive evolution is routinely used to improve, for instance, industrially-relevant bacterial strains.

By tapping into biodiversity - both natural and laboratory-induced, those methods perfectly complement the traditional synthetic biology approach, by not only providing

an alternative way of engineering biological systems but also they constitute an invaluable source of new genetic parts.

1.2 Natural biodiversity

This section reviews different approaches to accessing natural biological diversity and provides a comprehensive overview of metagenomics in the context of novel biomolecule discovery. It has been adapted from:

***Jajesniak P**, Wong TS. (2015) Tapping into biodiversity: From metagenomics to industrial enzymes. In: Biodiversity and Conservation (Vol. 2). USA: Studium Press LLC.*

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

1.3 Directed evolution

This section introduces the concept of protein engineering as a powerful tool for expanding genetic diversity, beyond its natural pool, and demonstrates that it perfectly complements metagenomics. The section also emphasises the importance of molecular cloning in the context of directed evolution experiments. It has been adapted from:

***Jajesniak P**, Wong TS. (2015) Tapping into biodiversity: From metagenomics to industrial enzymes. In: Biodiversity and Conservation (Vol. 2). USA: Studium Press LLC.*

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

REMOVED FROM THE E-THESIS

DUE TO COPYRIGHT

1.4 Diversity of non-coding DNA sequences

This section expands the discussion of genetic diversity beyond genes and proteins - to other non-coding DNA sequences. Section 1.4 focuses specifically on bacterial promoters as arguably one of the most crucial types of non-coding DNA sequences and demonstrates their importance in the context of synthetic biology. In addition, the information presented in here facilitated the judicious choice of a suitable promoter for Multi-Genius method, as described in Chapter 4. Finally, this literature review provides a necessary context for the subsequent discussion of new methods of engineering prokaryotic organisms, via simultaneous alteration of multiple genes, presented in Section 1.5. It has been adapted from:

***Jajesniak P**, Wong TS. (2015) From genetic circuits to industrial-scale biomanufacturing: bacterial promoters as a cornerstone of biotechnology. AIMS Bioengineering, 2(3), 277-296.*

1.4.1 Introduction

Bacteria are widely known as one of the most abundant and ubiquitous groups of living organisms known to man, and their importance to life sciences research over the past decades cannot be overstated. The prevalence of bacteria can be attributed not only to their high diversity, exceptional evolvability and short generation time but also to their robust mechanism of gene regulation. Individual bacterial cell can quickly and efficiently adapt to new environment by redirecting its metabolic flux and adjusting its protein content in response to external stimuli (e.g., production of β-galactosidase, lactose permease, and galactoside O-acetyltransferase by *Escherichia coli* in response to the presence of lactose in glucose-deficient medium) (Gorke and Stulke, 2008; Jacob and

Monod, 1961). Such a tight control of catabolic, biosynthetic and stress-response pathways (to name but a few) within a cell allows for a highly efficient utilization of naturally scarce resources.

Recent advances in our understanding of bacterial systems have provided the scientific community with an expanding molecular toolbox for altering and manipulating the genetic makeup of various bacterial species. Our ability to engineer biological organisms for useful purposes has sparked the development of synthetic biology and metabolic engineering. Despite the growing number of enabling technologies, transcriptional regulation still takes the centre stage in engineering prokaryotic organisms. Careful selection of promoter plays a critical role in maximizing bioproduct yield and, as a result, is key to biomanufacturing industry (Terpe, 2006; Weickert et al., 1996). While bioprocess specialists are, more often than not, interested in strong, inducible promoters with low or nonexistent basal expression profiles, metabolic engineers and synthetic biologists utilize various types of promoters for their work, which includes, but is not limited to, the design of genetic circuits and construction of heterologous biosynthetic pathways (Blazeck and Alper, 2013). Consequently, the demand for reliable and well-characterized promoters of different types is steadily increasing in various areas of biological research. As basic techniques of promoter engineering and characterization are becoming more and more accessible, there has been a large number of exciting developments in the field of transcriptional regulation in recent years.

In this section, I aim to provide a comprehensive overview of bacterial promoters research and engineering, with an emphasis on their application in recombinant protein

production, synthetic biology and metabolic engineering. First, I give a brief outline of prokaryotic gene regulation from a theoretical standpoint, highlighting the ramifications of biological complexity and ubiquity of nonlinear biomolecular interactions in the context of promoter engineering and design. Subsequently, I explore the role of promoters in large-scale protein production, discussing promoter selection criteria and recent advances in this area. This is followed by a summary of scientific endeavours in the field of synthetic biology and metabolic engineering that are centred around transcriptional regulation. I delve into the subject of promoter databases and associated bioinformatics tools, and the idea of describing promoters as standardized biological parts is carefully scrutinized. The review concludes by highlighting emerging technologies that complement transcriptional regulation and exploring future prospects of this research area.

1.4.2 Theoretical background

Over the past half a century, gene regulation in bacteria has been extensively studied. Although many details concerning their complex regulatory mechanisms are still unknown, major progress has been made in our understanding of their cellular machineries. Detailed mechanism of DNA transcription, promoter architecture and three-dimensional structures of transcription factors are only a few examples of what the scientific community has gathered, over this short time period, through combination of ingenious experimentation and rigorous data analyses. This information has enabled not only the development of more advanced molecular tools for engineering prokaryotic

organisms but also shed light on the far more complex mechanisms of eukaryotic gene regulation.

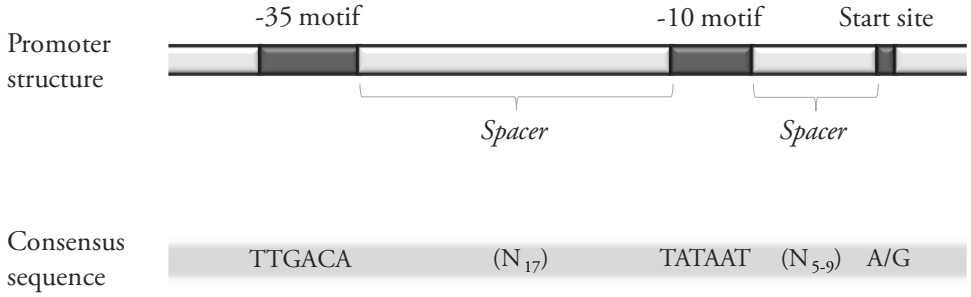

Figure 1.6 Outline of prokaryotic promoter structure - simplified structure and consensus sequence of *E. coli* σ^{70} promoter. During promoter recognition, σ^4 and σ^2 domains of σ^{70} factor (subunit of RNA polymerase holoenzyme) bind to -35 and -10 motifs, respectively, allowing for transcription initiation (Busby and Ebright, 1994). After the two DNA strands are separated by RNA polymerase, a single nucleoside triphosphate pairs itself with a nucleotide constituting transcriptional start site (+1 position) and becomes 5' end of growing RNA transcript, created during the elongation process.

Escherichia coli, the workhorse in laboratory, has been extensively studied as a model system of bacterial gene regulation. Since the inception of gene transcription studies, promoter has been described as one of the most fundamental regulatory elements present in bacterial operons. Based on numerous sequencing experiments, their relatively simple structure has been determined and is widely available (Figure 6) (Harley and Reynolds, 1987; Hawley and McClure, 1983; Oliphant and Struhl, 1988). Promoters, however, should rarely be analysed in isolation - it is the complex interplay between a promoter (including its operators), RNA polymerase holoenzyme, transcription factors and effector molecules that gives rise to different rates of transcription initiation (Busby

and Ebright, 1994; Ishihama, 1993), which, more often than not, significantly influence the final expression levels of regulated proteins. Even a single-base substitution or deletion can have a detrimental effect on binding energies of DNA-protein and protein-protein complexes (Ebright et al., 1984; Lewis et al., 1996) and, consequently, alter the properties of the whole system. Moreover, interactions between individual elements of the regulatory system are not always linear. For example, stronger binding between RNA polymerase holoenzyme and the promoter sequence is frequently associated with a higher rate of transcription initiation, however, increased stabilization of RNA polymerase complex sometimes has the opposite effect – transcription inhibition can be achieved by preventing RNA polymerase from leaving a promoter (*e.g.*, phage φ29 regulatory protein p4 that binds to A2c promoter acts as a transcription repressor via this mechanism) (Monsalve et al., 1998). The outlined biological complexity makes engineering of promoters a very challenging task and favours the use of random and semi-rational strategies such as directed evolution for this purpose.

The subject of transcriptional gene regulation in bacteria has been covered from a theoretical standpoint in many excellent reviews (Balleza et al., 2009; Browning and Busby, 2004) and textbooks devoted to general biochemistry and molecular biology. As such, I kindly invite any readers interested in exploring, in more details, theoretical basis of the outlined mechanisms to consult the relevant resources.

1.4.3 Promoter engineering

In light of biological complexity, outlined in the previous section, *de novo* design is still a largely unfeasible method of acquiring novel prokaryotic promoters. Most of the new

bacterial promoters of biotechnological significance are a result of promoter engineering or, more specifically, the three main strategies of prokaryotic promoter engineering – saturation mutagenesis of spacer regions (Figure 6), error-prone PCR and hybrid promoter engineering (Blazeck and Alper, 2013).

During saturation mutagenesis of spacer regions, mutagenic oligonucleotides are designed to vary the nucleotide sequence surrounding -35 and -10 motifs (Figure 6). This semi-rational strategy relies on the fact that extensive changes to promoter consensus regions, in most cases, significantly decrease binding of RNA polymerase molecule and, consequently, leaving them unchanged allows for a dramatic decrease in the number of non-functional promoter variants. Moreover, varying the spacer length is often unnecessary as the variable region between the two motifs has a constant optimal length (17 base pairs in the case of many *E. coli* promoters) (Aoyama et al., 1983; Mulligan et al., 1985). The discussed method has been used, for example, to create a library of synthetic promoters of varying strength for bacteria *Lactococcus lactis* (Jensen and Hammer, 1998) and *Lactobacillus plantarum* (Rud et al., 2006).

A complementary strategy to saturation mutagenesis of spacer regions is error-prone PCR, which was used, in combination with green fluorescent protein (GFP) expression and flow cytometry, to modify constitutive bacteriophage P_L promoter and create a library of 22 synthetic promoters of varying strength (Alper et al., 2005). Both methods, saturation mutagenesis of spacer regions and error-prone PCR, are claimed to be fairly equal in terms of the number of advantages and disadvantages, with a major difference

being that error-prone PCR involves screening libraries with a lower fraction of functional promoters (Table 1.3).

Table 1.3 Comparison between two main strategies of promoter engineering - saturation mutagenesis of spacer regions and error-prone PCR (based on the information provided in (Alper et al., 2005; Hammer et al., 2006; Jensen and Hammer, 1998)).

	Saturation mutagenesis of spacer regions	Error-prone PCR
Target	Spacer regions	Whole promoter
Theoretical library size	$>4^{17}$	$>4^{35}$
Fraction of functional library members	>50%	<1%
Reported screening technologies	Colorimetric assays	Colorimetric assays Flow-cytometry
Reported variation in promoter strength	400-fold	196-fold

The third strategy of obtaining new synthetic promoters is hybrid promoter engineering, which involves merging promoter parts from different sources. For example, the widely-used *tac* promoter is a hybrid of *trp* and *lac* promoters (de Boer et al., 1983). The region upstream of -20 position, with respect to transcriptional start site, derives from *trp* promoter whereas the rest originates from *lac*UV5 promoter (mutant of *lac* promoter that is no longer sensitive to catabolite repression). The resultant promoter is 3 and 11 times more efficient than *trp* and *lac* promoter, respectively, and remains inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG). Other examples of hybrid promoters include *Bacillus subtilis* P_{SPAC} promoter (Yansura and Henner, 1984) and *E. coli rha*P_{BAD} expression system (Haldimann et al., 1998).

It is worth mentioning that advances in the field of promoter engineering come not only from new methods of creating combinatorial libraries but also from developments in screening technologies. As mentioned, use of GFP expression along with flow cytometry allowed for high-throughput screening of promoter library and identifying promoter variants of different strengths (Alper et al., 2005). This approach facilitates screening of large combinatorial libraries, the size of which, however, is limited by efficiency of bacterial transformation. *In vitro* compartmentalization, a method of utilizing a water-in-oil emulsion and *in vitro* transcription/translation to create vast combinatorial libraries (Tawfik and Griffiths, 1998), was successfully used for directed evolution of peptides (Yonezawa et al., 2003), ribozymes (Levy et al., 2005) and proteins (Ghadessy et al., 2001; Griffiths and Tawfik, 2003, 2006; Sepp and Choo, 2005). Recently, this technique was demonstrated to be applicable to promoter engineering – use of *in vitro* compartmentalization led to identification of T7 promoter variant with a 10 times higher *in vitro* transcriptional activity (Paul et al., 2013).

1.4.4 Protein expression and promoter selection

Production of proteins, especially industrial enzymes, monoclonal antibodies and other biopharmaceuticals, remains a core activity of current biotechnological industry. Owing to their short generation times and relative ease of genetic manipulations, bacteria are widely utilized as protein expression platforms, despite their inability to perform post-translational modifications (Porro et al., 2011). *E. coli* is, without a doubt, a workhorse of bacterial protein manufacturing, accounting for production of nearly a third of 211 biopharmaceuticals approved by the end of year 2011 (Berlec and Strukelj, 2013;

Walsh, 2014). Several *Bacillus* species (i.e., *Bacillus brevis, Bacillus megaterium* and *Bacillus subtilis*) are also used as common biomanufacturing platforms, however, their utilization is mainly limited to production of homologous enzymes (*e.g.*, proteases and amylases) (Liu et al., 2013; Westers et al., 2004). Rapid advances in genetic engineering and constant development of new molecular tools have contributed to identification of new platforms for recombinant protein production – for examples, several bacterial species belonging to *Pseudomonas genus (*e.g.¸ *Pseudomonas fluorescens)* have been reported to provide protein yield comparable or higher than E. coli and were utilised for biopharmaceutical production *(Chen, 2012; Retallack et al., 2012).* Various other bacterial species, distinct form *E. coli*, are sometimes used for industrial-scale protein production, however, the prevalence of such cases still remains quite low (Chen, 2012; Terpe, 2006).

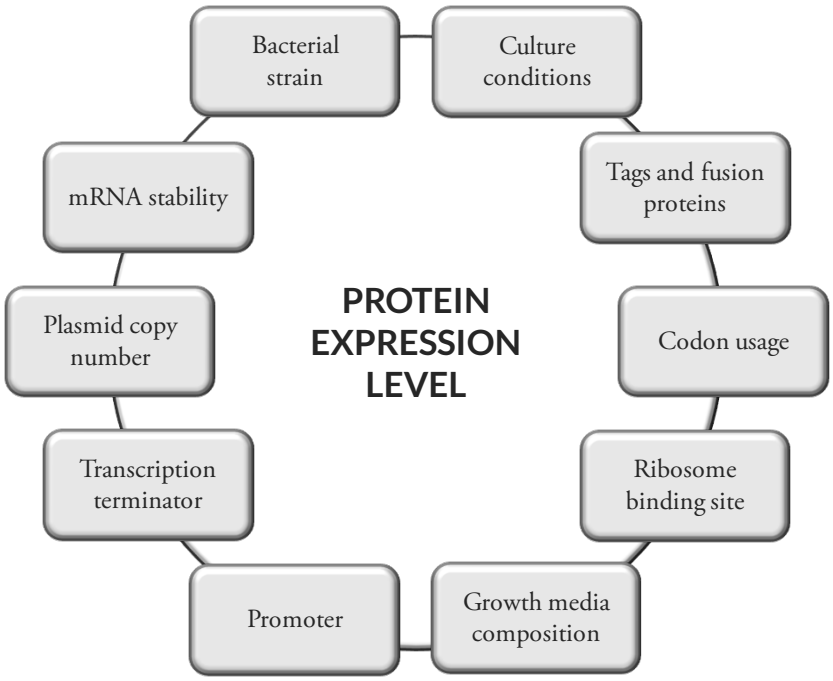

Figure 1.7 Main factors affecting protein expression levels during heterologous protein production.

Selecting the right host organism is only one example among a myriad of different variables (Figure 7) that one has to consider to maximize the production yield of functional protein (Gopal and Kumar, 2013; Liu et al., 2013; Rosano and Ceccarelli, 2014; Sorensen and Mortensen, 2005; Weickert et al., 1996). Among these decisions, an appropriate promoter is, arguably, one of the most crucial and impactful factors. In many cases, when protein production is concerned, the desired specifications of a promoter are well-defined – it should allow for robust, responsive and low-cost induction, be characterised by relatively low basal expression level, and maximize the final yield of functional protein. Unfortunately, looking at the list of most common *E. coli* expression systems and their properties (Table 1.4), it becomes immediately apparent that these criteria are often at odds with one another. For example, T7 expression system, characterised by the highest expression level, is associated with one of the most expensive inducers (i.e., IPTG) and a considerable basal expression level (Pan and Malcolm, 2000). To be fully functional, T7 promoter also requires the presence of T7 RNA polymerase, which has to be encoded on an additional plasmid or, more commonly, within the bacterial chromosome - e.g., *E. coli* BL21 (DE3), C41 (DE3) and C43 (DE3). Further, it is vital to note that the strength of a promoter does not always correlate with the final yield of functional protein. T7 promoter is notorious for producing a high fraction of insoluble protein, which frequently aggregates into intracellular inclusion bodies (Balzer et al., 2013; Tegel et al., 2011). Finally, a low expression level or low promoter leakiness is sometimes advantageous when expressing membrane proteins or other toxic biomolecules (Rosano and Ceccarelli, 2014; Terpe, 2006).

Table 1.4 Most commonly-used inducible *E. coli* expression systems and their important properties (adapted from (Terpe, 2006) and (Weickert et al., 1996)); inducer prices are based on the information provided by Sigma-Aldrich Corporation, http://sigmaaldrich.com).

Name	Expression level	Basal expression	Catabolite repression	Inducer	Inducer cost	References
lac	Low-medium	High	Yes	IPTG	High (63.90 $/g)	(Gronenborn, 1976)
*lac*UV5	Low-medium	High	Reduced [1]	IPTG	High (63.90 $/g)	(Silverstone et al., 1970; Wanner et al., 1977)
trp	High	High	No	IAA [2]	High (61.20 $/g)	(Bass and Yansura, 2000; Somerville, 1988)
tac	High	High	Reduced [1]	IPTG	High (63.90 $/g)	(de Boer et al., 1983)
trc	High	High	Reduced [1]	IPTG	High (63.90 $/g)	(Brosius et al., 1985)
phoA	High	Low	No	Phosphate starvation	-	(Craig et al., 1991; Neubauer and Winter, 2001)
P_L	High	Low [3]	No	Temperature shift	-	(Elvin et al., 1990; Valdez-Cruz et al., 2010)
tetA	Medium-high [4]	Low	No	Anhydrotetracycline	Low (1.67 $/mg) [5]	(Skerra, 1994)
araBAD (P_{BAD})	Low-high [4]	Low	Yes	L-arabinose	Low (2.30 $/g)	(Guzman et al., 1995; Siegele and Hu, 1997)

$rhaP_{BAD}$	Low-high [4]	Low	Yes	L-rhamnose	High (61.30 $/g)	(Haldimann et al., 1998)
T5/*lac*	Very high	High	No	IPTG	High (63.90 $/g)	(Deuschle et al., 1986; Gentz and Bujard, 1985; Samuelson, 2011)
T7 [6]	Very high	High [7]	Reduced [1]	IPTG	High (63.90 $/g)	(Studier and Moffatt, 1986)
T7/*proU*	Very high	Low	No	NaCl	Very low (40 $/kg)	(Donahue Jr and Bebee, 1999)

[1] The mutation of *lac* promoter reduces its sensitivity to catabolite repression, however, it does not fully eliminate it (Rosano and Ceccarelli, 2014).

[2] IAA denotes *3*-β-*indoleacrylic acid*. The promoter can also be induced by low intracellular level of L-tryptophan (nutrient starvation) (Somerville, 1988; Weickert et al., 1996).

[3] At 29-30˚C - the temperature at which repressor cI857 is fully functional (Lowman and Bina, 1990).

[4] Titratable.

[5] The amount of anhydrotetracycline commonly used for full induction of *tetA* promoter (200 μg/l) is, on average, 100-1000 lower than the amount of chemical inducer required for induction of the remaining promoters (Bass and Yansura, 2000; Terpe, 2006).

[6] Includes both *lac*- and *lac*UV5-based expression systems.

[7] New strategies, which aim to provide lower basal expression levels of T7 expression system, have been developed (Pan and Malcolm, 2000; Studier, 1991).

As a result, the choice of an appropriate expression system is often dependent on the target protein and, consequently, promoter selection remains a challenging, yet crucial task. To enable easier navigation among the various selection criteria and facilitate the process of promoter selection, decision diagram presented in Figure 8 provides an easy way of determining which of the commonly used inducible expression systems is the most adequate for a given application.

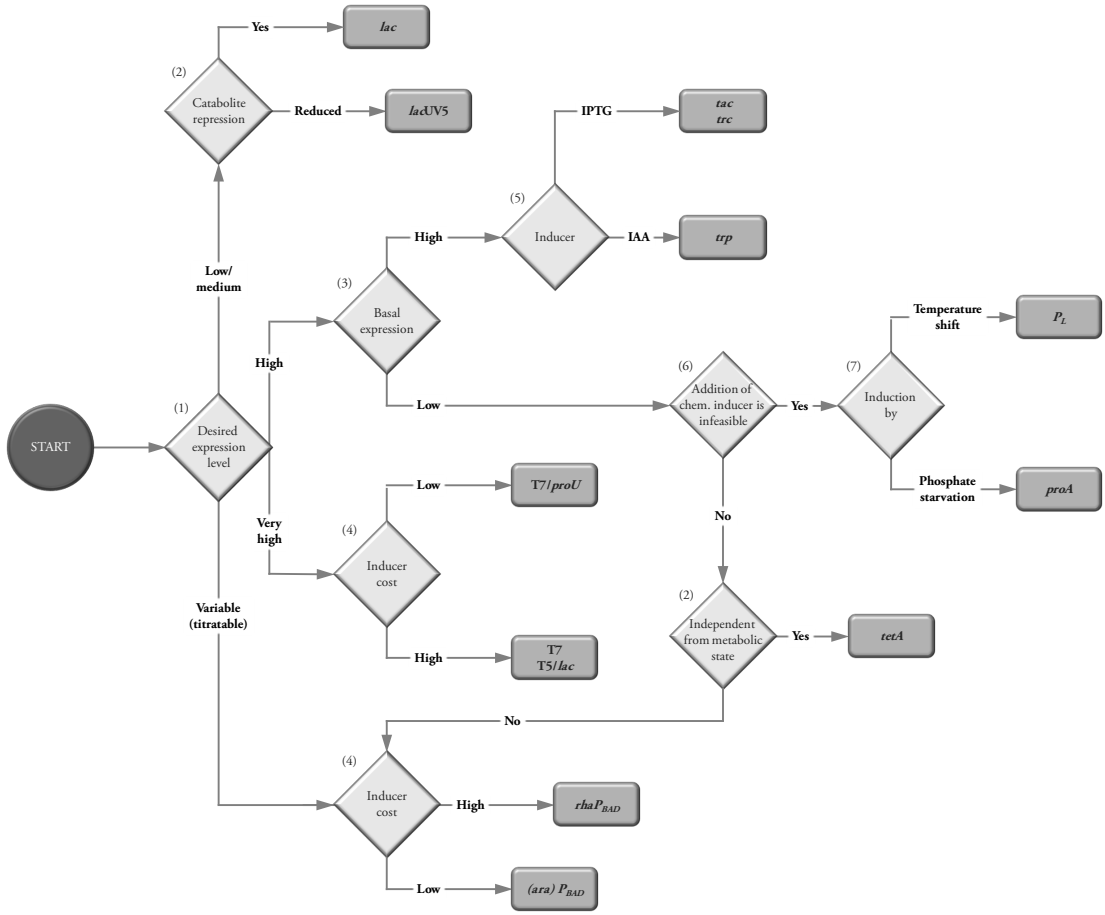

Figure 1.8 Proposed flowchart for a quick selection of an appropriate inducible promoter (based on the list of 13 most common inducible *E. coli* expression systems listed in Table 1.2). (1) At the start of the selection process, desired expression level should be chosen, taking into account protein solubility and toxicity – in the case of poorly soluble or toxic proteins, final product yield is often maximized by choosing a promoter

capable of low expression. (2) When glucose-based medium is to be used for cell culture, expression systems with a high dependency on catabolite repression system should be avoided. (3) Selecting promoter with a low basal expression is vital for expression of membrane proteins or other toxic biomolecules. (4) The cost of chemical inducer is a vital factor for large-scale protein production. (5) Choice between IPTG- and IAA-dependent promoters is likely to be based on the availability of the particular chemical inducer. (6) When addition of chemical inducer during cell culture is deemed to be infeasible, other induction strategies should be considered. (7) In the situation where temperature change of bacterial culture can be easily achieved, use of P_L expression system is recommended as it does not impose any limitations on media composition. Otherwise, *phoA* expression system should be used.

As many promoters exhibit, at least some, undesirable properties, it is not surprising that different strategies, including promoter engineering, have been utilized to enhance their characteristics or even develop novel expression systems (e.g. propionate-inducible (Lee and Keasling, 2005) and cumate-inducible (Choi et al., 2010) expression systems). Among different targets of these endeavours, efficient production of toxic proteins is prioritized (Saida et al., 2006). Development of *E. coli* expression system dependent on ferric uptake regulator (Fur), which exhibits relatively high expression and tight regulation, illustrates the rapid progress in this area (Guan et al., 2013). Table 1.5 provides a comprehensive overview of some recently-reported expression systems developed for three prokaryotic groups of major bioprocessing importance - *E. coli*, *Bacillus* spp. and *Pseudomonas spp.*

Table 1.5 Novel expression systems developed for *Escherichia coli*, *Bacillus* spp. and *Pseudomonas spp.* over the last decade.

Organism	Name	Important characteristics	Reference
E. coli	Propionate-inducible expression system (P_{prpB})	• Homogenous expression • Wide range of inducer-dependent expression level	(Lee and Keasling, 2005)
	Cumate-inducible expression system	• Tight regulation • High-level and homogenous expression • Wide range of inducer-dependent expression level	(Choi et al., 2010)
	Gene expression system Hsh	• Heat-shock inducible expression • Very high expression level	(Wu et al., 2010)
	pLAI expression system	• Expression triggered by high-cell density • Tight regulation • Low-level basal expression	(Nocadello and Swennen, 2012)
	Fur-dependent expression system (P_{fhuA})	• Tight regulation • Suitable for toxic protein synthesis	(Guan et al., 2013)
B. subtilis	Subtilin-regulated gene expression (SURE)	• Very high expression level • Low-level basal expression	(Bongers et al., 2005)

		• Not subjected to catabolite control	
	Maltose-inducible expression system	• High-level expression	(Ming-Ming et al., 2006; Ming et al., 2010)
	Cold-inducible expression system	• Temperature-inducible expression	(Thuy Le and Schumann, 2007)
B. megaterium	Sucrose-inducible promoter system (P_{sacB})	• Alternative to a well-established xylose-inducible promoter system (comparable expression levels)	(Biedendieck et al., 2007)
	T7 RNA polymerase-dependent expression system	• High-level expression • Reported difficulties with extracellular protein expression	(Gamer et al., 2009)
Pseudomonas spp.	P_{BAD}-based shuttle vectors	• Highly-regulated expression	(Qiu et al., 2008)

As addition of chemical inducer to bacterial culture is often problematic (e.g., potential of contamination) and expensive in large-scale protein production, much effort in enhancing current expression systems is concentrated around alleviating this problem. Use of temperature-inducible promoters, such as bacteriophage P_L promoter under the control of temperature-sensitive repressor cI857, constitutes a solution to this problem (Elvin et al., 1990; Valdez-Cruz et al., 2010), however, this strategy is not free from its inherent drawbacks. Heating large volumes of fermentation broth is not only challenging from an engineering and economic standpoint but can also have serious ramifications for the stability of temperature-sensitive proteins. As a result, the introduction of auto-induction media for T7 expression system has been a major breakthrough in protein expression (Studier, 2005). Auto-induction medium contains a substantial amount of glucose, which in combination with other specified culture conditions inhibits protein production at the early stage of cell growth; only after glucose is depleted, protein expression takes place (Studier, 2005, 2014). The widespread adoption of this method has motivated other researches to develop new strategies of auto-induction. For instance, utilization of a quorum sensing system from *Vibrio fischeri* allowed for creating *E. coli* expression system that couples protein production with cell density (Nocadello and Swennen, 2012).

Recent emergence of robust cell-free expression systems (Carlson et al., 2012; Katzen et al., 2005; Schwarz et al., 2008) necessitates development of promoters that are fully compatible with *in vitro* transcription/translation mixtures. As discussed above, the use of *in vitro* compartmentalization allowed for identification of T7 promoter mutant with

a 10 times higher *in vitro* transcriptional activity (Paul et al., 2013). Moreover, studies on transcription factors and RNA polymerase bring a wealth of information that could prove invaluable for promoter engineering. For example, T7 RNA polymerase mutation that decreases its propensity for abortive transcription and, consequently, increases its *in vitro* transcription efficiency, was reported (Guillerez et al., 2005).

1.4.5 Synthetic biology and metabolic engineering

The two emerging areas of biological research, synthetic biology and metabolic engineering, have been, for quite some time, a rich source of invention and scientific progress. Regardless of some differences between the two disciplines – e.g., synthetic biology is in principle more concerned with *de novo* design, whereas the efforts of metabolic engineers are concentrated on modifying existing biological pathways (both endogenous and heterologous) - they both share the common goal of devising useful biological systems. As a result, it is also not surprising that both disciplines are currently deeply intertwined (Yadav et al., 2012).

In contrast to heterologous protein production, metabolic engineering is characterized by a frequent use of constitutive promoters (Mijakovic et al., 2005). Because construction of optimized metabolic pathways usually requires expression of individual genes to be at different, yet well-defined levels, there exists a high demand for libraries of synthetic constitutive promoters of different strength. As mentioned in the promoter engineering section, both saturation mutagenesis of spacer regions and error-prone PCR have been extensively used to create such libraries (Alper et al., 2005; Jensen and Hammer, 1998; Rud et al., 2006).

Altering promoter strength is one of the main methods of optimizing metabolic pathways. Due to this fact, various inducible promoters play an essential role in the development of metabolic engineering. Use of inducible araBAD expression system allowed for modulating individual gene expression levels (by changing arabinose concentration) and, consequently, removing a bottleneck from *E. coli* mevalonate pathway *(Pitera et al., 2007)*. By utilizing promoters of different strength (*lac*, *lac*UV5 and *trc*), Anthony *et al.* alleviated two pathway bottlenecks and achieved a 5-fold increase in amorphadiene production, a precursor to anti-malarial compound (Anthony et al., 2009). Finally, T5, T7 and *trc* inducible promoters were used to carefully optimize two metabolic pathways in order to maximize the titers of taxadiene, a precursor to potent anticancer drug Taxol (Ajikumar et al., 2010). A 15000-fold increase in its production was reported.

Inducible expression systems are also a central part of synthetic biology. Genetic circuits constitute a great example of how simple biological components, including promoters, can be assembled into fully functional biological systems with complex properties, which significantly differ from the properties of the individual components. For instance, combining three transcriptional repressor systems resulted in construction of synthetic oscillatory network, with typical period of hours (Elowitz and Leibler, 2000). On the other hand, use of two repressible promoters arranged in a mutually inhibitory network allowed for the development of a genetic toggle switch in *E. coli* (Gardner et al., 2000). Recently, design and construction of genetic circuits moved towards closer integration between synthetic and endogenous circuitry (Elowitz and Leibler, 2000) – *e.g.*, *B. subtilis*

gene circuit responsible for inducing transient cell competency was analyzed in details and re-engineered to prevent the cells from exiting the competency state (Suel et al., 2006).

1.4.6 Databases and bioinformatics tools

Advances in promoter engineering and steady growth in the number of available promoters have sparked the development of complementary databases and bioinformatics tools. Arguably, one of the most known database of biological parts is the iGEM registry (recognized also under the name of the Registry of Standard Biological Parts; http://parts.igem.org). The database includes hundreds of functional promoters and is used each year by students around the world to construct functional biological systems as part of the International Genetically Engineered Machine (iGEM) competition. The listed promoters conform to the BioBrick standard, allowing for their quick assembly together with other functional components provided by the repository (Shetty et al., 2008). The database consists of both inducible and constitutive promoters, including a set of 20 constitutive promoters of various strengths. The vast majority of promoters listed in the iGEM registry are designed with *E. coli* in mind, however, the number of functional components conforming to BioBrick standard created for use with other organisms*, e.g., Bacillus subtilis,* is slowly increasing (Radeck et al., 2013). It should be noted that all promoter entries present in the database provide information about their DNA sequence, allowing for their in-house or commercial syntheses thereby alleviating the need of requesting them from the registry. Promoter databases other than the iGEM registry have been created – for example, PromEC

database provide DNA sequences of all endogenous *E. coli* promoters (excluding their regulatory elements) (Hershberg et al., 2001) and DBTBS database lists upstream regulatory elements of *B. subtilis* (Ishii et al., 2001; Makita et al., 2004; Sierro et al., 2008). Nevertheless, they often provide little to no information about properties of the listed promoters.

Much effort in the area of bioinformatics concerned with prokaryotic microorganisms has been devoted to the development of reliable promoter prediction tools. Their usefulness is certainly not limited to facilitating the identification of novel regulatory elements - they constitute an essential part of genome analysis and annotation. As a result, advances in promoter recognition and developments in the field of operon prediction software are often closely intertwined. Finally, ability to reliably detect promoter sequences can help immensely in preventing introduction of unwanted transcription initiation sequences when creating biomolecular constructs (Yao et al., 2013). A great example of program capable of predicting prokaryotic promoters and regulon is PePPER (de Jong et al., 2012). This web server allows for a quick identification of prokaryotic promoters based on the recognition of -35 and -10 promoter DNA motifs. This prediction is not error-free and, consequently, the existence of both undetected sequences and false positives should be taken into account when analyzing results from this program.

It is worth mentioning that the list of bioinformatics tools aiming to expedite the design of synthetic biological systems, from individual components, is steadily increasing and includes programs such as GenoCAD (Czar et al., 2009), TinkerCell (Chandran et al.,

2009) and Synthetic Biology Software Suite (SynBioSS) (Hill et al., 2008; Kaznessis, 2011). These programs aim to mimic computer-aided design (CAD), hugely popular among many engineering disciplines, and bridge the gap between the vast amount of biological data and computational modelling. In comparison to original engineering programs, however, these tools still lack certain functionalities. One of which is incorporating functional properties of individual biological part (e.g., promoter activity).

1.4.7 Standardization and quantification of promoter strength

Availability of standardized and well-defined biological components is one of the main premises of synthetic biology (Andrianantoandro et al., 2006; Endy, 2005) and one of the most promising methods of accelerating the development of biological research. The iGEM registry constitutes a first step towards standardization of biological components by providing a common method of assembling them into complex biomolecular constructs (all functional DNA sequences are flanked by well-defined restriction sites) (Shetty et al., 2008). However, the highlighted standardization is only limited to the assembly method, and does not encompass functional characteristics of each part, *e.g.*, promoter strength. As a result, judicious choice of a promoter remains a challenging task, especially when a well-defined expression level is desired.

The main obstacle to standardization of promoters and its activity is biological complexity. Most often than not, whenever a new promoter is discovered or engineered, its activity is determined indirectly, by measuring expression level of a reporter protein, e.g., green fluorescence protein (GFP). However, very frequently such a result is not very reliable as protein expression level is not only dependent on the rate of transcription

initiation but also on a myriad of different factors (Figure 7). As a result, the reported protein expression level can be reproduced only when all the other parameters remain the same, *e.g.*, amount of produced protein will increase when a plasmid with a higher copy number is used. Furthermore, there are significant differences in the accuracy of various methods of protein expression quantification. For example, measuring protein activity (e.g., fluorescence of GFP) usually does not provide any information about the amount of insoluble protein present in the cell – the in-depth comparison of T7, *trc* and *lac*UV5 promoters showed that *lac*UV5, the weakest promoter among the three, produces the highest fraction of soluble protein (Tegel et al., 2011). In a similar study, five expression systems (LacI/P_{T7}, LacI/P_{trc}, AraC/P_{BAD}, XylS/P_m and XylS/P_m ML1-17) were compared using a variety of different methods, including mRNA quantification, activity measurements, polyacrylamide gel electrophoresis and flow cytometry (Balzer et al., 2013). T7 promoter was confirmed to produce the highest amount of RNA transcript, which resulted in a correspondingly high production of insoluble protein. In addition to providing invaluable information about the five promoter systems, the study brought attention to the differences in protein expression between individual cells, indicating differences in culture homogeneity between the investigated promoter systems.

An important milestone in standardization of bacterial promoters and its activity has been the introduction of Relative Promoter Unit (RPU) (Kelly et al., 2009). In their research paper, Kelly *et al.* argue that the most representative and unbiased indicator of promoter activity is the rate of transcription initiation. Accurate determination of

absolute transcription initiation rate proves to be very challenging, if not impossible, and as a result its usefulness for describing promoter activity had been limited for a long time. It is shown, however, that under right experimental conditions relative rate of transcription initiation is approximated by the ratio of GFP synthesis rates of two promoters, the quantity that can be fairly easily determined experimentally. By measuring promoter activity relative to a standard promoter (BBa_J23101 in this case), the variation in the obtained results can be reduced by about 50%. As a result, the concept of RPUs allows for a more reliable comparison between different promoters and paves the way for comprehensive standardization of bacterial promoters.

1.4.8 Complementary technologies and future prospects

Regulation of transcription initiation is certainly not the only method of influencing the protein expression level. Modulating mRNA stability has been, for quite some time, one of the strategies of engineering gene expression (Carrier and Keasling, 1997, 1999; Gao et al., 2012; Smolke et al., 2000). Similar to promoters, ribosome binding sites (RBSs) can be designed to provide a specified rate of mRNA translation and, consequently, protein expression; in addition, based on thermodynamic calculations the strength of RBSs can be predicted *in silico* (Salis et al., 2009). Another approach, which allows for fine-tuning of gene expression, involves engineering of intergenic regions within a single operon - various post-transcriptional control elements were recombined and screened for desired expression, leading to 100-fold variation in the relative expression levels (Pfleger et al., 2006). Methods utilizing protein scaffolds (Dueber et al., 2009) and riboswitches (Topp et al., 2010) were also reported. Despite a great variety of presented strategies,

promoter selection still remains an integral part of these experiments and, as a result, advances in promoter engineering are still driving the development of genetic engineering and biotechnology. Additionally, the highlighted methods of modulating protein expression should be perceived as being complementary to promoter engineering rather than being in a direct competition to one another.

Judging by the variety of recombinant proteins produced using prokaryotic organisms and the diverse applications of bacterial promoters in the emerging fields of biological research, it would be unwise to assume that a single expression system, robust and versatile enough to meet all demands of the scientific community, will be constructed in the near future. It is much more plausible that the rapid expansion of promoter engineering (Blazeck and Alper, 2013), accelerated by the developments in the high-throughput screening technologies (Paul et al., 2013), will provide us with an abundance of bacterial promoters with unique characteristics. As the number of available expression systems is expected to rise continually, it is imperative that promoter selection procedures are expedited and streamlined by further developing enabling bioinformatics tools, expanding existing databases, and adopting a unified method of measuring and quantification of promoter activity. Only then, the key goal of synthetic biology – construction of functional and well-defined biological systems from standardized biological components – can be fully realized.

1.5 Genetic diversity and interactions between multiple genes

The review of genetic diversity has been, so far, limited to the discussion of single genetic elements. However, many phenotypes are encoded by multiple genes. For example, an ability to produce certain chemicals of commercial value is often encoded by vast metabolic pathways, spanning a whole host of closely-regulated genes. Intrinsic resistance of certain microorganisms to harsh environmental conditions - e.g., extreme temperature – can be attributed to the expression of more than one gene. What is more, the genes responsible for a particular phenotypes does not necessarily have to be located next to each other and might be located within distant genetic loci. This phenomenon significantly limits the usability of traditional function-based metagenomics when it comes to identifying genes encoding for complex phenotypes as only one DNA fragment can be screened at a time.

As a result, in order to be able to efficiently engineer the whole organisms, another approach is needed. Traditionally, this has been achieved through, so called, classical strain improvement. In this approach, organisms of interest is subjected to the action of physical or chemical mutagens (e.g., UV radiation) and constant selective conditions (e.g., increasing temperature), which depend on a specific phenotype that is being evolved or improved on. Despite its widespread use, the method is often very time-consuming and resource-intensive. Additionally, many mutations unrelated to the targeted phonotype are introduced in the process. The scale and precise nature of these genetic changes are often unknown, frequently leading to creation genetically-unstable

organisms. Finally, the use of harmful mutagens is also a major drawback that raises serious health and safety concerns.

To circumvent the addressed shortcomings, in recent years, novel methods allowing for simultaneous alteration of multiple genetic elements have been created to eventually provide a more robust way of identifying genes encoding for useful phenotypes and evolving organisms with desired properties. The reported methods can be broadly classified into three main groups depending on the utilised mechanism (Tee and Wong, 2014):

(1) DNA shuffling

(2) copy number

(3) transcription factors

(4) regulatory elements (e.g., promoters and ribosomal binding sites)

A representative example of the first group is whole-genome shuffling reported by Zhang *et al* (Zhang et al., 2002). The usefulness of the presented method was demonstrated by obtaining 9-fold increase in tylosin production by actinobacteria *Streptomyces fradiae*.

Chemically inducible chromosomal evolution (CIChE) is a copy-number-based method that utilises *E. coli recA* homologous recombination to create and screen bacterial cells containing different of gene copy number on their chromosomes (Tyo et al., 2009). The

use of the reported method led to obtaining 4-fold increase in biopolymer poly-3-hydroxybutyrate production and 10-fold increase in genetic stability.

Another novel approach to evolving useful traits involves engineering of prokaryotic transcription factors, e.g., the *E. coli* main sigma factor, σ^{70} (Alper and Stephanopoulos, 2007). Random mutagenesis of *rpoD,* encoding for σ^{70}, alters the promoter preferences of RNA polymerase and allows for identification of *E. coli* cells exhibiting abnormal phenotypes. Global transcription machinery engineering (gTME) was shown to be successful at evolving lycopene overproduction and increased tolerance to ethanol.

Finally, multiplex automated genome engineering (MAGE) targets many locations on a bacterial chromosome via the use of degenerate oligonucleotides (Wang et al., 2009b). Oligonucleotide-mediated sequence replacement can be used to target both genes and regulatory elements. Targeting of ribosome binding sites allowed for optimization of the 1-deoxy-d-xylulose-5-phosphate (DXP) biosynthesis pathway in *E. coli* and obtaining a 5-fold increase in lycopene production. It is worth noting that the whole experimental procedure was fully automated. On the other hand, trackable multiplex recombineering (TRMR) uses synthetic DNA cassettes to replace endogenous promoters or ribosome binding sites (Warner et al., 2010). This strategy, in combination with utilisation of molecular barcoding and microarrays, led to mapping of *E. coli* genes that confer growth advantage in various media and in the presence of several growth inhibitors.

Despite the great variety of the presented methods, their widespread use is frequently hindered by the complex protocol and high operating costs. One of the few methods offering a relatively simple workflow is based on Coexisting/Coexpressing

Genomic Libraries, CoGeLs, (Nicolaou et al., 2011). In this method, fragmented genomic DNA from an organism of interest is fragmented and transferred into individual plasmids (or fosmids) and, in contrast to standard genomic libraries, two plasmids are transformed into each cell (via two consecutive transformations). Both plasmids can normally coexist in the same cell as they contain different origins of replication. This approach allows for identification of interactions between distant genetic loci. As a proof of concept, the method has been used to identify genes encoding for complex acid-tolerance phenotype in *E. coli* and lead to identification of the novel combination of *arcZ* and *recA* genes that enhanced acid tolerance by 9000-fold (confirming that complex phenotypes such as acid tolerance can be encoded two-genes combinations).

Despite its great potential, the presented method is not free from inherent drawbacks. Necessity of two successive transformations considerably increases the complexity and duration of the whole method, and decrease its reliability (a significant fraction of screened cells is likely to contain more than two plasmids). Moreover, despite major advances in our understanding of regulatory elements and genetic circuits, they have not been used extensively in the area of multiple gene alterations. Finally, all the methods reported concentrate on altering genes, or their expression, that are native to the evolved organism. So far, there has been no major attempts to combine genetic information from other organisms – for example, utilising chromosomal DNA of thermophilic bacteria to evolve thermostability in common bacterial strains.

1.6 Scope and objectives

The aim of this thesis is to expand molecular toolbox for synthetic biology and allow a more efficient access to largely untapped potential of biological diversity. By reviewing the current landscape of synthetic biology and various methods of tapping into biodiversity, I identified two high-impact areas that are particularly lacking and, as such, constitute the focal point of my PhD project.

Gene cloning is a basic molecular technique that constitutes an important part of nearly every complex experiment associated with genetic engineering and is used extensively in research laboratories around the world. The impact of molecular cloning is particularly evident in areas such as protein engineering where, on many occasions, it is a major bottleneck to the process of directed evolution. Consequently, its performance is detrimental the rate with which new biological parts are discovered. Furthermore, faster and simpler molecular cloning techniques would address the issue of biological complexity, one of the main challenges of synthetic biology (Kwok, 2010), by facilitating the process of constructing and testing new biological circuits. As such, despite its seemingly simple concept, the importance of molecular cloning is paramount.

To address this problem, I aim to build upon recent advances in gene cloning and molecular biology, in general, to develop a relatively simple, fast and efficient method of integrating a DNA fragment of interest into an expression vector of choice. If possible, this strategy should also be applicable to cloning of combinatorial libraries and its usefulness to various branches of synthetic biology – including, but not limited to, protein engineering - should be clearly demonstrated.

The second high-impact area to be indentified is the field of microbial strain improvement and identifying genes encoding for complex phenotypes. In recent years, many novel methods for simultaneous alterations of multiple genes have been created to advance the progress of this area of research. The reported methods aim at improving biological organisms via a variety of different methods such as DNA shuffling, modification of gene copy number or utilisation of either transcription factors or regulatory elements such as promoters and ribosome binding sites. Unfortunately, most of the reported techniques are very complex and resource intensive, considerably limiting their general applicability.

To address the discussed shortcomings, I aim to expand the concept of genomic libraries to allow for a more robust way of identifying genes encoding for useful phenotypes and evolving microorganisms with desired properties. The functionality of the method should be proven by evolving a useful phonotype – e.g., halotolerance or thermotolerance – in an industrially-relevant bacterial species, such as *Escherichia coli*.

If successful, this PhD project should expand molecular toolbox available to synthetic biologists and enhance the way biodiversity – both natural and laboratory-induced – is being studied and utilised.

Chapter 2
QuickStep-Cloning

This chapter has been adapted from:

***Jajesniak P**, Wong TS. (2015) QuickStep-Cloning: a sequence-independent, ligation-free method for rapid construction of recombinant plasmids. Journal of Biological Engineering, 9, 15*

2.1 Abstract

Background: Molecular cloning is an essential step in biological engineering. Methods involving megaprimer-based PCR of a whole plasmid are promising alternatives to the traditional restriction-ligation-based molecular cloning. Their widespread use, however, is hampered by some of their inherent characteristics, e.g., linear amplification, use of self-annealing megaprimers and difficulty with performing point insertion of DNA. These limitations result in low product yield and reduced flexibility in the design of a genetic construct.

Result: Here, we present a novel technique of directional cloning, which overcomes these problems yet retaining the simplicity of whole-plasmid amplification. QuickStep-Cloning utilizes asymmetric PCRs to create a megaprimer pair with 3'-overhangs, and hence, facilitates the subsequent exponential whole-plasmid amplification. QuickStep-Cloning generates nicked-circular plasmids, thereby permitting direct bacterial transformation without DNA ligation. It allows DNA fragment integration into any plasmid at any position, in an efficient, time- and cost-effective manner, without tedious intermediate DNA gel purification, modified oligonucleotides, specialty enzymes and

ultra-competent cells. The method is compatible with competent *E. coli* cells prepared using the conventional calcium chloride method.

Conclusion: QuickStep-Cloning expands the versatility of megaprimer-based cloning. It is an excellent addition to the cloning toolbox, for the benefit of protein engineers, metabolic engineers and synthetic biologists.

2.2 Background

Gene cloning is an indispensable molecular biology technique that, since its first introduction, has been central to the development of genetic engineering and, consequently, the entire field of life sciences. Despite its widespread use, the traditional, restriction-ligation-based cloning protocol suffers from major problems, including, but not limited to: (i) low efficiency, (ii) dependency on the availability of unique restriction sites in a cloning vector and in the gene of interest, and (iii) time-consuming and labour-intensive process. In recent years, many novel approaches to molecular cloning have been proposed to expedite the procedure, enhance cloning efficiency and bypass the requirement of restriction sites (Lu, 2005; Tee and Wong, 2013). Homologous recombination (Court et al., 2002; Zhu et al., 2010), incorporation of phosphorothioate oligonucleotides (Blanusa et al., 2010) and use of zinc finger nucleases (Shinomiya et al., 2011) are only a few examples of different strategies utilized for this purpose.

Among the reported approaches to DNA cloning, methods involving megaprimer-based PCR of a whole plasmid, e.g., restriction site-free cloning (Chen et al., 2000), restriction-free (RF) cloning (van den Ent and Lowe, 2006), overlap extension PCR cloning (Bryksin and Matsumura, 2010) and MEGAWHOP cloning (Miyazaki, 2011), have attracted a significant interest among the scientific community. These methods were inspired by the hugely popular and easy-to-use QuikChange™ (Agilent) protocol for site-directed mutagenesis (Life Technologies, 2017). Despite their indisputable potential, megaprimer-based methods are inherited with several drawbacks that compromise their overall efficiency: (i) linear amplification of the recipient vector, (ii)

use of a completely complementary megaprimer pair, (iii) difficulty with performing point insertion of DNA, (iv) random mutations introduced by the DNA polymerase of choice during whole-plasmid amplification, and (v) poor amplification of GC-rich DNA fragments. The listed drawbacks significantly decrease the overall efficiency of the cloning method and, consequently, necessitate the use of enzyme-based DNA ligation and time-consuming optimisation of PCR conditions to achieve a sufficient number of transformants containing recombinant DNA of interest. Four recently proposed cloning methods, asymmetric bridge PCR with intramolecular homologous recombination [ABI-REC, (Bi et al., 2012)], recombination-assisted megaprimer (RAM) cloning (Mathieu et al., 2014) exponential megaprimer PCR (EMP) cloning (Ulrich et al., 2012), and inverse fusion PCR cloning [IFPC, (iGEM)], have been reported to achieve exponential amplification via incorporating additional oligonucleotides into megaprimer PCR. In all cases, however, the amplification results in generation of linearized plasmids instead of the more desirable circular DNA. ABI-REC and RAM are homologous recombination-dependent methods, relying on transformation of linearized plasmids and their repair *in vivo*, which usually provides significantly less transformants than transformation of nicked or intact plasmids. On the other hand, EMP and IFPC cloning protocols require phosphorylation and ligation to circularize the amplification products.

Here, we report QuickStep-Cloning, a novel method that builds upon the simplicity of QuikChange™. Not only it addresses major drawbacks of traditional DNA cloning, the method also circumvents the aforementioned problems of the existing megaprimer-

based cloning methods, including the problem of linear amplification and self-annealing megaprimers.

2.3 Results and Discussion

2.3.1 QuickStep-Cloning: Principle and molecular mechanism

In QuickStep-Cloning, DNA fragment of interest is amplified in two parallel asymmetric PCRs (Wang et al., 2009a), during which regions complementary to the integration site on the recipient plasmid are added to both ends of the amplified DNA fragment (Figure 2.1).

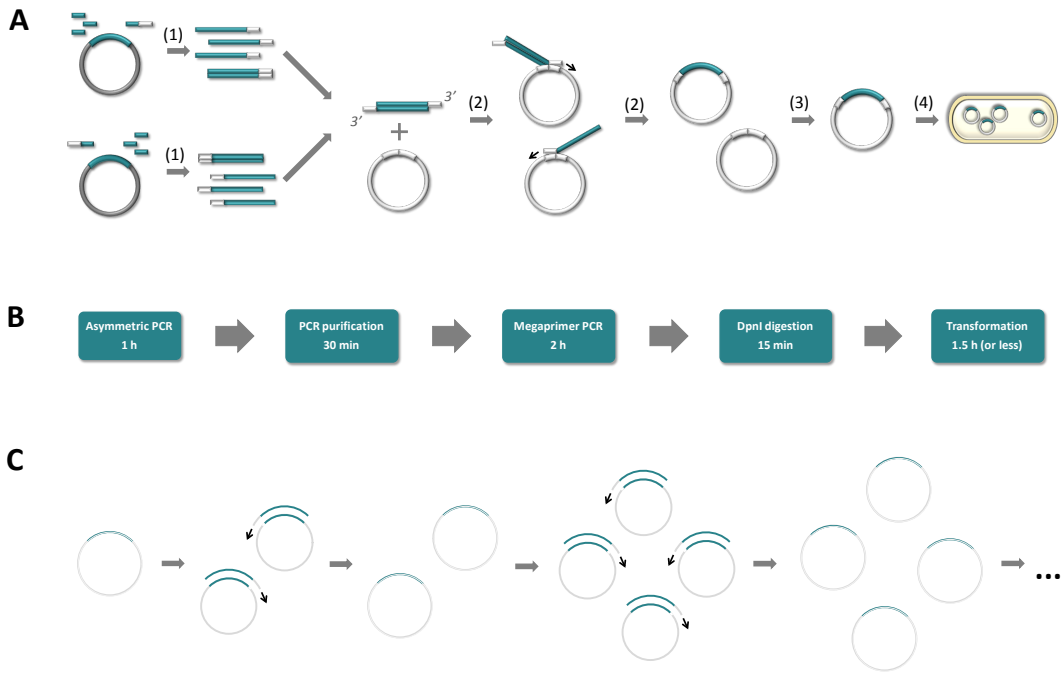

Figure 2.1 Overview of QuickStep-Cloning: (A) A schematic diagram presenting individual stages involved in the proposed method: (1) two parallel asymmetric PCRs of DNA of interest and PCR purification, (2) megaprimer-based PCR, (3) DpnI digestion, and (4) bacterial transformation. (B) Exemplary workflow for 1 kb insert and 7 kb cloning vector (exact duration of the asymmetric PCR depends on the length of cloned DNA fragment and the duration of megaprimer PCR is related to the size of the cloning vector). (C) Outline of exponential

amplification taking place during QuickStep-Cloning – megaprimers anneal themselves to the product of linear amplification and are extended by polymerase, producing further copies of the two single-stranded templates in an exponential manner. It should be noted that for the given mechanism, exponential amplification occurs in parallel with the linear process.

The products of the two asymmetric PCRs are purified, mixed and used as megaprimers for the consecutive PCR. In contrary to the traditional megaprimer-based PCR of a whole plasmid, the megaprimer pair in QuickStep-Cloning contains 3'-overhangs (instead of blunt ends) allowing it to anneal to the recipient plasmid even when the two megaprimers self-anneal. Megaprimers designed in this way facilitate an exponential amplification, which results in production of nicked-circular plasmids. After a short incubation with DpnI to remove methylated/hemimethylated recipient plasmids that do not contain gene of interest, the product of the megaprimer PCR can be directly used for transformation. For a 1-kb gene and a 7-kb recipient plasmid, for instance, the entire workflow can be completed in less than 6 hours (Figure 2.1).

2.3.2 Primer design for QuickStep-Cloning

QuickStep-Cloning allows point integration of a gene at any position of any recipient plasmid. This is achieved through judicious design of the four primers (denoted as *Fwd*, *Rev*, *IntA-Fwd* and *IntB-Rev*, Figure 2.2) that are used in the two parallel asymmetric PCRs. *Fwd* and *Rev* are primers derived from the target gene sequence only (18–22 bp). *IntA-Fwd* and *IntB-Rev*, are chimeric primers, carrying both the sequence upstream or downstream to the integration site and the target gene sequence (46–48 bp).

Asymmetric PCR with unbalanced concentration of *Fwd* (500 nM) and *IntB-Rev* (10 nM) primers results in sense strands with integration sequence at 3'-termini. Likewise in another asymmetric PCR using 10 nM of *IntA-Fwd* and 500 nM of *Rev*, antisense strands with integration sequence at 3'-termini are produced. When both strands from the two asymmetric PCRs are purified and mixed, megaprimer pairs with 3'-overhangs are produced for use in the subsequent megaprimer PCR step.

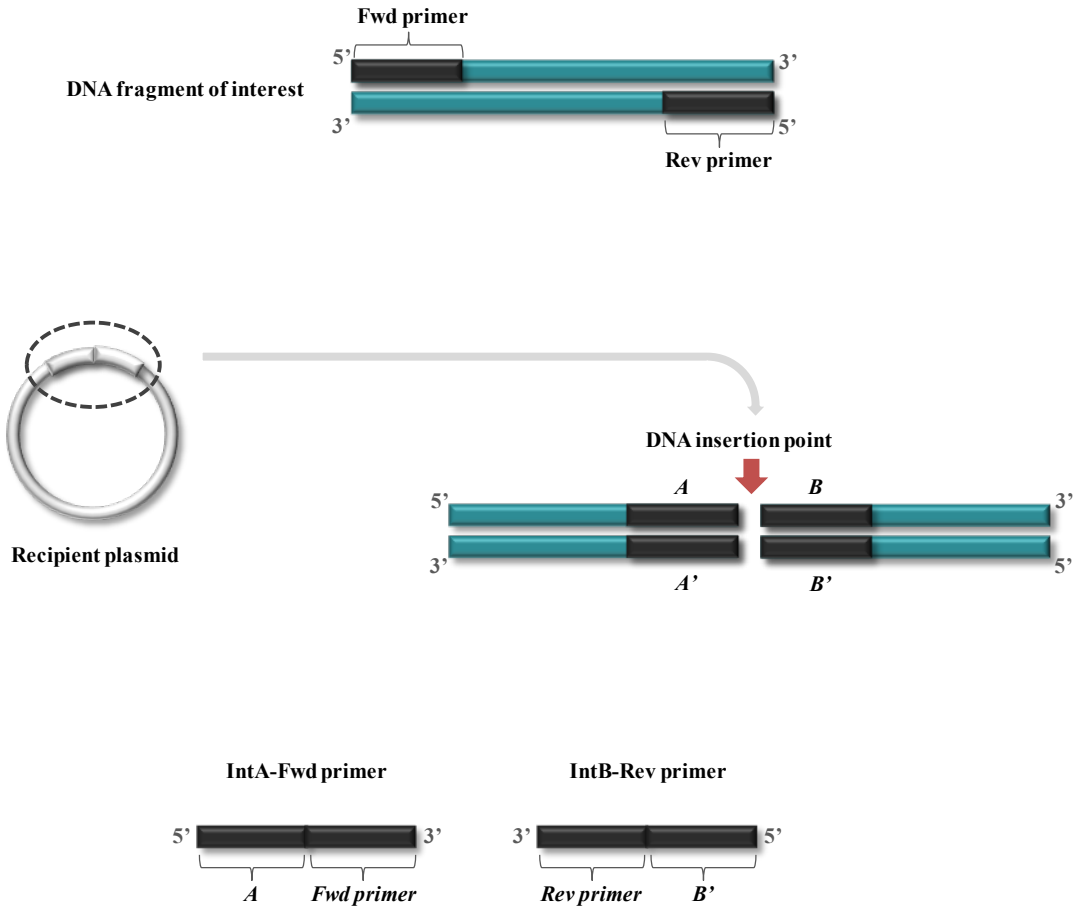

Figure 2.2 Outline of primer design for QuickStep-Cloning. The sequences of *Fwd* and *Rev* primers are derived from the target gene sequence. *IntA-Fwd* and *IntB-Rev* are chimeric primers, carrying sequence of integration site (5'-portion) and target gene sequence (3'-portion).

2.3.3 Demonstration of QuickStep-Cloning

To investigate the efficiency of the proposed design, QuickStep-Cloning was utilized to transfer a DNA fragment from pEGFP vector (containing ampicillin resistance gene; Figure 2.3) into pET24a-HLTEV-p53 plasmid (containing kanamycin resistance gene). The primers were designed to perform a point insertion of *egfp* gene just before the p53 open reading frame (Figure 2.4), producing kanamycin-resistant transformants capable of EGFP expression. After 30 cycles of asymmetric PCR and 25 cycles of megaprimer PCR, *E. coli* strains DH5α and C41 (DE3) were transformed with 5 µl of DpnI-digested PCR product and plated on agar plates supplemented with ampicillin or kanamycin and IPTG (Table 2.1).

Table 2.1 Results of *egfp* cloning experiment. Colony counts for *E. coli* strains DH5α and C41 (DE3) transformed with the products of RF cloning and of QuickStep-Cloning and plated on agar plates supplemented with: (i) 100 µg/ml ampicillin, (ii) 50 µg/ml kanamycin, and (iii) 50 µg/ml kanamycin and 1 mM IPTG. Transformation efficiency was determined based on concurrent transformation of 1 ng intact pET24a-HLTEV-p53 plasmid. Numbers in the brackets denote EGFP-expressing colonies, as determined by visual inspection using UV transilluminator. Lack of colonies observed on ampicillin-supplemented agar plates indicated that the final PCR mixture produced via QuickStep-Cloning, used directly for bacterial transformation, did not contain significant amount of donor plasmid.

Strain, Selection plate	QuickStep-Cloning	RF Cloning	Transformation efficiency [cfu/µg]
DH5α, Ampicillin	0	0	$3.8 \cdot 10^4$
DH5α, Kanamycin	476	35	$3.8 \cdot 10^4$
C41(DE3), Kan+IPTG	618(575)	160(7)	$4.2 \cdot 10^6$

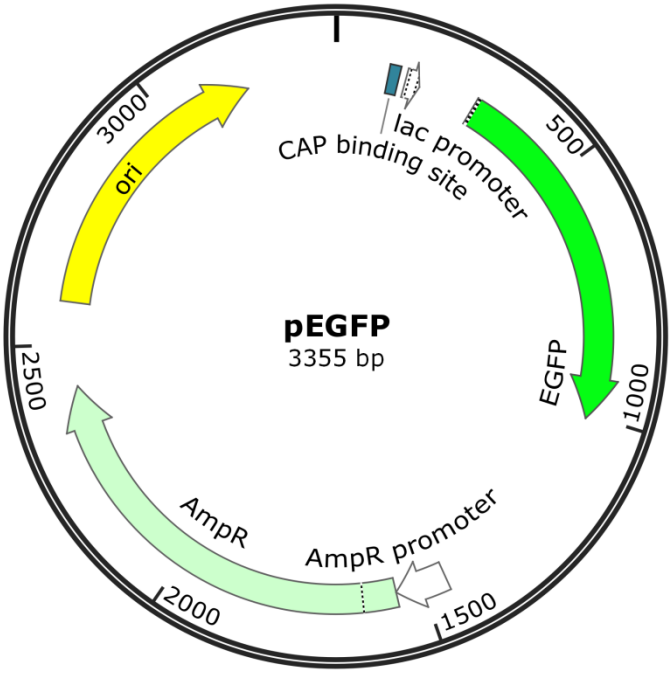

Figure 2.3 Plasmid map of pEGFP vector used in *egfp* cloning experiment (created with SnapGene).

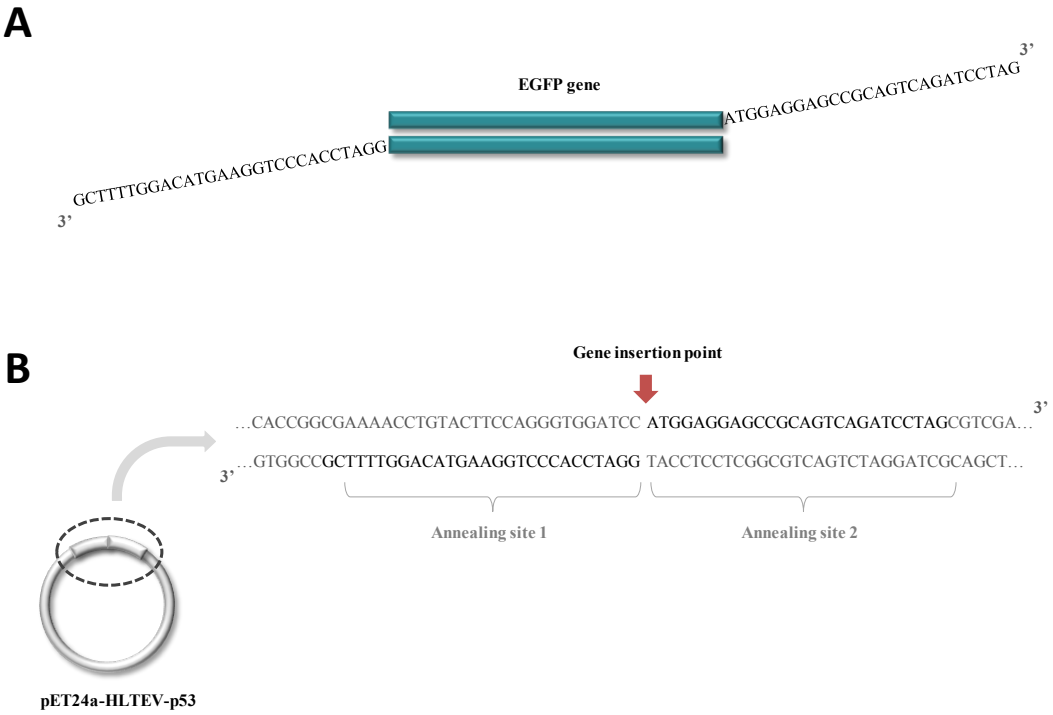

Figure 2.4 Outline of *egfp* gene cloning experiment. (A) Product of two asymmetric PCRs – *egfp* gene with 3' overhangs corresponding to the annealing sites flanking DNA insertion point present in pET24a-HLTEV-p53. (B) Sequence of the two megaprimer annealing sites.

EGFP-expressing colonies were easily discernible for C41 (DE3) grown on IPTG-supplemented plates (Figure 2.5). The accuracy of visual inspection of the transformants has also been further verified, by selecting randomly five EGFP-negative colonies and growing them at 30°C for 24 h in TB-based auto-induction media – no fluorescence was detected for all five clones (Figure 2.6). Worthy of note, there is no need to first remove *p53* gene that is pre-cloned into the recipient vector, highlighting a useful feature of QuickStep-Cloning. Further, the product of QuickStep-Cloning can be directly transformed into an expression strain [such as C41 (DE3)] for protein expression, bypassing the intermediate cloning strain (DH5α). Plasmids from ten randomly selected EGFP-expressing colonies were sequenced and the presence of DNA insert in the recombinant pET24a-HLTEV-p53 has been confirmed for all 10 clones. Worthy of note, no undesired mutation was found within the *egfp* gene in any of the 10 clones. In 9 cases, *egfp* gene was inserted at desired position in the right orientation. One plasmid contained two copies of *egfp* gene separated by a 28 bp sequence, containing partial sequence of *IntA-EGFP-Fwd* and *IntB-EGFP-Rev* primers. This construct is, most likely, a result of megaprimer-dimer formation during whole plasmid amplification. Concurrently, plasmids from five EGFP-negative colonies were sequenced – one clone contained no insert and the remaining four carried unwanted mutations in *egfp* gene. Three of them contained single base substitutions. One contained three single base substitutions and one 3 bp deletion, all present in the region where primers *EGFP-Fwd* or *IntA-EGFP-Fwd* bind.

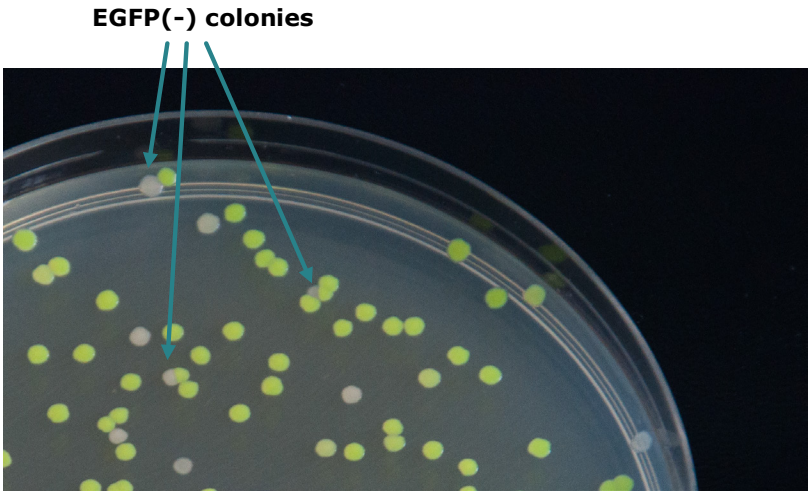

Figure 2.5 Photograph of *E. coli* C41 (DE3) colonies formed on agar plates supplemented with 50 µg/ml kanamycin and 1 mM IPTG after being transformed with the product of QuickStep-Cloning (as part of *egfp* cloning experiment). EGFP-expressing colonies are easily discernible even under visible light.

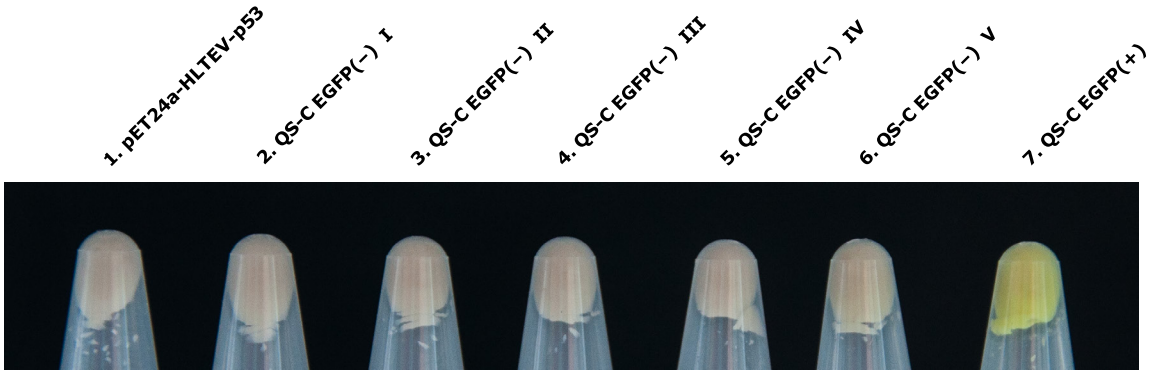

Figure 2.6 Cell pellets from cell cultures grown as part of *egfp* cloning experiment. Five EGFP-negative colonies together with one EGFP-expressing colony and one colony containing original pET24a-HLTEV-p53 were picked randomly and used to inoculate separate 5 ml aliquots of TB-based auto-induction media. After 24 h incubation at 30°C, 3 ml aliquots of cell culture were spun down and the resultant cell pellets were visually inspected for EGFP expression.

2.3.4 Optimizing QuickStep-Cloning

The success of QuickStep-Cloning is attributed to our ability to produce (1) ssDNA in sufficient quantity in the two asymmetric PCRs, and (2) high yield of megaprimer PCR. To address the former, the primer ratio in asymmetric PCRs (*i.e.*, the ratio of *Fwd-IntA* to *Rev* and the ratio of *Rev-IntB* to *Fwd*) was optimized (Figure 2.7).

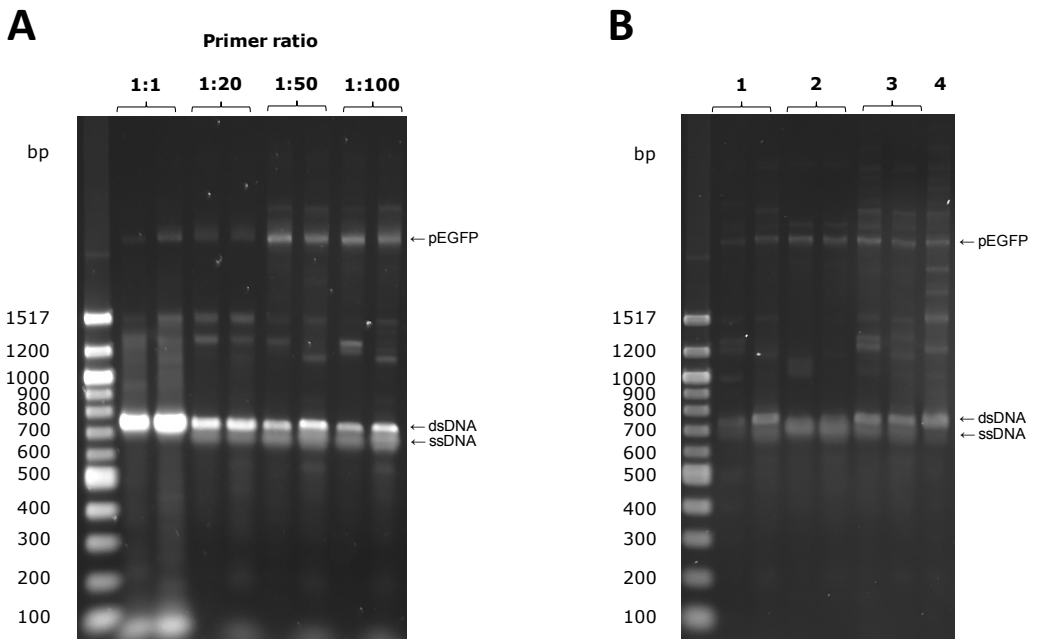

Figure 2.7 Investigation and optimization of asymmetric PCR stage of QuickStep-Cloning. (A) Yield of two parallel asymmetric PCRs for different primer ratios (represented by two separate rows for each ratio). (B) Identification of single stranded product of asymmetric PCR stage – (1) individual products of two parallel asymmetric PCRs, (2) both products after 2 min denaturation at 94°C, (3) renatured products, and (4) products of two parallel asymmetric PCRs after being mixed together. In all gel pictures, the appearance of low and high molecular weight bands could be attributed to non-specific binding of primers, commonly seen in regular PCRs.

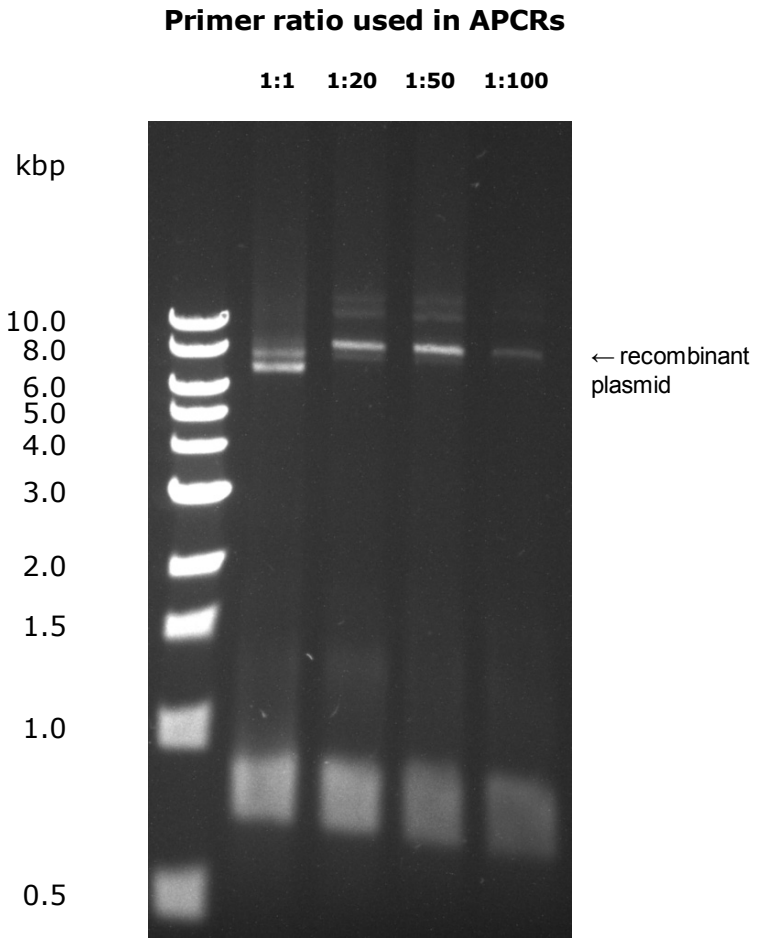

Figure 2.8 Yield of whole plasmid amplification for different primer ratios used during asymmetric PCRs.

At a ratio of 1:1, the PCR product was predominantly dsDNA, which was excellently stained by Diamond Nucleic Acid Dye from Promega (Figure 2.7A). At ratios of 1:20, 1:50 and 1:100, a lower gel band corresponding to ssDNA product started to appear. ssDNA migrates faster in agarose gel compared to its dsDNA counterpart and is less efficiently stained by fluorescent dye. A series of denaturation and annealing experiments were conducted (Figure 2.7B), confirming the identity of these lower gel bands. The effect of primer ratio on the efficiency of the proposed cloning method has further been investigated by analyzing subsequent whole plasmid amplification (Figure 2.8) and, based on the results obtained, 1:50 ratio was concluded to be the most optimal. To

obtain good product yield in megaprimer PCR, three parameters were carefully optimised, namely number of PCR cycles, concentration of recipient plasmid and megaprimer concentration (Figure 2.9).

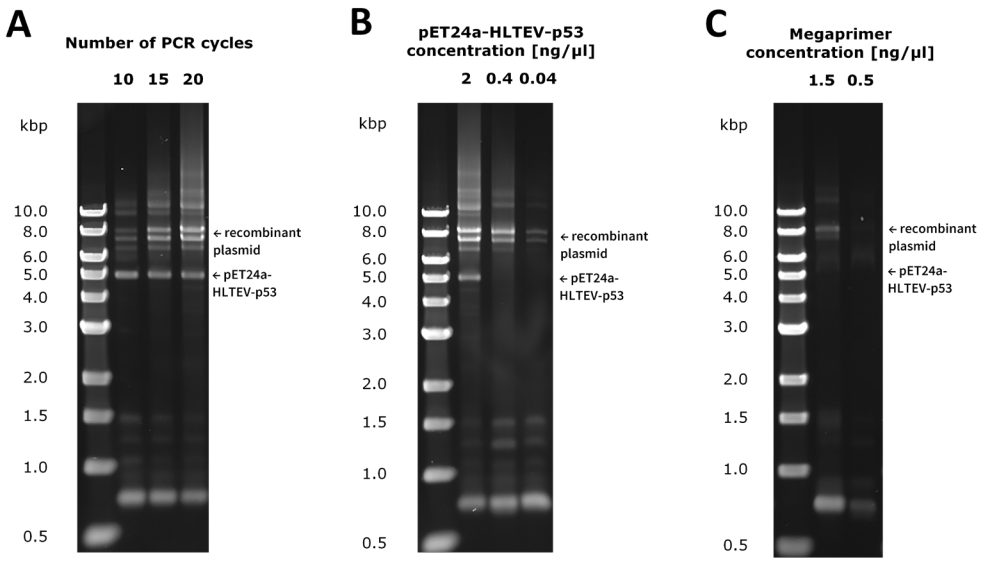

Figure 2.9 Optimization of megaprimer PCR stage of QuickStep-Cloning. (A) Yield of megaprimer PCR for varying number of PCR cycles. (B) Yield of megaprimer PCR for different concentrations of recipient plasmid. (C) Yield of megaprimer PCR for different concentrations of megaprimer. In all gel pictures, the appearance of low and high molecular weight bands could be attributed to non-specific binding of primers, commonly seen in regular PCRs.

2.3.5 Comparison to restriction-free (RF) cloning

To demonstrate the superior performance of QuickStep-Cloning, restriction-free (RF) cloning was carried out in parallel for comparison, using identical reaction conditions and primer design. QuickStep-Cloning provided much higher number of transformants - 93% of which contained recombinant plasmid (Table 2.1). RF cloning provided 160 transformants, only 4% of which displayed fluorescence. Five out of only seven EGFP-

expressing colonies obtained via RF cloning were used for subsequent sequencing – four plasmids contained the desired insert at the right orientation. One of the plasmids included not only a single mutation within the *egfp* gene but also an additional long (>100 bp) DNA fragment located between *egfp* and *p53* genes, containing partial sequence of *IntA-EGFP-Fwd* and *IntB-EGFP-Rev* primers. Poor efficiency of RF cloning might be attributed to lack of DNA ligation and, most importantly, inherent difficulties with point insertion of DNA, characteristic to many cloning methods relying on megaprimer-based PCR of a whole plasmid. Important to note, in the case of RF cloning it is advised to have a distance of 50 to several hundred base pairs between the two annealing sites on the recipient plasmid (Bryksin and Matsumura, 2010), necessitating removal of a short DNA sequence between both annealing sites during cloning. Sequencing 5 plasmids isolated from EGFP-negative colonies showed that three of them contained a relatively short (~30 bp) DNA insert instead of a desired *egfp* gene, plausibly a result of mispriming and primer-dimer formation. The remaining two did not return readable sequence.

2.3.6 General applicability of QuickStep-Cloning

QuickStep-Cloning is not limited to transfer of gene between two plasmids carrying distinct selection markers (in the case of *egfp* cloning, gene was transferred from Amp^r-pEGFP to Kan^r-pET24a-HLTEV-p53). To investigate the robustness of the developed protocol, QuickStep-Cloning method was applied to another system. The *rfp* gene from Kan^r-pBbA8k-RFP (containing *rfp* gene, under the control of arabinose-inducible promoter; Figure 2.10) was successfully cloned into Kan^r-pET24a-HLTEV-p53 using

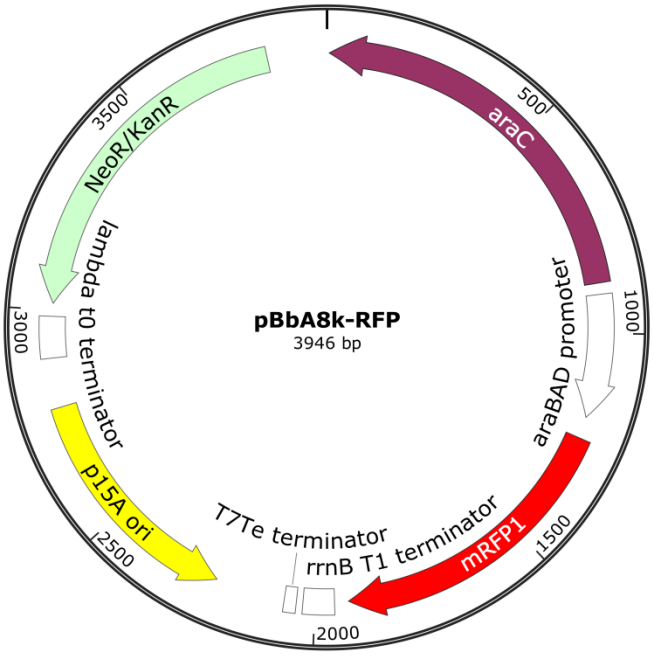

Figure 2.10 Plasmid map of pBbA8k-RFP vector used in *rfp* cloning experiment (created with SnapGene).

Table 2.2 Results of *rfp* cloning experiment. Colony counts for *E. coli* strains DH5α and C41 (DE3) transformed with the products of RF cloning and of QuickStep-Cloning and plated on agar plates supplemented with: (i) 50 µg/ml kanamycin and (ii) 50 µg/ml kanamycin and 1 mM IPTG. Transformation efficiency was determined based on concurrent transformation of 1 ng intact pET24a-HLTEV-p53 plasmid. Numbers in the brackets denote RFP-expressing colonies, as determined by visual inspection of the plates.

Strain, Selection plate	QuickStep-Cloning	RF Cloning	Transformation efficiency [cfu/µg]
DH5α, Kanamycin	334	26	$3.8 \cdot 10^4$
C41(DE3), Kan+IPTG	418(404)	113(103)	$4.2 \cdot 10^6$

QuickStep-Cloning. The only differences in the protocol from *egfp* cloning experiment have been the use of a new set of four primers, designed following general guidelines presented in this paper, and the corresponding annealing temperatures. Without any further optimization, QuickStep-Cloning again exhibited superior performance in comparison to RF cloning (Table 2.2), providing 418 colonies, 97% of which expressed RFP (Figure 2.11).

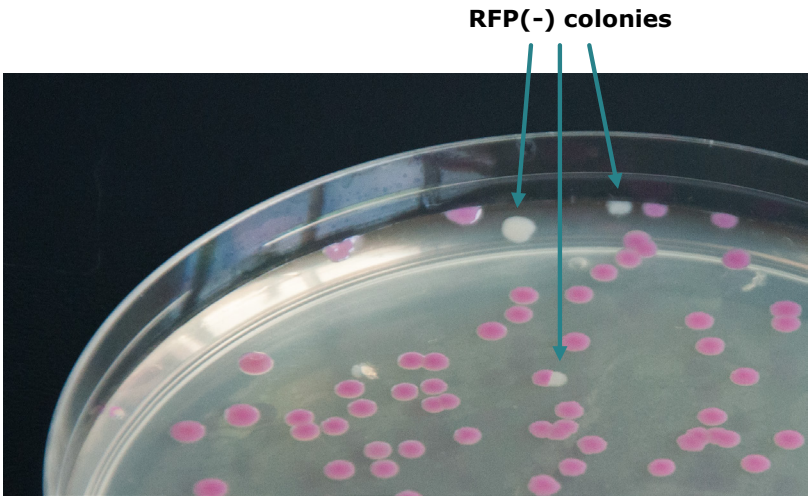

Figure 2.11 Photograph of *E. coli* C41 (DE3) colonies formed on agar plates supplemented with 50 μg/ml kanamycin and 1 mM IPTG after being transformed with the product of QuickStep-Cloning (as part of *rfp* cloning experiment). RFP-expressing colonies are easily discernible under visible light.

The accuracy of visual inspection of agar plates has been confirmed by further expression studies (Figure 2.12 and 2.13). Sequencing has shown that out of five investigated RFP-expressing transformants, all five of them contained pET24a vector with *rfp* insert. Only one clone contained unwanted mutation, namely, a 5-bp deletion downstream of *egfp* gene (*i.e.*, at the vector integration site). Interestingly, plasmids from five out of only 15 observed RFP-negative colonies have also been scrutinized and all of them had short

deletions at or close to start codon of *rfp* gene, where primers *RFP-Fwd* or *IntA-RFP-Fwd* bind. The localization of these unwanted mutations within the primer-binding region is unlikely to be purely coincidental. It is hypothesized that these artifacts could be derived from the impurities (*e.g.*, deletion products) present in the synthetic oligonucleotides. As the occurrence of occasional mutations, especially deletions, is a widely-known shortcoming of long, desalted primers, it is envisaged that the use of HPLC-purified primers can further improve the already exceptionally high efficiency of QuickStep-Cloning.

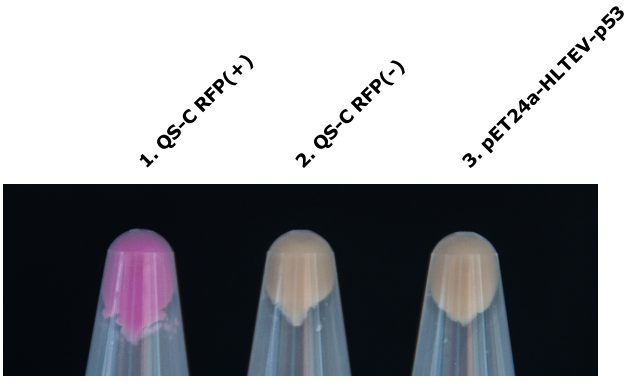

Figure 2.12 Cell pellets from cell cultures grown as part of *rfp* cloning experiment. One RFP-expressing colony together with one RFP-negative colony and one colony containing original pET24a-HLTEV-p53 were picked randomly and used to inoculate separate 5 ml aliquots of 2×TY media supplemented with 1 mM IPTG. After 48 h incubation at 30°C, 3 ml aliquots of cell culture were spun down and the resultant cell pellets were visually inspected for RFP expression.

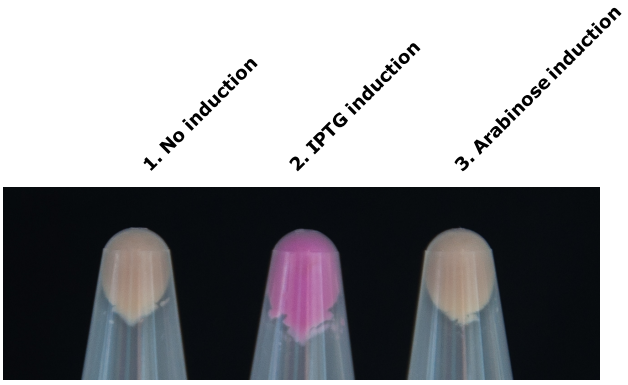

Figure 2.13 Cell pellets from cell cultures grown as part of *rfp* cloning experiment. Three RFP-expressing colonies were picked randomly and used to inoculate separate 5 ml aliquots of 2×TY media. One sample contained media only (1), one was supplemented with 1 mM IPTG (2) and 0.1% w/v arabinose was added to the remaining one (3). After 48 h incubation at 30°C, 3 ml aliquots of cell culture were spun down and the resultant cell pellets were visually inspected for RFP expression.

2.3.7 Comparison to other cloning methods

In order to highlight the novelty of and the benefits offered by QuickStep-Cloning, the proposed method was compared with four recently reported strategies of exponential megaprimer-based cloning (ABI-REC, RAM cloning, EMP cloning and IFPC); the results of this comparison are summarized in Table 2.3. QuickStep-Cloning is one of the first cloning methods fully optimised for use with the recently developed Q5 High-Fidelity DNA Polymerase (New England Biolabs), which is characterised by its ultra-low error rate (200× higher fidelity than *Taq* polymerase and approximately 2× higher fidelity than the widely-used Phusion polymerase), very high speed of DNA replication

Table 2.3 A comparison of QuickStep-Cloning to other recently reported megaprimer-based cloning methods. Desirable features are highlighted in orange to facilitate comparison.

Cloning method	QuickStep-Cloning	RF	ABI-REC	RAM	EMP	IFPC
Cloning strategy	Megaprimer	Megaprimer	Megaprimer	Megaprimer	Megaprimer	Megaprimer
Amplification mode	Exponential	Linear	Exponential	Exponential	Exponential	Exponential
Transformed product	Nicked-circular plasmid (2 nicks per plasmid)	Nicked-circular plasmid (2 nicks per plasmid)	Linear DNA	Linear DNA	Closed-circular plasmid	Closed-circular plasmid
E. coli cells used	Chemically competent DH5α and C41 (DE3)	Electrocompetent TG1	Chemically competent DH5α	Strain type not reported	Chemically competent DH5α	Chemically competent TOP10
In vivo homologous recombination	No	No	Yes	Yes	No	No
Enzymatic phosphorylation-ligation	No	No	No	No	Yes	Yes
Number of primers required	4	2	3	3	3	3

Gel purification	No	No	No	1×	No	Strongly recommended
PCR purification	1×	1×	No	No	2×	No
Estimated cloning time[1]	5 h 15 min	14 h	7 h 45 min	7 h 45 min	7 h 15 min	6 h 30 min
Reported cloning efficiency [2]	93-97%	~90% [3]	93–100%	75-94%	10–100%	~90%
Reference	-	*(van den Ent and Lowe, 2006)*	*(Bi et al., 2012)*	*(Registry of Standard Biological Parts)*	*(Ulrich et al., 2012)*	*(iGEM)*

[1]As estimated for cloning 1 kb DNA fragment into 7 kb plasmid according to originally reported protocol (for more information see Additional documentation).

[2] Judging by the percentages reported, all methods are capable of delivering similar efficiency. Worthy of note, these numbers are dependent on the approaches used by the authors to evaluate cloning efficiency.

[3] As reported in the original paper (van den Ent and Lowe, 2006). Ulrich *et al.* (Ulrich et al., 2012) and Mathieu *et al.* (Registry of Standard Biological Parts) demonstrate, respectively, 27% and 16% efficiency for RF cloning.

(6 kb/min) and superior performance for a broad range of amplicons, including DNA with a high GC content. The presented method was demonstrated to be suitable for direct transformation of not only widely used *E. coli* cloning strain (DH5α) but also a common expression strain, C41 (DE3). Worthy of note, the two distinct experiments utilizing QuickStep-Cloning (cloning of *egfp* and *rfp* genes) provided hundreds transformants (Table 2.1 and 2.2), despite the use of a relatively simple transformation method (allowing for transformation efficiencies in the range of just 10^4-10^6 cfu per μg of intact plasmid). In comparison, many of the previously-reported methods were investigated based solely on highly-efficient transformation protocols. For example, overlap extension PCR cloning, a method utilizing the principles of RF cloning, was reported to produce up to 600 colonies from small aliquots of final PCR mixture (Bryksin and Matsumura, 2010). However, the chemically competent *E. coli* cells used in that study had been prepared via Inoue method, a time-consuming protocol which allows to achieve transformation efficiencies exceeding 10^9 cfu/μg (Wong et al., 2004a). In stark contrast to the other four recently reported methods, QuickStep-Cloning does not rely on either undesirable *in vivo* homologous recombination or enzymatic phosphorylation-ligation process. The whole cloning procedure requires only one PCR purification step, whereas both RAM cloning and IFPC involve time-consuming gel purification. Based on a rough estimate of time needed to integrate a 1-kb DNA fragment into a 7 kb plasmid using the six different megaprimer-based cloning methods, QuickStep-Cloning emerges as an unquestionable winner when it comes to overall cloning time. Most importantly, its cloning efficiency compares favourably to the values reported for the remaining five methods. The only drawback of QuickStep-Cloning is

its requirement of four distinct primers (difference of one additional short primer in comparison to the other exponential cloning methods). Even though there is a chance of accidental DNA misinsertion (no such cases have yet been identified throughout our study), *Fwd* and *Rev* primers can be useful in colony PCR for quick identification of plasmids with gene insert. Of course, the use of primers complementary to vector regions flanking the insertion site is most appropriate for identifying clones with gene inserted at desired location.

Based on the presented facts, QuickStep-Cloning fares exceptionally well in comparison to other, previously-reported megaprimer-based cloning methods. However, what about more popular cloning methods such Gibson Assembly cloning (Gibson, 2011; Tyo et al., 2009) or Ligation Independent Cloning (LIC) (Zhang et al., 2002)? Both of these methods are often advertised as being able to achieve full cloning in less than an hour and less than 3 hours, respectively, appearing to be much faster when compared to 6 hour duration time provided for QuickStep-Cloning. Surprisingly, these general estimates usually not only neglect the time needed for bacterial transformation but also assume that two DNA fragments to be joined already contain complementary terminal regions and that recipient plasmid is already linearized. The last point is particularly salient in the case of any general cloning experiment utilizing Gibson Assembly or LIC, as recipient plasmid has to be linearized, most often than not, with either restriction enzymes or inverse PCR. Use of restriction enzymes for this purpose introduces a host of problems inherent to the traditional, restriction-ligation-based cloning protocol, such as dependency on the availability of unique restriction sites in a cloning vector. Application

of inverse PCR allows for sequence-independent cloning, however, it provides some of the drawbacks associated with megaprimer-based cloning (*e.g.*, reliance on error-prone polymerase of choice and necessity of careful primer design). If Gibson assembly was to be used together with inverse PCR to clone 1 kb DNA fragment into 7 kb expression vector (analogous to the proof-of-concept *egfp* cloning experiment presented in here), according to our conservative estimates, about 3 hours would be needed to perform the inverse PCR and subsequent DpnI digestion (to remove any traces of parental vector) and plasmid purification. Adding to this the time needed to perform enzymatic assembly and bacterial transformation, the total time of performing cloning via Gibson Assembly appears to be comparable to QuickStep-Cloning. Worthy of note, recipient plasmid linearization and amplification of DNA insert combined with introduction of complementary overhangs require design of the same number of primers as QuickStep-Cloning. Taking into account the cost of enzymatic reaction components (T5 exonuclease, Taq ligase, suitable polymerase and appropriate buffer sustaining simultaneous activity of all three enzymes) and the need of synthesizing four different primers, Gibson Assembly cloning seems to be more costly and resource-intensive than QuickStep-Cloning. What is more, use of highly-competent bacterial strains for Gibson Assembly is highly recommended. Without shadow of a doubt, Gibson Assembly remains a powerful and highly versatile molecular-biology tool, which involves a broad range of applications including, but not limited to, multiple-fragment assembly and molecular cloning coupled to simultaneous deletion of a DNA fragment. The same argument applies equally well to Ligation Independent Cloning. In our opinion, however, for certain applications such as point insertion of long DNA stretches into a

cloning vector, QuickStep-Cloning provides an attractive alternative to even the most popular and established cloning methods.

2.4 Conclusions

Based on the presented experimental results, it can be claimed that QuickStep-Cloning is a rapid and highly efficient method of molecular cloning. A DNA fragment of interest can be inserted into any position on the recipient vector and fully cloned in less than 6 hours, without the need of DNA ligation and with only one simple PCR purification step. The usefulness of QuickStep-Cloning is certainly not limited to standard cloning experiments, involving transfer of a gene sequence from a donor vector to a recipient plasmid. The developed method could be especially useful for protein tagging or, potentially, cloning DNA fragments directly from genomic DNA. The method should also facilitate the vital process of constructing and testing new biological circuits by allowing faster integration of genetic elements. We envisage that QuickStep-Cloning would find its applications in the developing fields of protein engineering, metabolic engineering and synthetic biology.

2.5 Methods

2.5.1 Materials

All enzymes, deoxyribonucleotides and DNA ladders were purchased from New England Biolabs (Ipswich, USA).

2.5.2 Primers

Primers used in this study were synthesized by Eurofins Genomics (Ebersberg, Germany). Melting temperatures of oligonucleotides were determined using the New England Biolabs T_m Calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator). Four primers were used in *egfp* cloning experiment: *EGFP-Fwd* (5'-ATGGTGAGCAAGGGCGAG-3', 18 bp), *IntA-EGFP-Fwd* (5'- CGAAAACCTGTAC TTCCAGGGTGGATCCATGGTGAGCAAGGGCGAG-3', 46 bp), *EGFP-Rev* (5'-TTACTTGTACAGCTCGTCCATG-3', 22 bp) and *IntB-EGFP-Rev* (5'- CTAGGAT CTGACTGCGGCTCCTCCATTTACTTGTACAGCTCGTCCATG-3', 48 bp). Underlined parts of *IntA-Fwd* and *IntB-Rev* are identical to *Fwd* and *Rev* primers, respectively, and the remaining parts correspond to the two megaprimer annealing sites flanking DNA insertion point present in pET24a-HLTEV-p53. Similarly, the following four primers were used for *rfp* cloning experiment: *RFP-Fwd* (5'-ATGGCGAGTAGCGAAGACG-3', 19 bp), *IntA-RFP-Fwd* (5'-CGAAAACCTGTAC TTCCAGGGTGGATCCATGGCGAGTAGCGAAGACG-3', 47 bp), *RFP-Rev* (5'-TTAAGCACCGGTGGAGTGACG-3', 21 bp) and *IntB-RFP-Rev* (5'- CTAGGATCT GACTGCGGCTCCTCCATTTAAGCACCGGTGGAGTGACG-3', 47 bp).

2.5.3 QuickStep-Cloning

To transfer *egfp* gene from pEGFP (Clontech Laboratories, Mountain View, USA) into pET24a-HLTEV-p53 plasmid, two asymmetric PCRs were carried out in parallel. Asymmetric PCR mixture I (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *EGFP-Fwd* primer, 10 nM *IntB-EGFP-Rev* primer, 0.2 ng pEFGP, and 1 U Q5 High-Fidelity DNA Polymerase. Asymmetric PCR mixture II (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 10 nM *IntA-EGFP-Fwd* primer, 500 nM *EGFP-Rev* primer, 0.2 ng pEFGP and 1 U Q5 High-Fidelity DNA Polymerase. Both mixtures were thermocycled using the following conditions: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 65°C and 30 s extension at 72°C. The two PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and their DNA concentration was determined using NanoDrop 2000 (Thermo Scientific, Wilmington, USA). For megaprimer PCR, the mixture (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 200 ng of purified asymmetric PCR product I, 200 ng of purified asymmetric PCR product II, 20 ng pET24a-HLTEV-p53 and 1 U Q5 High-Fidelity DNA Polymerase. The mixture was thermocycled according to the following program: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were subsequently added to the PCR mixture and incubated at 37°C for 15 min to remove the parental pET24a-HLTEV-p53 plasmids. To clone *rfp* gene from pBbA8k-RFP, purchased from Addgene (plasmid #35273), into pET24a-HLTEV-p53

plasmid, the same protocol was followed, using a dedicated primer set (*RFP-Fwd*, *IntA-RFP-Fwd*, *RFP-Rev* and *IntB-RFP-Rev*) and corresponding annealing temperature of 68°C (provided by New England Biolabs T_m Calculator) for the two asymmetric PCRs.

2.5.4 Restriction-free (RF) cloning

PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *IntA-Fwd* primer, 500 nM *IntB-Rev* primer, 0.2 ng pEFGP, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the same program as asymmetric PCR in QuickStep-Cloning: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 65°C and 30 s extension at 72°C. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using NanoDrop 2000. Megaprimer PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 nM of each dNTP, 400 ng purified PCR product, 20 ng pET24a-HLTEV-p53 and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled in the same conditions as QuickStep-Cloning megaprimer PCR: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were added to the PCR mixture and incubated at 37°C for 15 min to remove the parental pET24a-HLTEV-p53 plasmids. To clone *rfp* gene from pBbA8k-RFP into pET24a-HLTEV-p53 plasmid, the same protocol was followed, using a dedicated primer set (*RFP-Fwd*, *IntA-RFP-Fwd*, *RFP-Rev* and *IntB-RFP-Rev*) and corresponding annealing temperature of 68°C for the first PCR.

2.3.5 DNA gel electrophoresis

PCR products were analyzed using either 0.7% or 1.5% agarose gel. DNA was stained using Diamond Nucleic Acid Dye (Promega, Madison, USA). DNA ladders used were Quick-Load 1 kb DNA Ladder and Quick-Load 100 bp DNA Ladder.

2.5.6 Transformation and clone analysis

E. coli DH5α and C41 (DE3) were transformed with 5 µl of DpnI-digested products of QuickStep-Cloning or RF cloning, using a standard chemical transformation protocol (Minagawa et al., 2007). Concurrently, the two bacterial strains were transformed with 1 µl of 1 ng/µl intact pET24a-HLTEV-p53 to estimate transformation efficiency. Transformed bacteria were plated on *TYE* agar plates (10 g/l tryptone, 5 g/l yeast extract, 8 g/l sodium chloride and 15 g/l agar) supplemented with: (i) 100 µg/ml ampicillin, (ii) 50 µg/ml kanamycin, and (iii) 50 µg/ml kanamycin and 1 mM IPTG. The plates were incubated overnight at 37°C and for further 12 h at 30°C. The number of EGFP-expressing colonies was determined by visual inspection using UV transilluminator. Five EGFP-negative colonies, together with one EGFP-expressing colony and one colony containing original pET24a-HLTEV-p53 were used to inoculate separate 5 ml aliquots of TB-based auto-induction media (12 g/l tryptone, 24 g/l yeast extract, 3.3 g/l $(NH_4)_2SO_4$, 6.8 g/l KH_2PO_4, 7.1 g/l Na_2HPO_4, 0.5 g/l glucose, 2.1 g/l α-Lactose monohydrate and 0.31 g/l $MgSO_4 \cdot 7H_2O$). After 24 h incubation at 30°C, 3 ml aliquots of cell culture were spun down in 1.5 ml microcentrifuge tubes and the resultant cell pellets were visually inspected for EGFP expression. Ten EGFP-expressing colonies and five EGFP-negative colonies obtained using QuickStep-Cloning, and five

EGFP-expressing colonies and five EGFP-negative colonies obtained using RF cloning [picked randomly from C41 (DE3) Kan+IPTG plate] were grown overnight at 37°C in 5 ml 2×TY media (16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl). The recombinant plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen) and sequenced by Source BioScience (Nottingham, UK). The same transformation protocol was used for *rfp* cloning experiment. The number of RFP-expressing colonies was determined by visual inspection. Three RFP-expressing colonies were used to inoculate: (i) 5 ml 2×TY media, (ii) 5 ml 2×TY media supplemented with 1 mM IPTG, and (iii) 5 ml 2×TY media supplemented with 0.1% w/v arabinose. After 48 h incubation at 30°C, 3 ml aliquots of cell culture were spun down in 1.5 ml microcentrifuge tubes and the resultant cell pellets were visually inspected for RFP expression. Plasmids from five RFP-expressing colonies and five RFP-negative colonies obtained using QuickStep-Cloning were isolated and sent for sequencing.

2.5.7 Estimated cloning times reported in Table 2.3 – calculations

Cloning times have been determined for the experiment involving insertion of 1 kb DNA fragment into 7 kb recipient plasmid. Time needed to prepare necessary PCR mixtures has not been included in the calculations (as it is very difficult to be accurately estimated). Exact times of particular steps involved in the PCR and recommended time of DpnI digestion and enzymatic phosphorylation-ligation have been sourced from references provided in Table 2.3. As a large fraction of PCR duration involves changing the temperature of reaction mixtures, additional 45 s has been added to each cycle to account for this fact – e.g., estimated time of one cycle involving 7 s denaturation, 20 s

annealing and 30 s extension is 102 s (57 s of incubation + 45 s needed for thermocycling).

1. QuickStep-Cloning

1st PCR:	- 30 cycles: 7 s denaturation, 20 s annealing, 30 s/kb extension - Final extension: - Total PCR duration time for 1 kb insert: 1 h
Purification:	*Column-based PCR purification* Estimated time: 30 min
2nd PCR:	- 25 cycles; 10 s denaturation, 20 s annealing, 30 s/kb extension - final extension: 2 min Total PCR duration time for 7 kb recipient plasmid: 2 h
DpnI digestion	15 min
Phosphorylation-ligation	None
Transformation	1 h 30 min
Total time	**5 h 15 min**

2. RF cloning

1st PCR:	- 20 cycles: 7 s denaturation, 20 s annealing, 30 s/kb extension - Final extension: -

	Total PCR duration time for 1 kb insert: 40 min
Purification:	*Column-based PCR purification* Estimated time: 30 min
2nd PCR:	- 35 cycles; 30 s denaturation, 60 s annealing, 120 s/kb extension - Final extension: - Total PCR duration time for 7 kb recipient plasmid: 9 h 30 min
DpnI digestion	2 h
Phosphorylation-ligation	None
Transformation	1 h 30 min
Total time	**14 h**

3. ABI-REC

1st PCR:	- 30 cycles: 15 s denaturation, 30 s annealing, 60 s/kb extension - Final extension: 2 min Total time for 1 kb insert and 7 kb recipient plasmid: 4 h 15 min
Purification:	None
2nd PCR:	None
DpnI digestion	2 h
Phosphorylation-ligation	None

Transformation	1 h 30 min
Total time	**7 h 45 min**

4. RAM cloning

1st PCR:	- 20 cycles: 35 s denaturation, 35 s annealing, 30 s/kb extension - Final extension: 5 min Total PCR duration time for 1 kb insert: 1 h
Purification:	*Gel extraction* Estimated time: 1h 30 min
2nd PCR:	- 15 cycles; 35 s denaturation, 35 s annealing, 30 s/kb extension - Final extension: 5 min Total PCR duration time for 7 kb recipient plasmid: 1 h 30 min
DpnI digestion	2 h 20 min
Phosphorylation-ligation	None
Transformation	1 h 30 min
Total time	**7 h 45 min**

5. EMP cloning

1st PCR:	- 25 cycles: 10 s denaturation, 30 s annealing, 15 s/kb extension - Final extension: -

	Total PCR duration time for 1 kb insert: 45 min
Purification:	*Column-based PCR purification*
	Estimated time: 30 min
2nd PCR:	- 25 cycles; 10 s denaturation, 30 s annealing, 30 s/kb extension
	- Final extension: 2 min
	Total PCR duration time for 7 kb recipient plasmid: 2 h
DpnI digestion	30 min
Phosphorylation-ligation	2 h
Transformation	1 h 30 min
Total time	**7 h 15 min**

5. IFPC

1st PCR:	- 25 cycles: 15 s denaturation, 20 s annealing, 30 s/kb extension
	- Final extension: 7 min
	Total PCR duration time for 1 kb insert: 1 h
Purification:	Gel extraction
	Estimated time: 1h 30 min
2nd PCR:	- 25 cycles; 20 s denaturation, 30 s annealing, 30 s/kb extension
	- Final extension: 7 min
	Total PCR duration time for 7 kb recipient plasmid: 2 h

DpnI digestion	None
Phosphorylation-ligation	30 min
Transformation	1 h 30 min
Total time	**6 h 30 min**

Chapter 3
QuickStepS-Cloning

3.1 Brief introduction

QuickStep-Cloning presented in Chapter 2 builds upon the previously-reported methods to expedite the process of molecular cloning. After QuickStep-Cloning was developed, it was identified that, in practice, the first part of the method, involving the synthesis of megaprimer primer with 3' overhangs, constitutes a significant bottleneck for the whole process of gene cloning. This is caused by the requirement of performing two parallel asymmetric PCRs to obtain the desired DNA fragments.

As a result, it is predicted that QuickStep-Cloning could be significantly improved, if the synthesis of the primer was achieved in an easier and simpler way.

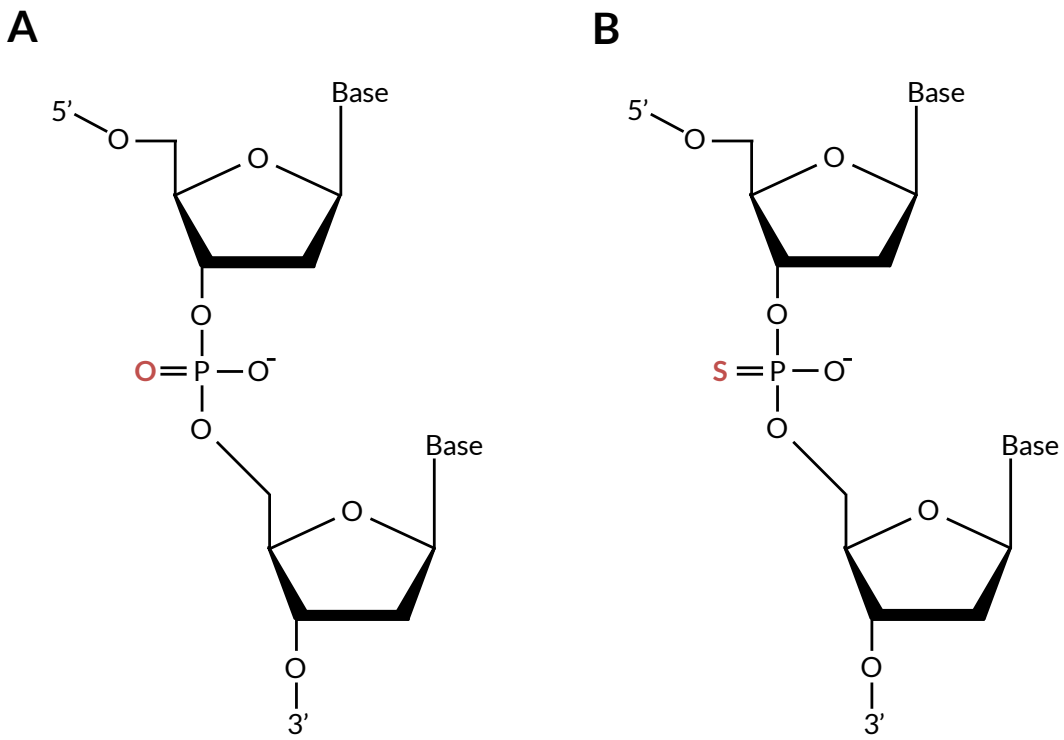

Figure 3.1 Chemical structure of (A) DNA, containing standard phosphodiester bond between two neighbouring nucleosides, and (B) DNA containing phosphorothioate bond. The difference in chemical structure is clearly marked in red

Use of phosphorothioate oligonucleotides could potentially provide such opportunity. Replacement of a usual phosphodiester internucleotide linkage with a phosphorothioate bond (Figure 3.1), constitutes one of the most common type of oligonucleotide modification, used most frequently to make a given oligonucleotide resistant to the action of nucleases. However, phosphorothioate bonds have got another interesting characteristic – they are susceptible to iodine cleavage in alkaline solutions (Gish and Eckstein, 1988). This property has been utilised in the development of several molecular tools, including mutagenesis (Dennig et al., 2011; Wong et al., 2004b) and gene cloning (Blanusa et al., 2010).

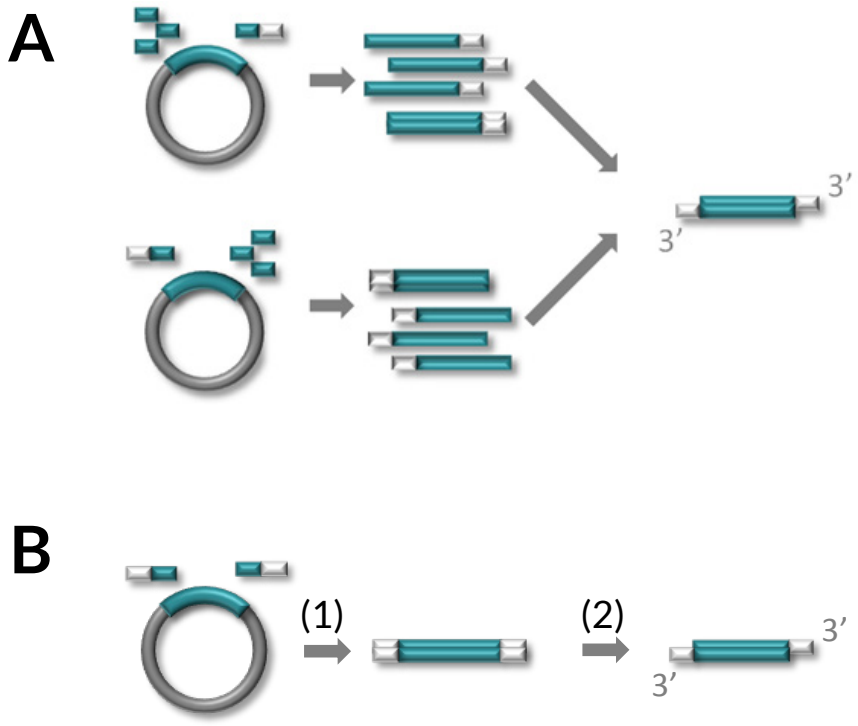

Figure 3.2 Comparison of the megaprimer synthesis stage between (A) original QuickStep-Cloning and (B) improved version of the method, utilising phosphorothioate oligonucleotides, where (1) denotes a standard PCR with phosphorothioate oligonucleotides and (2) represents 5 min incubation at 70°C in 10 mM iodine.

The incorporation of phosphorothioate oligonucleotides into QuickStep-Cloning should significantly streamline the whole procedure (Figure 3.2). In the new version of the method, called QuickStepS-Cloning, the two asymmetric PCRs are replaced with a single PCR and subsequent 5 min incubation of the PCR product in 10 mM iodine and 10% ethanol, at 70°C – as described in (Blanusa et al., 2010).

It should be also noted that, in QuickStep-Cloning, the two asymmetric PCRs produce not only the desired product but also a certain fraction of megaprimer pairs that are completely complementary (identically, to the megaprimer design for RF cloning). This fraction of unwanted side-products should be much lower when the strategy involving phosphorothioate oligonucleotides is utilised. As a result, it is envisaged that the new method design will not only expedite and simplify the whole process but it will also increase the overall cloning efficiency.

3.2 Results and Discussion

3.2.1 Optimisation

QuickStepS-Cloning was optimised based on *rfp* cloning experiment, described in Chapter 2, where *rfp* gene from Kan^r-pBbA8k-RFP (Figure 2.10) is cloned into Kan^r-pET24a-HLTEV-p53.

After the first PCR with phosphorothioate oligonucleotides, 6.25 µl of 0.5 M Tris-HCl buffer (pH 9) and 6.25 µl 100 mM iodine in absolute ethanol were added to 50 µl of the PCR product. After mixing, the resultant mixture was incubated for 5 min at 70°C and then snap cooled on ice to prevent the cleaved DNA fragments from properly re-annealing to the uncovered 3' overhangs. After a quick PCR purification (using a silica-based spin column), the megaprimer PCR was conducted as described in Chapter 2.

During the method optimisation, the influence of different parameters on the yield of QuickStepS-Cloning was investigated (Figure 3.3). Firstly, it was shown that there is no discernible difference between 5 minutes and 2 h incubation with iodine, showing that 5 min incubation is sufficient for cleavage of phosphorothioate bonds to go to completion (Figure 3.3A). It was also investigated whether there is a need for an additional purification after the first PCR, however, it was shown that the addition of the extra step only barely slightly the final yield (Figure 3.3B).

Finally, similarly to Chapter 2, the influence of megaprimer concentration, number of PCR cycles and vector concentration was thoroughly examined (Figure 3.3C-E). It is worth noting that in the case of 35 cycles, the increased amount of the product, visible

on the agarose gel, did not translate into higher cloning efficiency. This was most likely caused by a high amount of side-products that interfered with the subsequent chemical transformation.

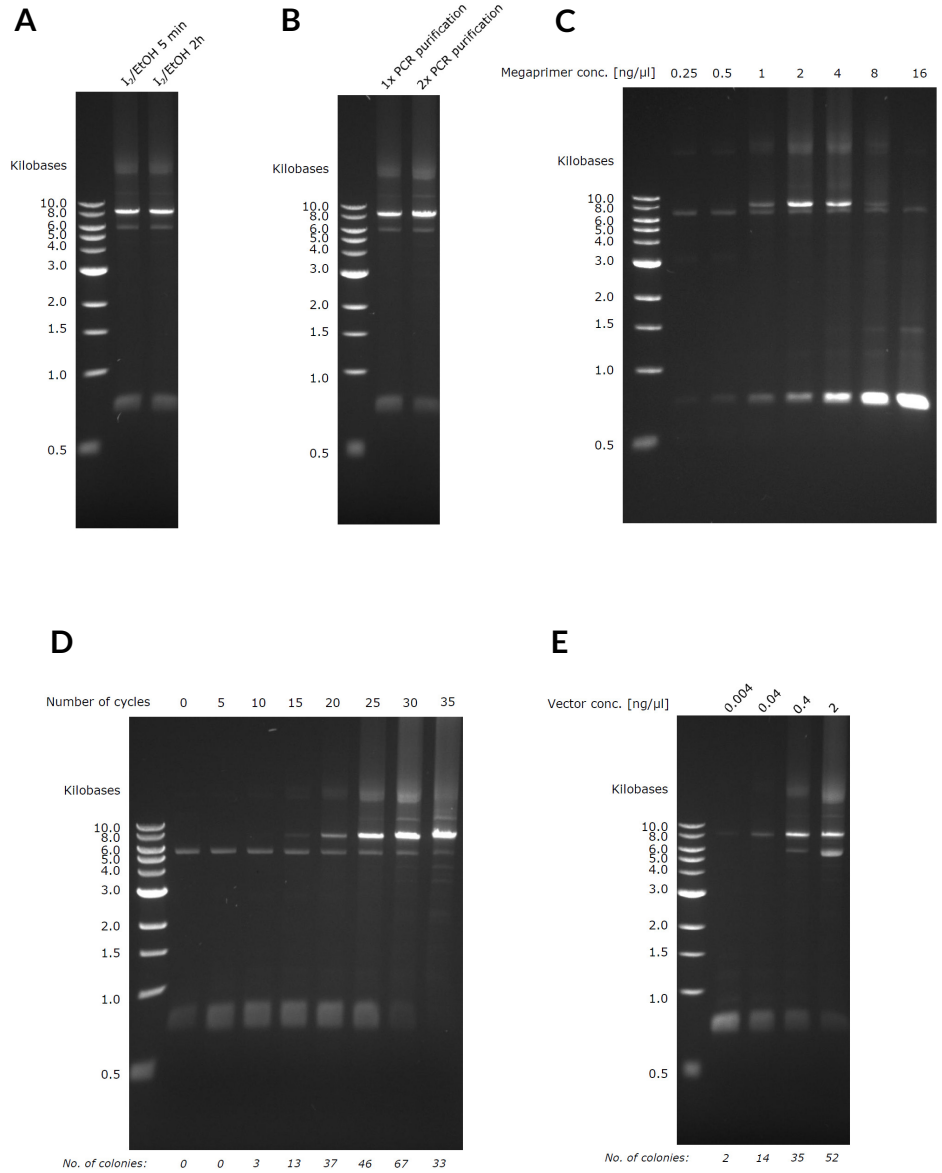

Figure 3.3 Optimisation of QuickStepS-Cloning. Yield of megaprimer PCR: (A) for different durations of incubation with iodine, (B) with and without additional purification stage before the incubation with iodine, (C) for different concentrations of megaprimer (D) different number of cycles of megaprimer PCR and (E) vector concentration. The band at 8 kb represents the product, whereas the band at 6 kb shows the template plasmid.

3.2.2 Method comparison

The optimised protocol was compared with RF cloning and the previous version of the method (Table 3.1 and Figure.3.4).

Table 3.1 Comparison of QuickStepS-Cloning, using 25 or 30 cycles of megaprimer PCR, with RF cloning and QuickStep-Cloning, as determined by *rfp* cloning experiment. Colony counts for *E. coli* C41 (DE3) transformed with the products of aforementioned cloning methods and plated on agar plates supplemented with: 50 μg/ml kanamycin and 1 mM IPTG. Transformation efficiency was determined based on concurrent transformation of 1 ng intact pET24a-HLTEV-p53 plasmid. Numbers in the brackets denote RFP-expressing colonies, as determined by visual inspection of the plates.

QuickStep-Cloning	RF Cloning	QuickStepS-Cloning (25c)	QuickStepS-Cloning (30 c)	Transformation efficiency [cfu/μg]
9(9)	0	19(16)	43(41)	$2 \cdot 10^4$

The presented results clearly show that QuickStepS-Cloning exhibits superior performance in comparison to the other two method, resulting in 2-4 fold improvement (depending on the number of cycles used for megaprimer PCR) in cloning efficiency in comparison to QuickStep-Cloning. At the same time, the percentage of RFP-positive colonies remains exceptionally high. Surprisingly, RF cloning did not produce any clones. This was likely caused by a low transformation efficiency achieved during this experiment. A higher transformation efficiency would probably be needed to observe any colonies produced by RF cloning.

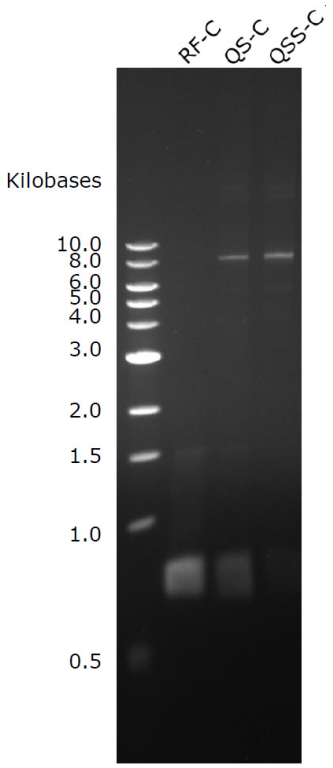

Figure 3.4 Yield of RF cloning, QuickStep-Cloning and QuickStepS-Cloning (25 cycles of megaprimer PCR), as determined for *rfp* cloning experiment.

3.2.3 Method validation

To prove that the developed protocol is applicable to other cloning experiments, without the need of further optimisation, QuickStepS-Cloning was used to clone P450 BM3 reductase gene from pCWori-BM3 plasmid into pETM11-BM3 WT plasmid (Figure 3.5). This cloning can be classified as relatively challenging due to high lengths of both the cloned gene and the recipient plasmid – 2 and 7 kb respectively (usually, the overall number of transformants provided by a cloning process significantly decreases when longer DNA fragments are being combined). To act as a comparison, RF cloning was performed in parallel.

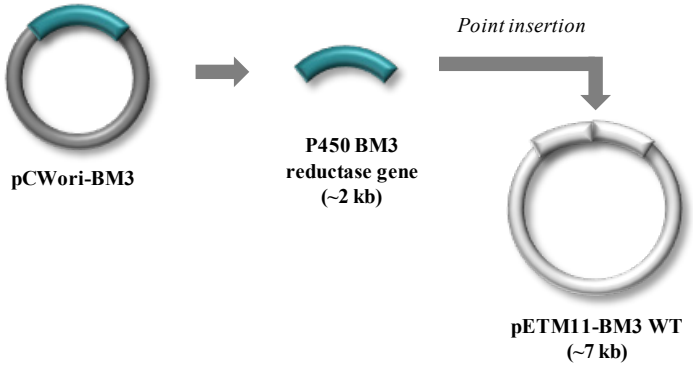

Figure 3.5 Outline of P450 BM3 reductase gene cloning experiment.

Owing to high length of the cloned gene and the recipient plasmid, the amount of resultant DNA was much lower (Figure 3.6) in comparison to the *rfp* cloning experiment (Figure 3.4).

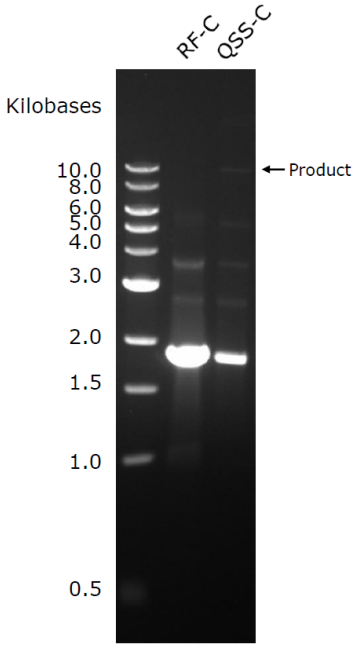

Figure 3.6 Yield of RF cloning and QuickStepS-Cloning, as determined for P450 BM3 reductase gene cloning experiment.

Despite the low amount of recombinant DNA produced, as determined by the gel electrophoresis, a significant number colonies were obtained when *E. coli* was transformed with the product of QuickStepS-Cloning. Five of the 49 available colonies

were picked at random and the presence of the insert was confirmed in all 5 of them via a colony PCR. Again, RF cloning did not produce any colonies, which this time was caused by both the insufficient transformation efficiency and the complexity of the molecular cloning involved. Consequently, it was shown that the developed method performs much better than the state-of-the-art method.

Table 3.2 Results of P450 BM3 reductase gene cloning experiment. Colony counts for *E. coli* strains DH5α transformed with the products of RF cloning and of QuickStep-Cloning and plated on agar plates supplemented with 50 µg/ml kanamycin. Transformation efficiency was determined based on concurrent transformation of 1 ng intact pETM11-BM3 WT plasmid.

QuickStepS-Cloning	RF Cloning	Transformation efficiency [cfu/µg]
49	0	$4 \cdot 10^4$

3.2.4 Application to directed evolution

To make use of the exceptionally high cloning efficiency of QuickStepS-Cloning, the applicability of the method to protein engineering was investigated.

Firstly, *rfp* cloning experiment was performed once again using QuickStepS-Cloning, however, this time, the mutagenic were incorporated into the first PCR – *rfp* gene was amplified using Taq polymerase (NEB) in the presence of imbalanced concentration of dNTPs and high $MnCl_2$ concentration.

In parallel, MEGAWHOP was used to perform similar mutagenesis and, subsequently, integrate the resultant gene library into the plasmid. It should be noted that

MEGAWHOP differs slightly from RF cloning and MEGAWHOP cloning, discussed in Chapter 2. In the MEGAWHOP protocol the recipient plasmid is replaced with a vector that already contains the gene to be mutagenised (Miyazaki and Takenouchi, 2002). As such, normal cloning needs to be performed in advance in order to utilise the method. The trade-off is significantly higher number of transformants produced by the method, due to a much better annealing of the megaprimer (the annealing is no longer limited to the terminal regions of the megaprimer). As a result, MEGAWHOP is a popular method of choice for many directed evolution experiments where a large library is needed.

E. coli C41 (DE3) was transformed with a purified product of both QuickStepS-Cloning and MEGAWHOP via electroporation. The results of the experiment are presented in Table 3.3. As shown, the QuickStep-Cloning produces a combinatorial library of a similar size to MEGAWHOP, without the need of performing prior cloning (utilising a plasmid without the target gene as a template)

Table 3.3 Comparison of the library size when performing an experiment involving *rfp* mutagenesis via QuickStepS-Cloning and MEGAWHOP. Colony counts for *E. coli* C41 (DE3) transformed with the products of QuickStepS-Cloning and MEGAWHOP via electroporation and plated on agar plates supplemented with 50 μg/ml kanamycin and 1 mM IPTG. Transformation efficiency was determined based on concurrent transformation of 1 pg intact pET24a-HLTEV-p53 plasmid.

QuickStepS-Cloning	MEGAWHOP	Transformation efficiency [cfu/μg]
$1.0 \cdot 10^4$	$1.5 \cdot 10^4$	$2 \cdot 10^6$

To further demonstrate the usefulness of QuickStepS-Cloning to directed evolution, three protein variants exhibiting a different colour under visible light were isolated during the described experiment (Figure 3.7). The difference in colour was further confirmed by absorbance measurements (Figure 3.8). The genes encoding for the three proteins were sequenced and associated mutations were identified .(M1: T195S; M2: F91L, M3:Q66L and T202A).

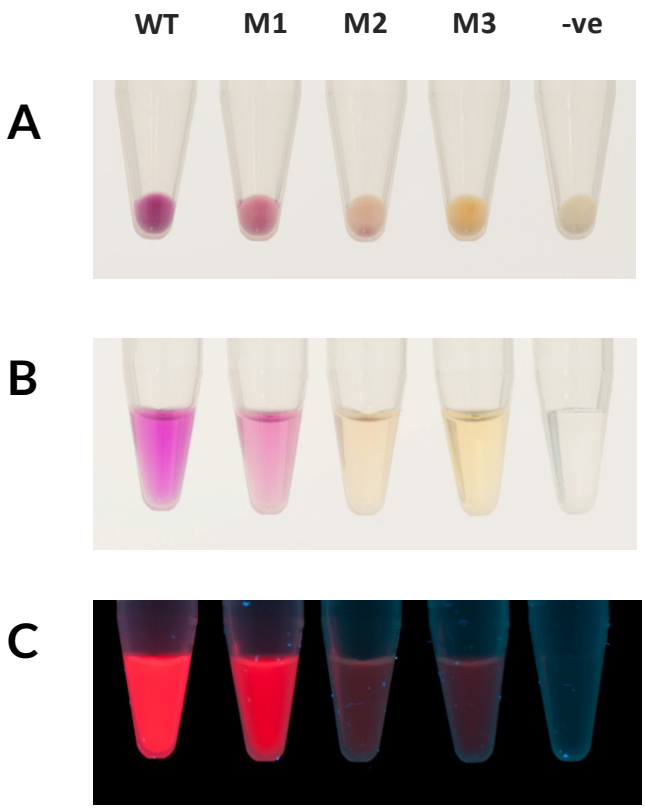

Figure 3.7 Comparison of the wild-type RFP and three isolated mutants: (A) cell pellets from the expressing *E. coli* C41 (DE3), (B) purified protein (purification achieved with Ni-NTA Spin Columns; Qiagen), and (C) purified protein under UV light. (-ve) denotes negative control: *E. coli* C41 (DE3) without a plasmid in (A) and PB buffer in (B) and (C).

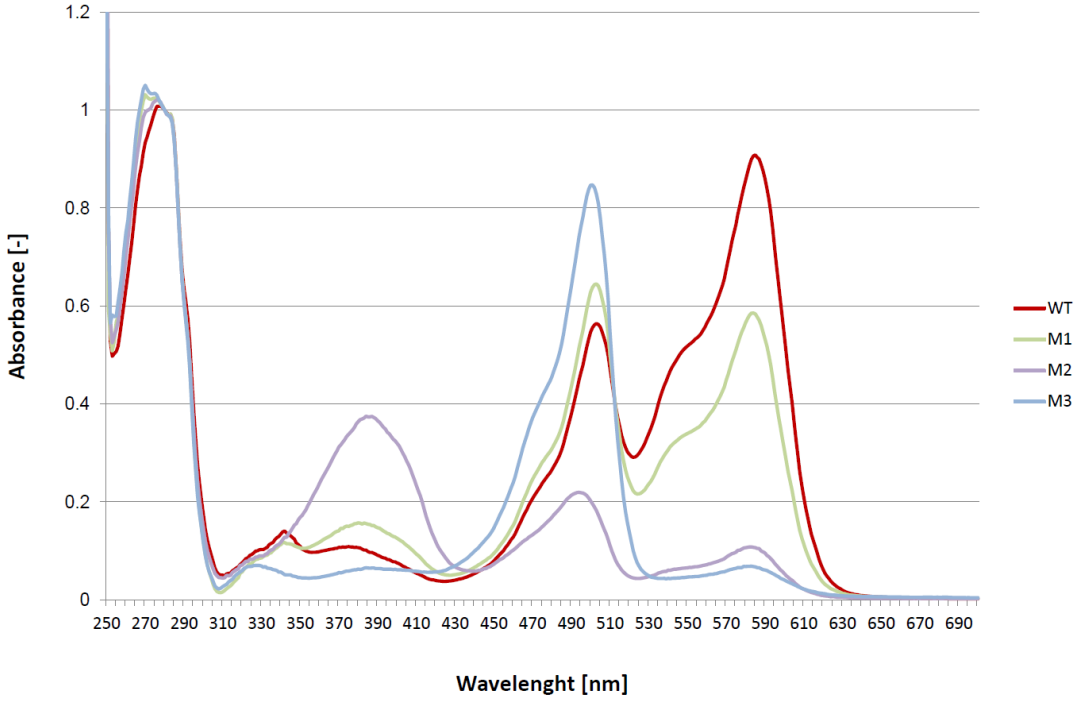

Figure 3.8 Absorbance spectra of the wild-type RFP and three isolated mutants. Proteins were purified using Ni-NTA Spin Columns; (Qiagen).

3.3 Conclusion

QuickStepS-Cloning utilises phosphorothioate oligonucleotides to improve the cloning method presented in Chapter 2. By targeting the identified bottleneck of megaprimer synthesis, the new protocol not only expedites and simplifies the whole procedure of gene cloning but also achieves significantly higher cloning efficiency. It is demonstrated that QuickStepS-Cloning cloning can compete with even the most efficient methods of incorporating combinatorial libraries such as MEGAWHOP. Finally, the usefulness of the method to protein engineering is clearly demonstrated by incorporating random mutagenesis into method and performing a proof-of-concept directed evolution experiment. The same approach can also be used for fine tuning gene expression by quickly incorporating a library of mutagenised promoters into a biological circuit.

All in all, QuickStepS-Cloning should be of great interest to synthetic biology because not only it facilitates the construction of biological circuits (expanding the capabilities and the throughput of the previous method) but also, when used in combination with directed evolution, it can expedite the process of acquiring new, well-defined biological parts.

3.4 Methods

3.4.1 Materials

All enzymes, deoxyribonucleotides and DNA ladders were purchased from New England Biolabs (Ipswich, USA).

3.4.2 Primers

Primers used in this study were synthesized by Eurofins Genomics (Ebersberg, Germany). Melting temperatures of oligonucleotides were determined using the New England Biolabs T_m Calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator). Six primers, in total, depending on the method of choice were used for *rfp* cloning experiment. *RFP-Fwd* (5'-ATGGCGAGTAGCGAAGACG-3', 19 bp), *IntA-RFP-Fwd* (5'-CGAAAACCTGTACTTCCAGGGTGGATCCATGGCGAGTAGCGAAGACG-3', 47 bp), *RFP-Rev* (5'-TTAAGCACCGGTGGAGTGACG-3', 21 bp), *IntB-RFP-Rev* (5'- CTAGGATCTGACTGCGGCTCCTCCATTTAAGCACCGGTGGAGTGACG-3', 47 bp), *IntA-RFP-Fwd-P* (5'- CGAAAACCTGTACTtCCAGGGTGGATCCaTGGCGAGTAGCGAAGACG-3', 47 bp) and *IntB-RFP-Rev-P* (5'- CTAGGATCTGACTgCGGCTCCTCCATtTAAGCACCGGTGGAGTGACG-3', 47 bp). The small letters in the DNA sequence signify the location of the phosphorothioate bonds.

Underlined parts of *IntA-Fwd* and *IntB-Rev* are identical to *Fwd* and *Rev* primers, respectively, and the remaining parts correspond to the two megaprimer annealing sites flanking DNA insertion point present in pET24a-HLTEV-p53.

The following two primers were used for P450 BM3 reductase gene cloning experiment: *P450_IntB_Rev* (5'- GTCGACGGAGcTCGAATTCTTaCCCAGCCCACACGTCT TTTGC -3', 43 bp) and *P450_IntA_Fwd* (5'- GGTAAAAGCAAAATcGAAAAAAATT CCGCTtGGCGGTATTCCTTCACCTAGCACTG -3', 56 bp). The small letters in the DNA sequence signify the location of the phosphorothioate bonds.

3.4.3 QuickStep-Cloning

To clone *rfp* gene from pBbA8k-RFP into pET24a-HLTEV-p53 plasmid, two asymmetric PCRs were carried out in parallel. Asymmetric PCR mixture I (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *RFP-Fwd* primer, 10 nM *IntB-RFP-Rev* primer, 0.2 ng pBbA8k-RFP, and 1 U Q5 High-Fidelity DNA Polymerase. Asymmetric PCR mixture II (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 10 nM *IntA-RFP-Fwd* primer, 500 nM *RFP-Rev* primer, 0.2 ng pRGP and 1 U Q5 High-Fidelity DNA Polymerase. Both mixtures were thermocycled using the following conditions: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 68°C and 30 s extension at 72°C. The two PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and their DNA concentration was determined using VersaWave (Expedeon). For megaprimer PCR, the mixture (50 µl) contained 1× Q5 Reaction Buffer, 200 µM of each dNTP, 200 ng of purified asymmetric PCR product I, 200 ng of purified asymmetric PCR product II, 20 ng pET24a-HLTEV-p53 and 1 U Q5 High-Fidelity DNA Polymerase. The mixture was thermocycled according to the following program: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min

annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were subsequently added to the PCR mixture and incubated at 37°C for 15 min to remove the parental pET24a-HLTEV-p53 plasmids.

3.4.4 Restriction-free (RF) cloning

PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *IntA-RFP-Fwd* primer, 500 nM *IntB-RFP-Rev* primer, 0.2 ng pBbA8k-RFP, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the same program as asymmetric PCR in QuickStep-Cloning: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 68°C and 30 s extension at 72°C. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon). Megaprimer PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 nM of each dNTP, 400 ng purified PCR product, 20 ng pET24a-HLTEV-p53 and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled with the following conditions: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were added to the PCR mixture and incubated at 37°C for 15 min to remove the parental pET24a-HLTEV-p53 plasmids.

To clone P450 BM3 reductase gene from pCWori-BM3 into pETM11-BM3 WT plasmid, the same protocol was followed, using a dedicated primer set (*P450_IntA_Fwd* and *P450_IntB_Rev*) and, for the first PCR, the following thermocycling conditions

were used: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 80 s annealing and extension at 72°C.

3.4.5 QuickStepS-Cloning

PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *IntA-RFP-Fwd-P* primer, 500 nM *IntB-RFP-Rev-P* primer, 0.2 ng pBbA8k-RFP, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the same program as asymmetric PCR in QuickStep-Cloning: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 68°C and 30 s extension at 72°C.

6.25 µl of 0.5 M Tris-HCl buffer (pH 9) and 6.25 µl iodine in absolute ethanol were added to 50 µl of the PCR product. After brief mixing by pipetting, the resultant mixture was incubated for 5 min at 70°C and then snap cooled on ice.

Subsequently, it was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon). Megaprimer PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 nM of each dNTP, 400 ng purified PCR product, 20 ng pET24a-HLTEV-p53 and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled in the same conditions as QuickStep-Cloning megaprimer PCR: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were added to the PCR mixture and incubated at 37°C for 15 min to remove the parental pET24a-HLTEV-p53 plasmids.

To clone P450 BM3 reductase gene from pCWori-BM3 into pETM11-BM3 WT plasmid, the same protocol was followed, using a dedicated primer set (*P450_IntA_Fwd* and *P450_IntB_Rev*) and, for the first PCR, the following thermocycling conditions were used: (i) 30 s initial denaturation at 98°C and (ii) 30 cycles of 7 s denaturation at 98°C, 80 s annealing and extension at 72°C.

3.4.6 DNA gel electrophoresis

PCR products were analyzed using either 0.7% agarose gel. For visualisation purposes, ethidium bromide was added to the gel. Quick-Load 1 kb DNA Ladder was used as a DNA ladder.

3.4.7 Chemical transformation and clone analysis

E. coli DH5α and C41 (DE3) were transformed with 5 µl of DpnI-digested products of QuickStep-Cloning, QuickStepS-Cloning or RF cloning, using a standard chemical transformation protocol (Minagawa et al., 2007). Concurrently, the bacterial strains were transformed with 1 µl of 1 ng/µl intact pET24a-HLTEV-p53 or pETM11-BM3 WT to estimate transformation efficiency. Transformed bacteria were plated on *TYE* agar plates (10 g/l tryptone, 5 g/l yeast extract, 8 g/l sodium chloride and 15 g/l agar) supplemented with: 50 µg/ml kanamycin and 1 mM IPTG (in the case of *rfp* cloning experiment). The plates were incubated overnight at 37°C and for further 12 h at 30°C. The number of RFP-expressing colonies was determined by visual inspection using UV transilluminator.

3.4.8 Error-prone PCR

PCR mixture (50 µl) containing 1× Standard Taq (Mg-free) reaction buffer, 7mM $MgCl_2$, 0.05 mM $MnCl_2$, 200 µM of dATP, 200 µM of dGTP, 1 mM of dGTP, 1 mM of dCTP 400 nM *IntA-RFP-Fwd-P* or *IntA-RFP-Fwd* primer (depending whether QuickStepS-Cloning or MEGAWHOP is being used, respectively) library, 400 nM *IntB-RFP-Rev-P* or *IntB-RFP-Rev* primer, 50 ng pBbA8k-RFP, and 1.25 U Taq DNA Polymerase was thermocycled using the same program as asymmetric PCR in QuickStep-Cloning: (i) 30 s initial denaturation at 95°C and (ii) 30 cycles of 20 s denaturation at 95°C, 30 s annealing at 55°C and 45 s extension at 68°C, (iii) final extension at 68°C for 5 min. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon).

3.4.9 MEGAWHOP

Megaprimer PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 nM of each dNTP, 500 ng megaprimer, 50 ng pET24a-HLTEV-RFP-p53 and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled in the following conditions: (i) 30 s initial denaturation at 98°C, (ii) 25 cycles of 10 s denaturation at 98°C, 4 min annealing and extension at 72°C, and (iii) 2 min final extension at 72°C. Forty units of DpnI were added to the PCR mixture and incubated at 37°C for 15 min to remove the parental plasmids.

3.4.10 Transformation of *rfp* library

The product of QuicStepS-Cloning and MEGAWHOP were purified using QIAquick PCR Purification Kit. 1 µl of the purified mixture was used to transform *E. coli* DH5α cells using a standard protocol. After the transformation, the cell were plated at different dilution on LB agar plates (10 g/l tryptone, 5 g/l yeast extract, 10 g/l sodium chloride and 15 g/l agar) supplemented with: 50 µg/ml kanamycin and 1 mM IPTG (in the case of *rfp* cloning experiment). The plates were incubated overnight at 37°C and for further 12 h at 30°C. Three colonies exhibiting a slightly different colour under visible light were identified by visual inspection and were grown overnight at 37°C in 5 ml 2×TY media (16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl) supplemented with 50 µg/ml kanamycin.

3.4.11 Protein expression and purification

250 µl of overnight cultures of *E. coli* corresponding to three isolated mutants and wild-type RFP were used to inculcate 50 ml LB media supplemented with kanamycin and grown at 37°C with shaking. When OD_{600} reached 0.6, the expression was induced with 1mM IPTG. The cells were grown for further 16 h. The protein was isolated using Ni-NTA Spin Columns (Qiagen), according to the manufacturer's instructions. The absorbance measurements were conducted using UV-1600PC UV-Vis spectrophotometer (VWR).

Chapter 4
Multi-Genius:
Method development

4.1 Method outline

As discussed in section 1.5, there exist significant limitations and lacking features in the area of identifying genes encoding for useful phenotypes. To provide a robust alternative to existing methods, a new approach to simultaneous targeting of multiple genes is being proposed (Figure 4.1). In this method, DNA fragments from two genomic libraries are integrated into a vector containing two different inducible promoters (Figure 4.2). This should allow for identification of interactions between distant genes and significantly help in evolving organisms with a desired phenotype.

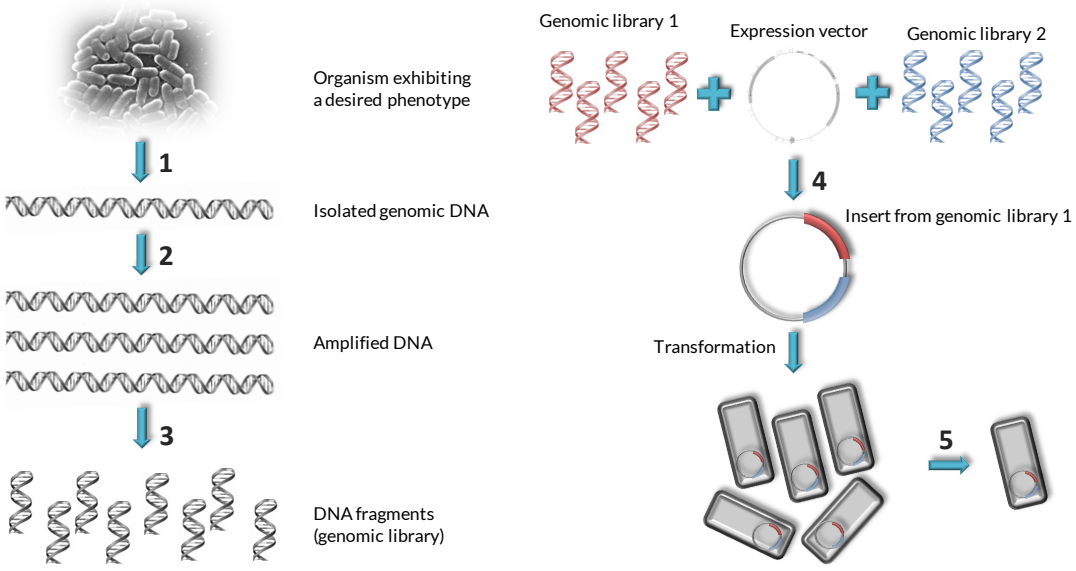

Figure 4.1 Overview of Multi-Genius: (1) genomic DNA extraction, (2) amplification of genomic DNA, (3) fragmentation, (4) integration of two genomic libraries into one plasmid containing two inducible promoters, and (5) selection.

In the proposed method, the first step is to isolate genomic material, by extracting it from an organism of interest and purifying it. The isolated DNA is then amplified and divided into smaller fragments. Fragmentation can be achieved through a variety of

different methods, each providing their own advantages and drawbacks. Next, the DNA fragments from each genomic library are cloned into the expression vector, so that each plasmid contains two fragments, each from a different libary. A proper cloning technique needs to be carefully chosen as this step is likely to constitute a primary bottleneck for the whole method. Resultant vectors are then transformed into a bacterial species that needs to be modified. Transformed cells are then incubated in selective conditions – for example high temperature or low pH. Plasmids can be then isolated from the selected cells and analysed to identify genes coding for useful phenotypes. Use of two inducible promoters allows for selection in [0,0], [0,1], [1,0] and [1,1] configurations making the gene identification much simpler. What is more, the use of two independent inducible promoters might prove invaluable in determining the optimal expression levels of a pair of identified genes. Namely, once a beneficial gene combination is identified, the screening can be performed in the presence of selective conditions and varying concentration of the two inducers (microtiter plates or gradient agar plates might utilised for this purpose). The optimal concentrations of the two inducers can be found in this way.

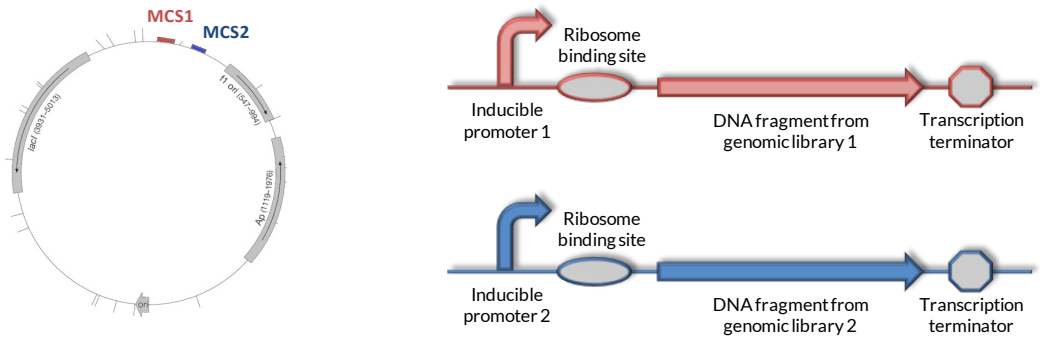

Figure 4.2 Design of expression vector for Multi-Genius. MSC stands for multiple cloning site.

There are several challenges that will need to be overcome to utilise the proposed method. Many vital and complex design decisions will need to made to ensure a proper operation of the method. For example, finding a suitable method for efficient integration of the DNA fragments from two genomic libraries into the expression vector will be a major challenge. Finally, a space of possible two-fragment combinations is so great that it will be vital to optimise the experimental protocol so that the size of library that can be screened with Multi-Genius is maximised. It should be also remembered that genomic DNA will be randomly fragmented causing some or even most of the them to be non-functional

Despite the listed challenges, this project offers many potential benefits. The key aim is to expand upon the concept of genomic libraries and develop a relatively simple, robust and cost-effective method for identifying interactions between distant genetic loci. The proposed method is of great scientific and industrial significance as it has a potential for evolving useful phenotypes in industrially-relevant bacterial strains.

4.2 Genomic DNA extraction

For the purpose of genomic DNA extraction, a standard Bacterial DNA Kit from OMEGA Bio-Tek was used. Genomic DNA was extracted from 10^9 *E. coli* cells, according to manufacturer's instructions, and its concentration and purity was analysed using VersaWave spectrophotometer (Table 9). The experimental results show that the selected kit provided DNA of sufficient concentration and quality for the proposed application. During the experiments, it was determined, however, that when more than 10^9 cells are used (e.g., 2 ml of the overnight culture) the HiBind DNA spin-column becomes oversaturated and the quality of obtained DNA is very poor (data not shown). As a result, it is imperative to measure cell density before genomic DNA isolation to make sure that the recommended number of cells is not exceeded when performing the DNA extraction.

Table 4.1 Concentration and purity of genomic bacterial DNA extracted from an overnight culture of *E. coli* DH5α and C41(DE3).

Bacterial strain	Final elution volume [µl]	Concentration [ng/µl]	260/280 ratio [-]	260/230 ratio [-]
E. coli DH5α	100	18.4	1.82	1.96
E. coli C41(DE3)	100	52.5	1.89	2.16

4.3 DNA amplification

The purpose of DNA amplification step is expand the applicability of Multi-Genius to situations when only minute amounts of genomic material are available – this might happen, for example, when dealing with microorganisms that are difficult to culture under laboratory conditions.

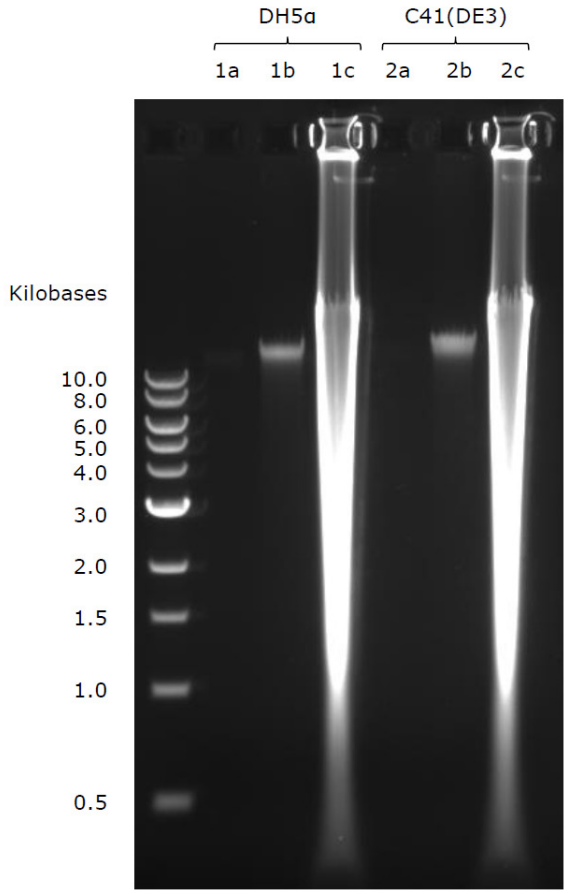

Figure 4.3 Product of genomic DNA amplification. Genomic DNA extracted from DH5α (denoted as 1) and C41(DE3) (denoted as 2) – (a) before amplification and at concentration of 0.5 ng/µl – initial concentration of DNA at the start of amplification reaction, (b) before amplification and at concentration of 10 ng/µl, (c) after amplification.

Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit was selected for DNA amplification. In order to test its performance, the genomic DNA from *E. coli* strains C41(DE3) and DH5α was amplified, according to manufacturer's instructions, and the products of the amplification reactions were analysed using gel electrophoresis (Figure 4.3). The results show that using the discussed kit and the corresponding experimental protocol, high concentrations of amplified genomic DNA can be obtained from very small amounts of template DNA (10 ng).

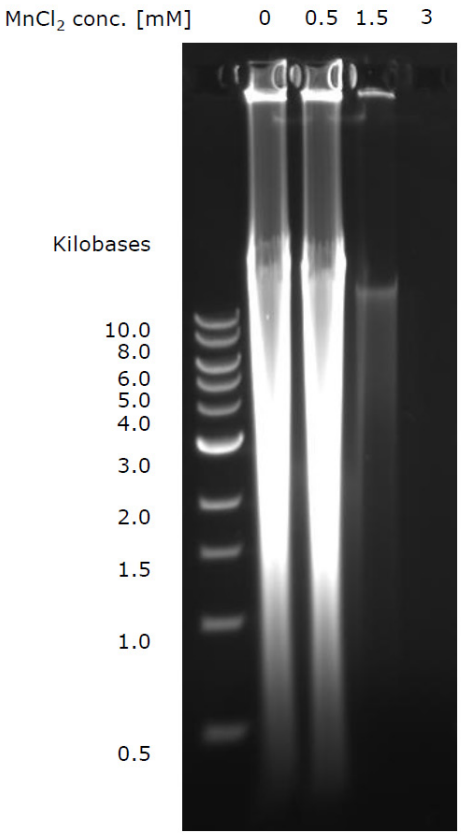

Figure 4.4 Yield of genomic DNA amplification carried out at different concentrations of $MnCl_2$. Genomic DNA was extracted from *E. coli* C41(DE3).

It was also briefly investigated whether it is possible to perform random mutagenesis of genomic DNA during amplification. As such, the amplification reaction was carried out in the presence of different concentrations of magnesium chloride, $MnCl_2$. The results show that concentration of 0.5 mM $MnCl_2$ does not have a significant effect on the amplification (Figure 4.4). However, for 1.5 mM $MnCl_2$ noticeable decrease in the total amount of amplification product is observed, implying that Phi29 polymerase (present in Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit) was affected by the presence of the added chemical. Addition of 3 mM $MnCl_2$ completely inhibits the amplification. The obtained results are consistent with results published in scientific literature (Fujii et al., 2004, 2006).

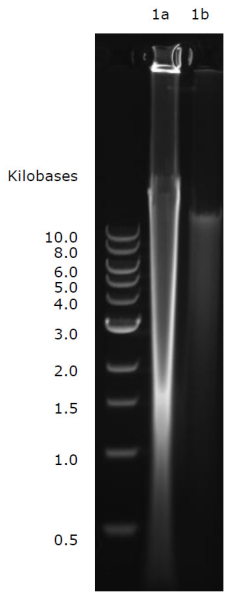

Figure 4.5 Amplified genomic DNA (1a) before and (1b) after purification with QIAquick PCR Purification Kit. Genomic DNA was extracted from *E. coli* C41(DE3).

The applicability of standard column-based DNA purification to amplified genomic DNA was also investigated (Figure 4.5). Significant decrease in the overall product

concentration was observed after purification. This observation was confirmed by DNA concentration measurements using NanoDrop. Based on the experimental results, it was decided that DNA purification using membrane-based columns is not suitable for purification of genomic DNA amplified using illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit and ethanol precipitation has to be used for this purpose.

4.4 DNA fragmentation

The method of choice for fragmentation of genomic DNA is NEBNext dsDNA Fragmentase, enzyme mixture developed and sold by New England Biolabs, capable of generating breaks in double stranded DNA in a time-dependent manner. The main advantage of Fragmentase in comparison to competing methods is its general accessibility – the use of a different method, such as a nebuliser, could significantly limit the overall accessibility of Multi-Genius.

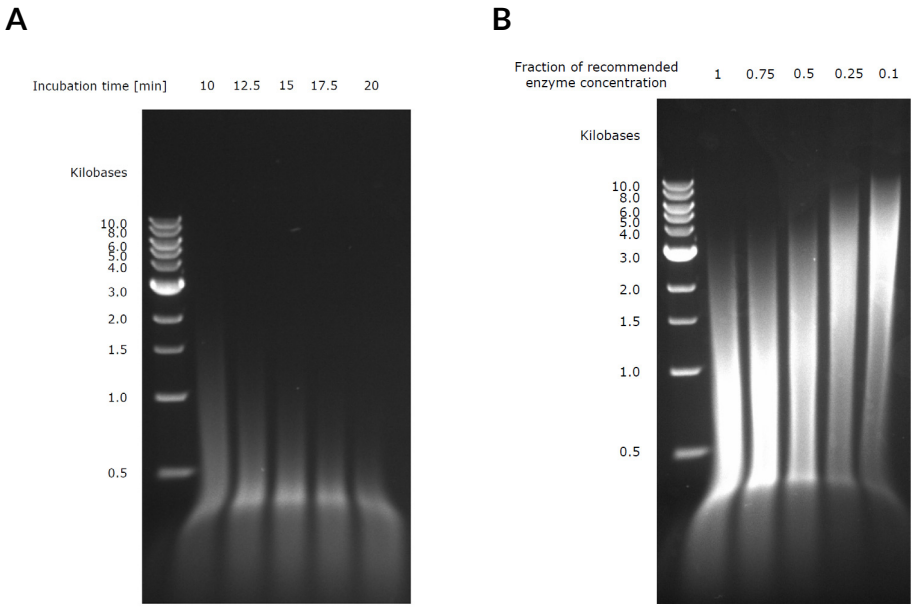

Figure 4.6 Optimisation of DNA fragmentation. Yield of Fragmentase reaction for: (A) varying incubation time, and (B) different enzyme concentration while the incubation is set to 10 min.

According to manufacturer's instructions, genomic DNA was incubated with Fragmentase for varying time (10-20 min) to find optimal reaction conditions for achieving the desired size distribution (1-2 kb). The results show that even when the minimum recommended incubation time of 10 min is used the produced fragments are

too short for the desired application. As using the incubation lower than 10 minutes could decrease the reproducibility of the fragmentation (due to irregularities in a way the reaction is stopped) , it was decided that instead of shortening the incubation time below 10 minutes lower concentration of the enzyme will be used. Based on the conducted experiment, it was demonstrated that 10 min incubation with 0.4-0.5 of the recommended enzyme concentration provides the optimal results.

It should be noted that gel extraction will be employed after genomic DNA fragmentation to isolate only the DNA fragments of the desired length (1-2 kb range). As a result, the observed wide size distribution while certainly not welcomed should not constitute a major problem in the following stages of Multi-Genius.

4.5 Vector design and construction

The expression vector containing two cloning sites with different inducible promoters is, undoubtedly, the cornerstone of the whole project. As such, a lot of thought was put into its design and construction. Several important decision had to made:

- selection marker – kanamycin was chosen as it is the most widely used antibiotic when it comes to protein expression, owing to its reliability and relative stability.

- plasmid copy number – origin of replication conferring a high copy number (pUC; copy number > 400) was selected to ensure that the cells will produce enough plasmid for the susequent genetic manipulation. It should be noted that high copy number can cause a significant metabolic burden when encoded protein is being expressed. However, this effect can be potentially offset by regulating the inducer concentration for the two promoters.

- promoters – the primary characteristic of the two promoters should be their orthogonality, i.e., when one promoter is being induced, the second promoter should not be affected. IPTG-inducible, T_5 promoter and rhamnose-inducible promoter, P_{rham}, are reported to being orthogonal (ATUM, 2017). Looking at the information gathered in section 1.4, the two promoters seem to a suitable choice for the given application.

As the vector with aforementioned design criteria did not exist, it needed to be constructed by combining at least two DNA fragments. The good choice of vector backbone has been pD441-SR, 4 kb vector sold by ATUM that contains T_5 promoter,

kanamycin resistance gene and the desired origin of replication (Figure 4.7). What is more, the vector is delivered in a linearised form, simplifying the subsequent molecular assembly.

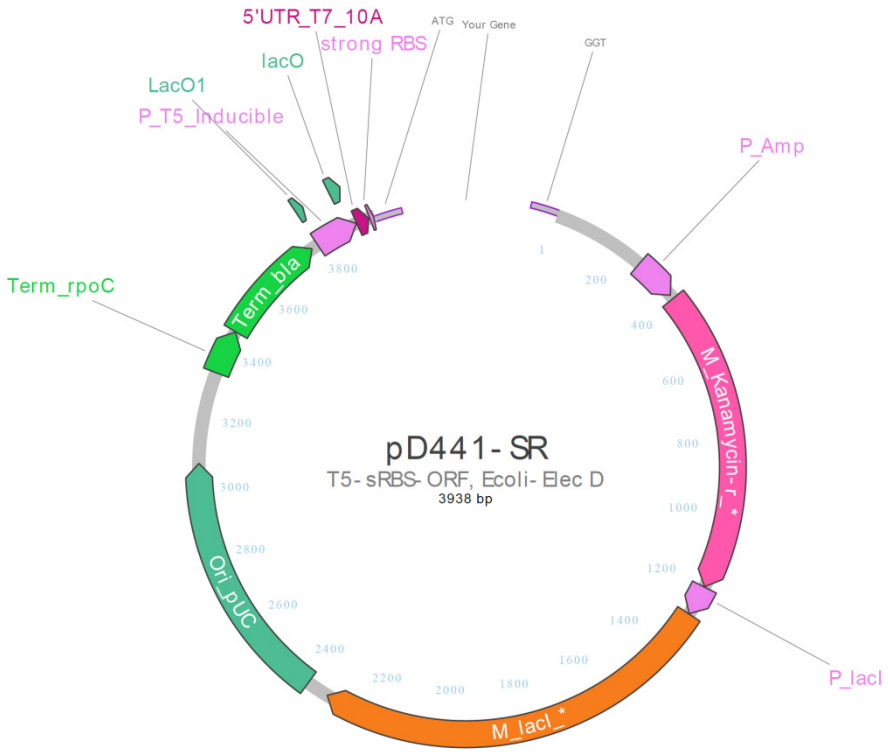

Figure 4.7 Outline of pD441-SR, used as a first building block for the assembly of Multi-Genius expression vector.

The second DNA fragment needs to contain a cloning site under a control of rhamnose-inducible promoter. Utilising the information about Expresso Rhamnose Cloning & Expression System (Lucigen, 2016), such DNA fragment was designed and synthesised by GeneScript. The fragment contained restriction sites complementary to the linearised pD441-SR (Figure 4.8). To perform the molecular assembly, the fragment was digested with corresponding restriction enzyme (SapI) and joined with the pD441-SR vector using traditional ligation and used for subsequent chemical transformation. The

resultant expression vector is presented in Figure 4.9. The successful integration was confirmed by sequencing of the isolated plasmid.

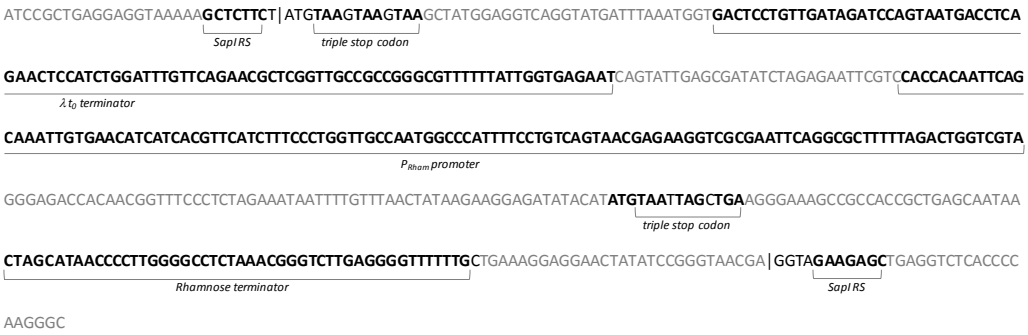

Figure 4.8 530 bp DNA fragment containing the cloning site under a control of rhamnose-inducible promoter, used as the second building block for the assembly of Multi-Genius expression vector.

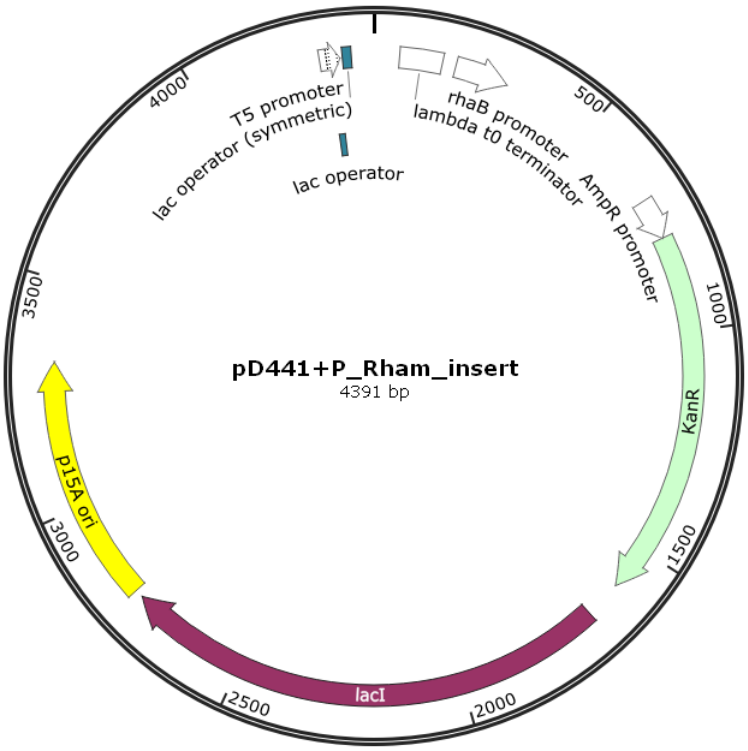

Figure 4.9 Outline of the Multi-Genius expression vector, containing two cloning sites under the control of T_5 and rhamnose-inducible promoter.

4.6 Molecular cloning

The process of optimising molecular cloning was undoubtedly the most time-consuming part of the whole development of Multi-Genius. As the sequence of the integrated DNA fragments is unknown, when cloning genomic libraries, most of the reported cloning method are unsuitable for this application. Also, due to the unknown nature of the cloned DNA sequence, half of the transformants, produced via any of the applicable methods, will contain a gene integrated in a opposite orientation.

At first glance, three main methods of molecular cloning are available: TA cloning, use of DNA adapters and blunt-end cloning. Examining more closely the design of Multi-Genius, TA cloning seem to be unsuitable due to a difficulty of introducing T-overhangs into the second cloning site (once the first fragment has bee integrated). Due to a perceived complexity of utilising the adapters, the blunt-end ligation became the method of choice. Outline of the designed cloning strategy is presented in Figure 4.10.

To assess the functionality of the designed strategy, it was decided that it will be used to clone a pool of EGFP genes into one cloning site and RFP into the second one. In this way, the cloning efficiency should be easy to quantify when the transformed cells are plated on agar plates supplemented with IPTG and rhamnose.

In the early stages of the method optimisation, the fraction of the colonies exhibiting RFP or EGFP fluorescence was less than 1% using a standard blunt-end ligation protocol (the maximum number is 50%).

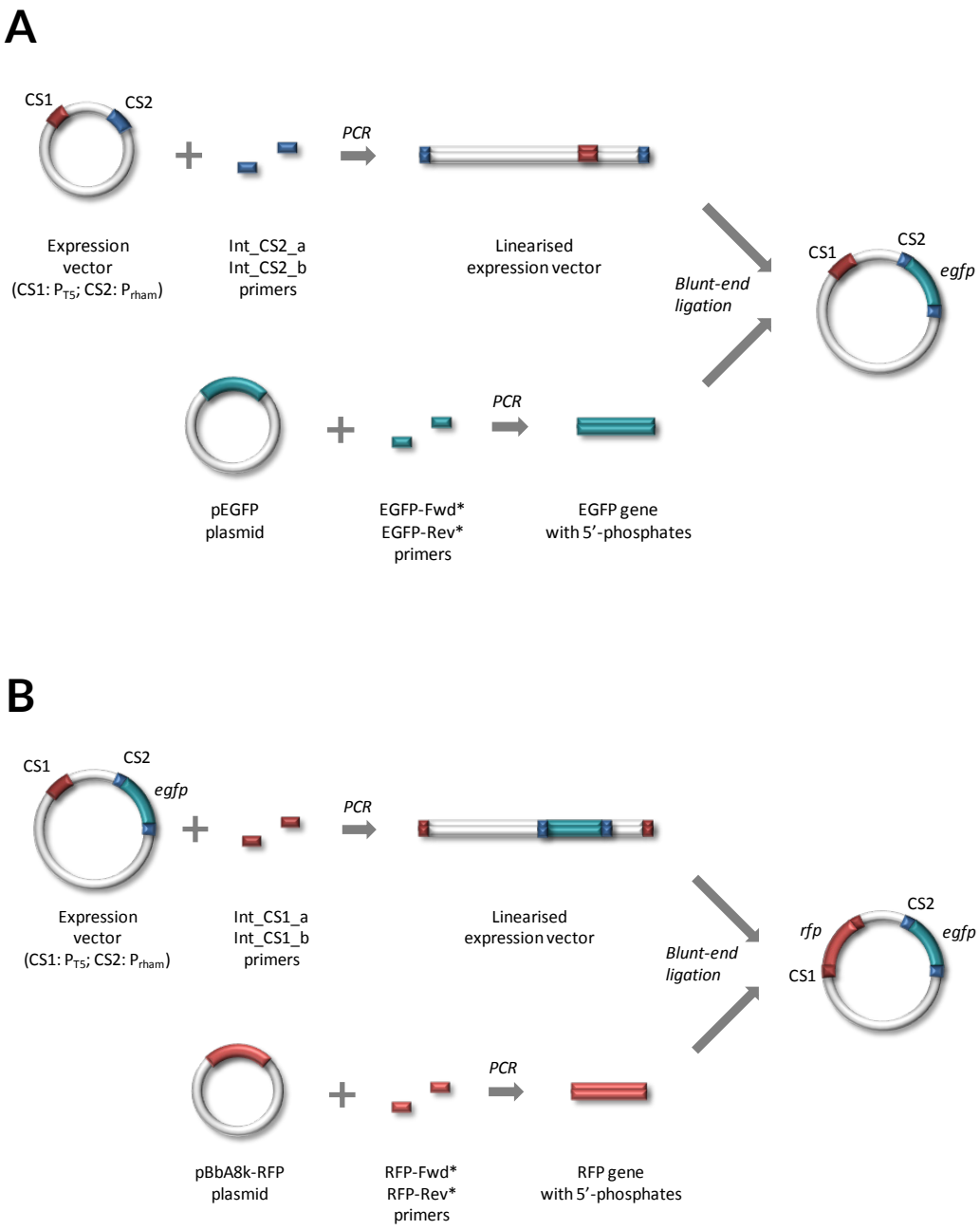

Figure 4.10 Outline of the cloning strategy for Multi-Genius, as shown on the example of (A) integrating *egfp* gene into cloning site number, and (B) subsequent integration of *rfp* gene into cloning site number 1. (*) denotes phosphorylated oligonucleotides.

Various different strategies were tested in order to increase overall cloning efficiency. In the end, four strategies contributed to achieving transformation efficiency close to 50% and increasing the overall reproducibility:

- the number of doublings in the PCRs used to linearise the vector was decreased, by increasing the initial vector concentration and lowering the number of cycles from 25 to 30. This should lower, in the process, the chances of incorporating random mutations into the vector backbone during amplification.
- the primers used for vector linearisation were HPLC purified as truncated oligonucleotides seemingly interfered with subsequent blunt-end ligation.
- the linearised vector was dephosphorylated before the blunt-end ligation. At first glance, the linearised vector should not contain 5'-phosphate as it is a product of a PCR with unmodified oligonucleotides. However, the inclusion of the dephosphorylation step significantly increased the cloning efficiency.
- the reaction parameters of blunt-end ligation were optimised. Some of the important changes to the protocol include: increasing the ligase concentration four times, inclusion of polyethylene glycol in the reaction mixture and increasing the insert and vector concentration two times.

All in all, the initial cloning protocol was significantly altered to increase the overall transformation efficiency. The optimised protocol was utilised to incorporate RFP and EGFP genes into Multi-Genius expression vector.

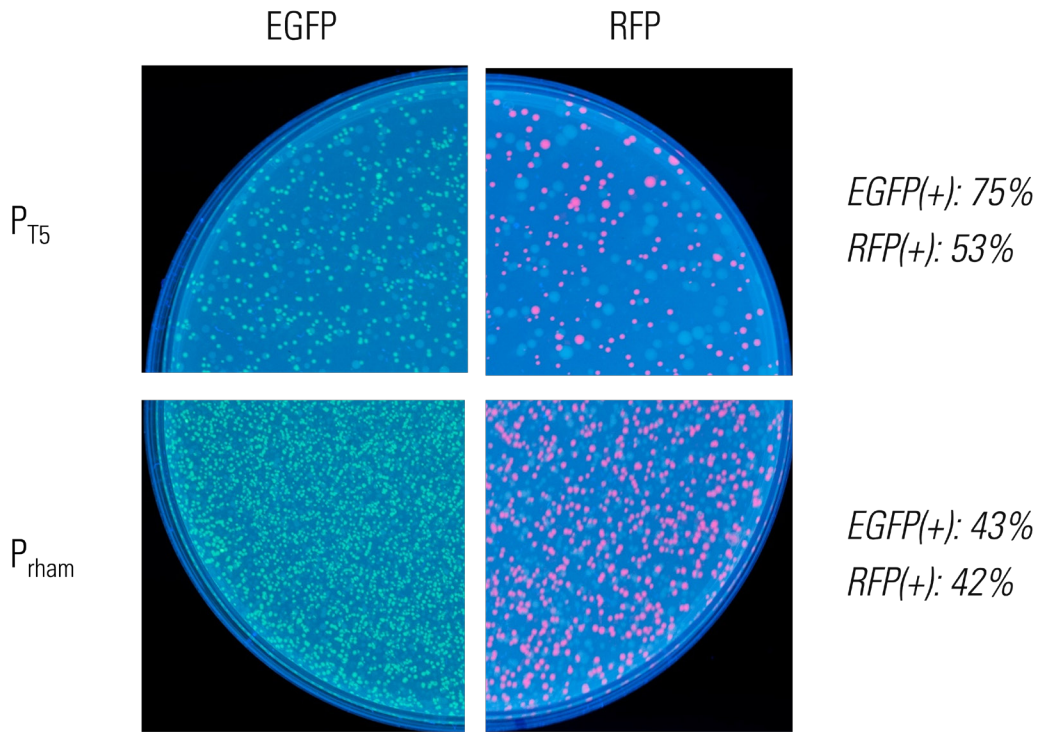

Figure 4.11 The result of applying optimised cloning protocol for incorporating two genomic libraries into Multi-Genius vector via blunt-end ligation; the integration of the first fragment. The RFP and EGFP represent the two libraries. The achieved transformation efficiencies are given on the right.

Table 4.2 The result of applying optimised cloning protocol for incorporating two genomic libraries into Multi-Genius vector via blunt-end ligation; the integration of the first fragment (presented in Figure 4.11) – achieved number of transformants.

	EGFP	RFP
P_{T5}	$4 \cdot 10^5$	10^5
P_{rham}	10^6	$6 \cdot 10^5$
Transformation efficiency: $5 \cdot 10^8$ *cfu/µg*		

Firstly, each of the two fluorescent proteins was integrated into each cloning site – resulting in four combinations (EGFP/CS1, EGFP/CS2, RFP/CS1 and RFP/CS1). The results are presented in Figures 4.11 and Table 4.2. The results clearly demonstrate an excellent performance of the developed protocol. It should be noted that transformation efficiency of more than 50% is, most likely, caused by a poor growth of cells containing the inverted EGFP and RFP genes, due to potential misfolding and formation of inclusion bodies.

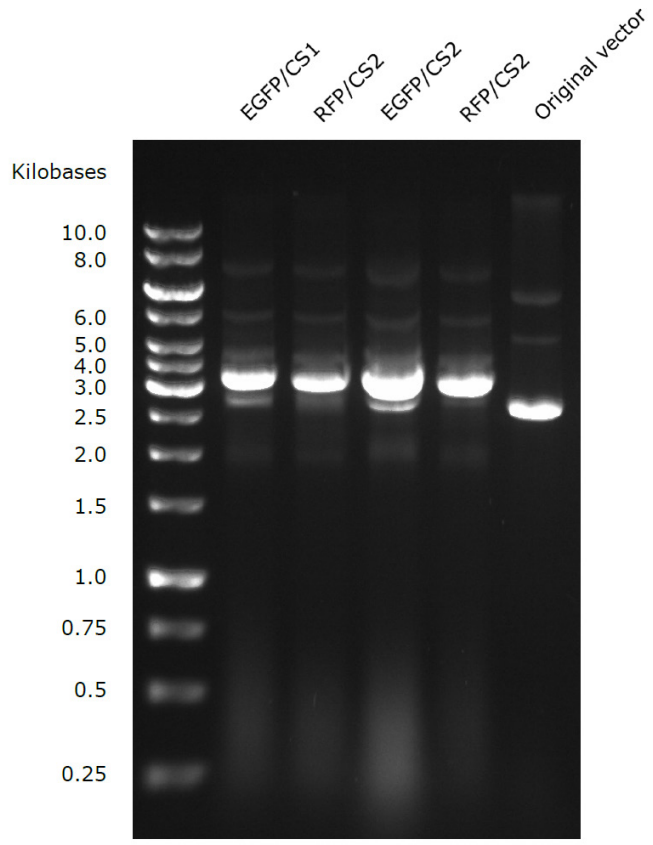

Figure 4.12 Plasmids extracted from four different population cells resulting from applying the optimised cloning protocol to integrate: (1) EGFP into cloning site 1, (2) RFP into cloning site 1, (3) EGFP into cloning site 1, and (4) RFP into cloning site 2.

To further investigate the efficiency of the method, plasmids extracted from each population were analysed on agarose gel (Figure 4.12). The gel image shows quite clearly that the vast majority of plasmids, and consequently the transformants, contain an insert. What is more, the number of plasmids containing more than one insert is negligible.

Subsequently, the optimised protocol was used to integrate RFP into cloning site 2 of EGFP/CS1 plasmid mixture and EGFP into cloning site 2 of RFP/CS1 plasmid mixture. The results are presented in Figure 4.13. Once again, the cloning efficiency proved to be very high.

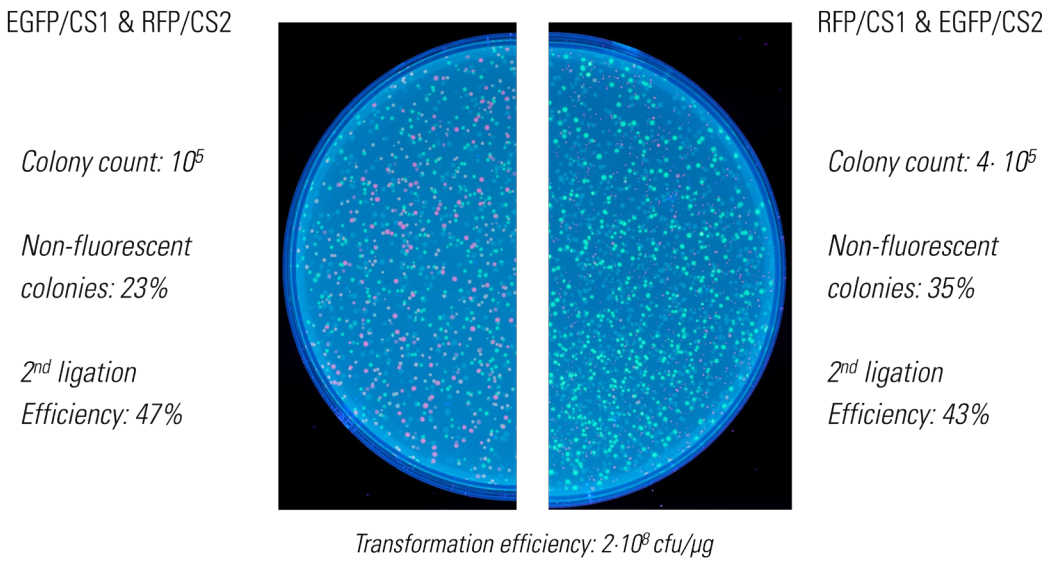

Figure 4.13 The result of applying optimised cloning protocol for incorporating two genomic libraries into Multi-Genius vector via blunt-end ligation; the integration of the second fragment. The RFP and EGFP represent the two libraries.

4.7 Methods

4.7.1 Materials

All enzymes, deoxyribonucleotides and DNA ladders were purchased from New England Biolabs (Ipswich, USA).

4.7.2 Primers

Primers used in this study were synthesized by Eurofins Genomics (Ebersberg, Germany). Melting temperatures of oligonucleotides were determined using the New England Biolabs T_m Calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator). To amplify EGFP and RFP fragments the following four primers were used:

EGFP_Fwd_2 (5'- GTGAGCAAGGGCGAGG -3'; 16 bp)

EGFP-Rev (5'-TTACTTGTACAGCTCGTCCATG-3', 22 bp)

RFP_Fwd_2 (5'- GCGAGTAGCGAAGACGTTATCAAAGAG -3'; 27 bp)

RFP-Rev (5'-TTAAGCACCGGTGGAGTGACG-3', 21 bp)

The primers were phosphorylated by T4 Kinase and subsequently purified with QIAquick Nucleotide Removal Kit (Qiagen).

To linearise Multi-Genius expression vector, the following four HPLC-prurified primers were used:

Int_CS1_a_HPLC (5'- CATTTTTTACCTCCTTAAAAG-3'; 21 bp)

Int_CS1_b_HPLC (5'- TAAGTAAGTAAGCTATGGAG -3'; 20 bp)

Int_CS2_a_HPLC (5'- CATATGTATATCTCCTTCTTATAGTTAAAC -3'; 30 bp)

Int2_CS2_b_HPLC (5'- TAATTAGCTGAAGGGAAAGC -3'; 57 bp)

4.7.3 Genomic DNA extraction

Genomic DNA was extracted using Bacterial DNA Kit from OMEGA Bio-Tek, in accordance with manufacturer's instructions.

4.7.4 Genomic DNA amplification

Extracted genomic DNA was amplified using illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit, in accordance with manufacturer's instructions.

4.7.5 DNA fragmentation

Genomic DNA was fragmented using Fragmentase from NEB, in accordance with manufacturer's instructions. To achieve the desired size distribution, 20µl reaction mixtures were prepared, containing: 1X Fragmentase Reaction Buffer, up 3 µg of genomic DNA and 0.8 µl of Fragmentase. The reaction mixture was incubated for 10 min at 37°C. After 10 min, the reaction was stopped by adding 5 µl of 0.5 M EDTA.

4.7.6 Vector construction

Linearised pD441-SR vector was ordered from ATUM and 530 bp DNA insert was synthesised by GeneScript. The DNA fragment was digested overnight with SapI restriction enzyme. The restriction enzyme was deactivated by 20 min incubation at 65°C. The two fragments were assembled together in 20 µl reaction mixture containing: 1X T4 DNA Ligase Buffer , 0.02 pmol of linearised pD441-SR vector, 0.06 pmol of the insert and 1 µl T4 DNA ligase. The ligation mixture was incubated

overnight at 16°C. 5 µl of the ligation mixture was for chemical transformation of *E. coli* DH5α. The presence of the insert was confirmed by sequencing.

4.7.7 DNA cloning

To amplify EGFP gene, PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM phosphorylated *EGFP_Fwd_2* primer, 500 nM *EGFP-Rev* phosphorylated primer, 0.2 ng pEGFP, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the following programme: (i) 30 s initial denaturation at 98°C, (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 65°C and 30 s extension at 72°C and (iii) 2 min final extension at 72°C. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon).

To amplify RFP gene, PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM phosphorylated *RFP_Fwd_2* primer, 500 nM *RFP-Rev* phosphorylated primer, 0.2 ng pBbA8k-RFP, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the following programme: (i) 30 s initial denaturation at 98°C, (ii) 30 cycles of 7 s denaturation at 98°C, 20 s annealing at 70°C and 30 s extension at 72°C and (iii) 2 min final extension at 72°C. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon).

To linearise Multi-Genius expression vector at the cloning site 1 (under the control of T_5 promoter), PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each

dNTP, 500 nM *Int_CS1_a_HPLC* primer, 500 nM *Int_CS1_b_HPLC* primer, 1 ng Multi-Genius expression vector, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the following programme: (i) 30 s initial denaturation at 98°C, (ii) 24 cycles of 7 s denaturation at 98°C, 20 s annealing at 55°C and 2 min 15 s extension at 72°C and (iii) 4 min final extension at 72°C. . Forty units of DpnI were subsequently added to the PCR mixture and incubated at 37°C for 1 h to remove template DNA. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon).

To linearise Multi-Genius expression vector at the cloning site 2 (rhamnose-inducible), PCR mixture (50 µl) containing 1× Q5 Reaction Buffer, 200 µM of each dNTP, 500 nM *Int_CS2_a_HPLC* primer, 500 nM *Int_CS2_b_HPLC* primer, 1 ng Multi-Genius expression vector, and 1 U Q5 High-Fidelity DNA Polymerase was thermocycled using the following programme: (i) 30 s initial denaturation at 98°C, (ii) 24 cycles of 7 s denaturation at 98°C, 20 s annealing at 60°C and 2 min 15 s extension at 72°C and (iii) 4 min final extension at 72°C. . Forty units of DpnI were subsequently added to the PCR mixture and incubated at 37°C for 15 min to remove template DNA. The PCR product was purified using QIAquick PCR Purification Kit and its DNA concentrations was determined using VersaWave (Expedeon).

The linearised vector was dephosphorylated by incubation Shrimp Alkaline Phosphatase (rSAP) for 1.5 h at 37°C. To inactivate the enzyme, the mixture was subsequently heated to 65°C for 15 min.

The linearised vector and the inserts encoding for the two fluorescent were assembled together in 20 µl ligation mixtures containing: 1X T4 DNA Ligase Buffer, 6.8% PEG6000, 7.5% 1,2-propanediol, 0.04 pmol of linearised vector, 0.12 pmol of the insert and 1,600 U T4 DNA ligase. The ligation mixture was incubated at 16°C for 4 h. Afterwards, the ligation mixture was purified and concentrated to 5 µl using DNA Clean & Concentrator™-5 (Zymo Research).

E. coli DH5α was transformed via electroporation with 4 µl of the resultant ligation mixture (purified). Transformed cells were plated on LB agar plates (10 g/l tryptone, 5 g/l yeast extract, 10 g/l sodium chloride and 15 g/l agar) supplemented with: 50 µg/ml kanamycin, 1 mM IPTG and/or 0.2% (w/v) rhamnose.

4.7.8 DNA gel electrophoresis

PCR products were analyzed using either 0.7% agarose gel. For visualisation purposes, ethidium bromide was added to the gel. Quick-Load 1 kb DNA Ladder was used as a DNA ladder.

Chapter 5
Multi-Genius:
Validation and application

5.1 Expression studies

To further investigate the functionality of the developed method, after performing EGFP and RFP cloning experiments summarised in Section 4.6, two plasmids, exhibiting the simultaneous expression of both fluorescent proteins, were picked at random and the two plasmid were isolated: CS1:EGFP/CS2:RFP and CS1:RFP/CS2:EGFP, where CS1 denotes cloning site 1 under the control of T_5 promoter and CS2 represents the rhamnose-inducible cloning site. Firstly, both cloning sites in the two plasmids were sequenced and it was confirmed that the two fluorescent proteins were seamlessly integrated into the vector.

To investigate the orthogonality of the two cloning sites, the two vectors were cultured with different configurations of the two inducers – rhamnose and IPTG. The resultant cell pellets were carefully inspected and the fluorescence corresponding to EGFP and RFP expression was also measured (Figures 5.1 and 5.2; Tables 5.1 and 5.2)..

All in all, the result are quite promising and the expression vector seems to be working as intended. However, some rather unexpected observations need to be noted.

It appears that when two proteins are being co-expressed, the fluorescence of RFP significantly exceeds the fluorescence of the EGFP. At first glance, it might be assumed that the two proteins are present at fairly similar amounts and the observed discrepancy is caused by the difference in their brightness or an overlap of their emission-absorption spectra. However, the brightness of EGFP is three times higher than RFP (Campbell et al., 2002; NIC_Wiki, 2017) and the potential effects of spectral overlap seem to be insufficient to account for such a significant difference.

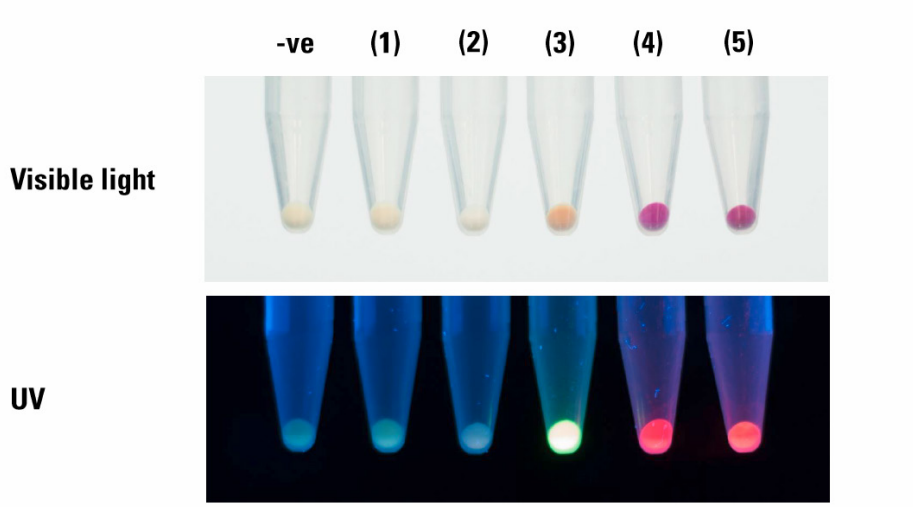

Figure 5.1 Cell pellets of CS1:EGFP/CS2:RFP-transformed cells after incubation for 12 h (at 37°C and 500 rpm) in LB media supplemented with: (1) kanamycin only, (2) Kan + 2% glucose, (3) Kan + 1mM IPTG, (4) Kan + 0.2% L-rhamnose, and (5) Kan + 1 mM IPTG + 0.2% L-rhamnose. (-ve) denotes cell pellet of DH5α cells with no plasmid.

Table 5.1 Fluorescence of CS1:EGFP/CS2:RFP-transformed cells after incubation for 12 h (at 37°C and 500 rpm) in LB media supplemented with: (1) kanamycin only, (2) Kan + 2% glucose, (3) Kan + 1mM IPTG, (4) Kan + 0.2% L-rhamnose, and (5) Kan + 1 mM IPTG + 0.2% L-rhamnose.

$\lambda_{exc}/\lambda_{em}$ [nm]	Fluorescence [10^3 RFU]				
	(1)	(2)	(3)	(4)	(5)
488/509 (EGFP)	0.44	0.05	16.29	0.13	3.83
584/607 (RFP)	0.10	0.11	1.58	13.64	14.10

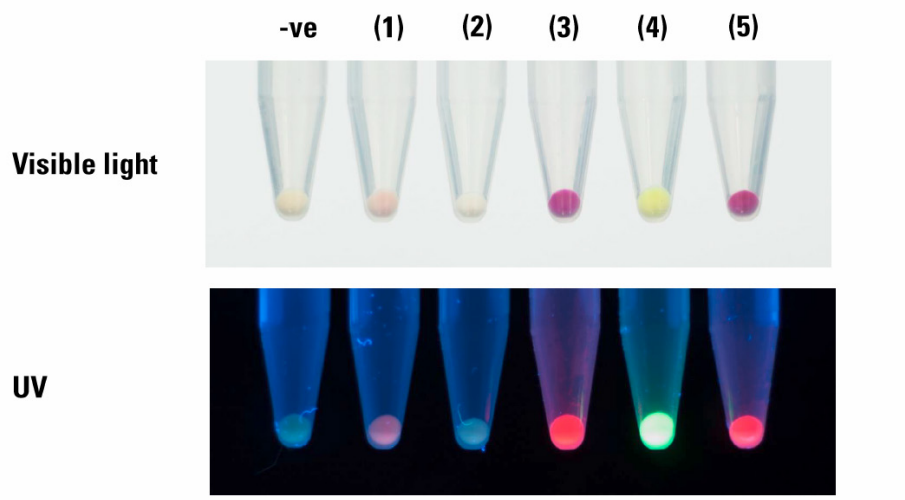

Figure 5.2 Cell pellets of CS1:RFP/CS2:EGFP-transformed cells after incubation for 12 h at (37°C and 500 rpm) in LB media supplemented with: (1) kanamycin only, (2) Kan + 2% glucose, (3) Kan + 1 mM IPTG, (4) Kan + 0.2% L-rhamnose, and (5) Kan + 1 mM IPTG + 0.2% L-rhamnose. (-ve) denotes cell pellet of DH5α cells with no plasmid.

Table 5.2 Fluorescence of CS1:RFP/CS2:EGFP-transformed cells after incubation for 12 h at (37°C and 500 rpm) in LB media supplemented with: (1) kanamycin only, (2) Kan + 2% glucose, (3) Kan + 1 mM IPTG, (4) Kan + 0.2% L-rhamnose, and (5) Kan + 1 mM IPTG + 0.2% L-rhamnose.

$\lambda_{exc}/\lambda_{em}$ [nm]	Fluorescence [10^3 RFU]				
	(1)	(2)	(3)	(4)	(5)
488/509 (EGFP)	0.05	0.04	0.11	17.80	1.21
584/607 (RFP)	0.55	0.10	14.27	0.05	12.87

It is rather the case that when the two proteins are being co-expressed the rate of EGFP expression is much lower than RFP. A situation when the expression of one protein overtakes the second one (leading to its poor expression) during co-expression, even when the two protein are well expressed in isolation, has been reported in literature (Diatchenko et al., 1996). This phenomenon has been attributed to different translational rates between the two proteins (Novagen, 2011). This is in complete agreement with the investigated situation, as the utilised EGFP gene has been optimised for expression in mammalian cells and due to the presence of rare codons its translational rate should be much lower than RFP optimised for expression in *E. coli*.

The second observation is that in the absence of its repressor, glucose, T_5 promoter seem to be slightly leaky as seen by a slight RFP expression in Figure 5.2 when no inducer is added (1).

Finally, the transcription terminator between the cloning site seem to be imperfect as seen by a noticeable expression of RFP when IPTG is added to CS1:EGFP/CS2:RFP-transformed cells (Table 5.1).

Further expression studies were conducted, to investigate expression profile with respect to varying concentration of one inducer (Figures 5.3-5.6). The results show that the expression from the two promoters is titratable over a nearly 5 fold expression range. In addition, the observation about the imperfect transcription terminator was further confirmed – resulting in unwanted expression from the second cloning site, when the first one is induced (Figure 5.3).

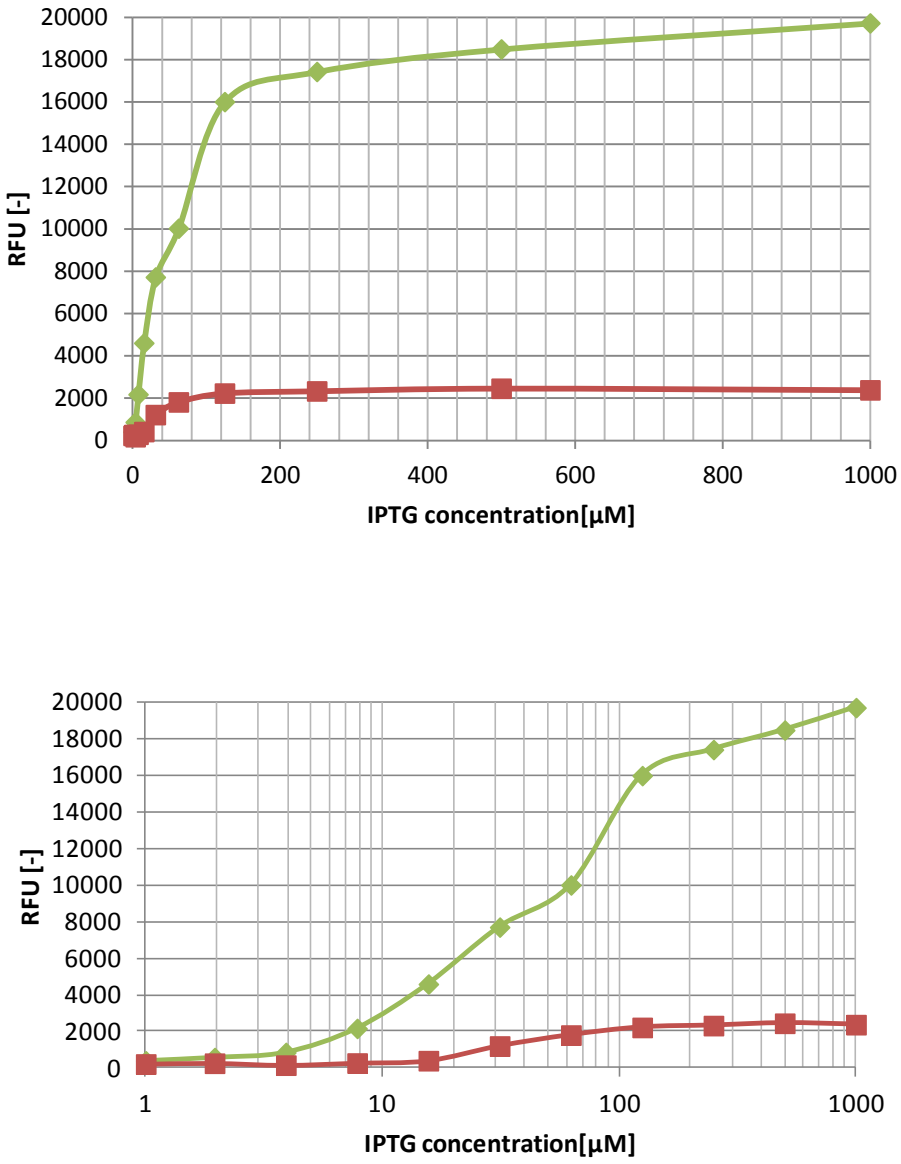

Figure 5.3 Fluorescence of CS1:EGFP/CS2:RFP-transformed cells after 12 h incubation with different concentration of IPTG. Green line denotes EGFP (λ_{exc} = 488 nm ; λ_{em} = 509 nm) and red line corresponds to RFP (λ_{exc} = 584 nm ; λ_{em} = 607 nm). First chart - linear scale; second chart – logarithmic scale.

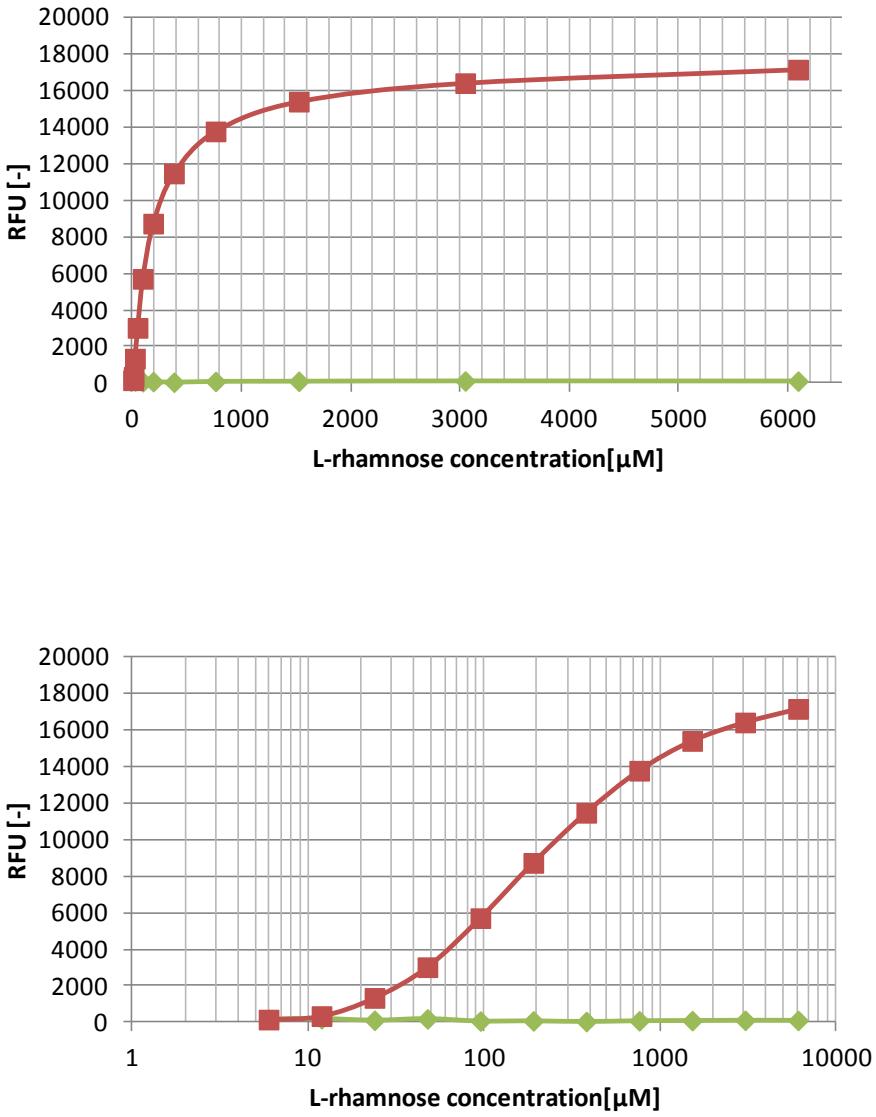

Figure 5.4 Fluorescence of CS1:EGFP/CS2:RFP-transformed cells after 12 h incubation with different concentration of L-rhamnose. Green line denotes EGFP (λ_{exc} = 488 nm ; λ_{em} = 509 nm) and red line corresponds to RFP (λ_{exc} = 584 nm ; λ_{em} = 607 nm). First chart - linear scale; second chart – logarithmic scale.

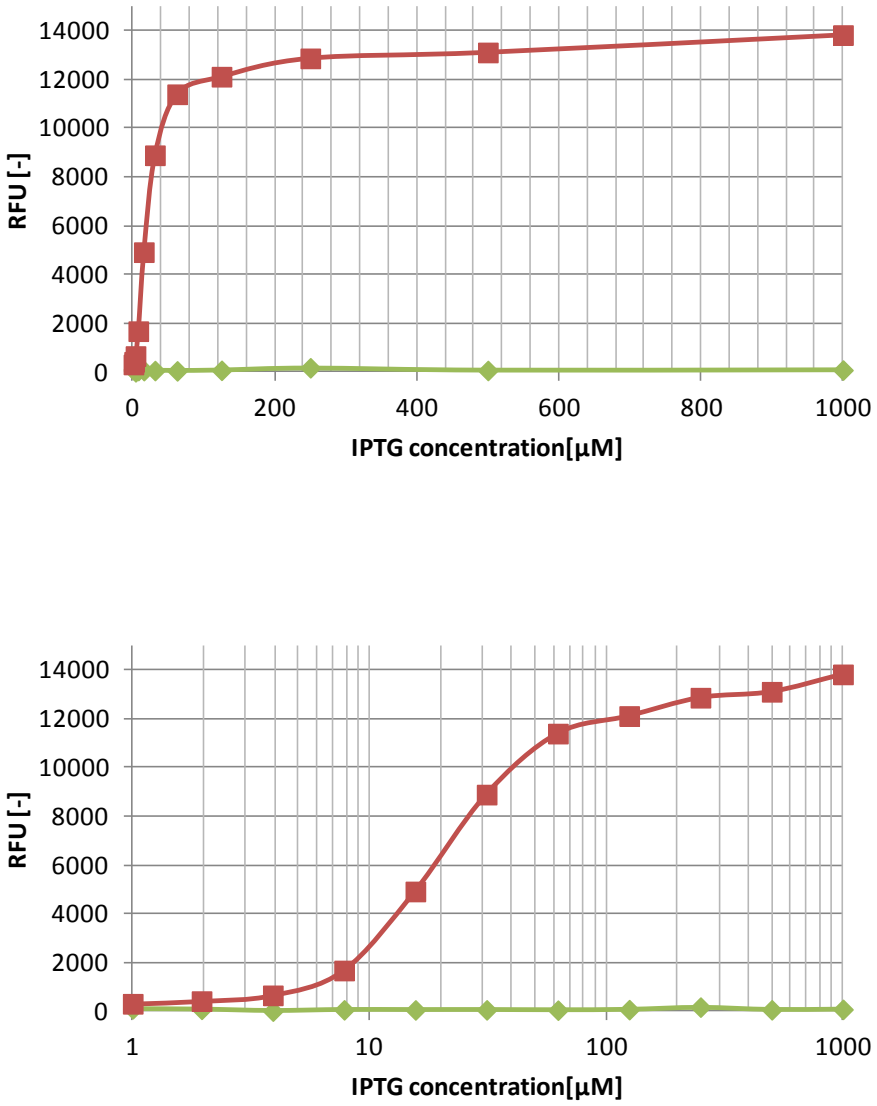

Figure 5.5 Fluorescence of CS1:RFP/CS2:EGFP-transformed cells after 12 h incubation with different concentration of IPTG. Green line denotes EGFP (λ_{exc} = 488 nm ; λ_{em} = 509 nm) and red line corresponds to RFP (λ_{exc} = 584 nm ; λ_{em} = 607 nm). First chart - linear scale; second chart – logarithmic scale.

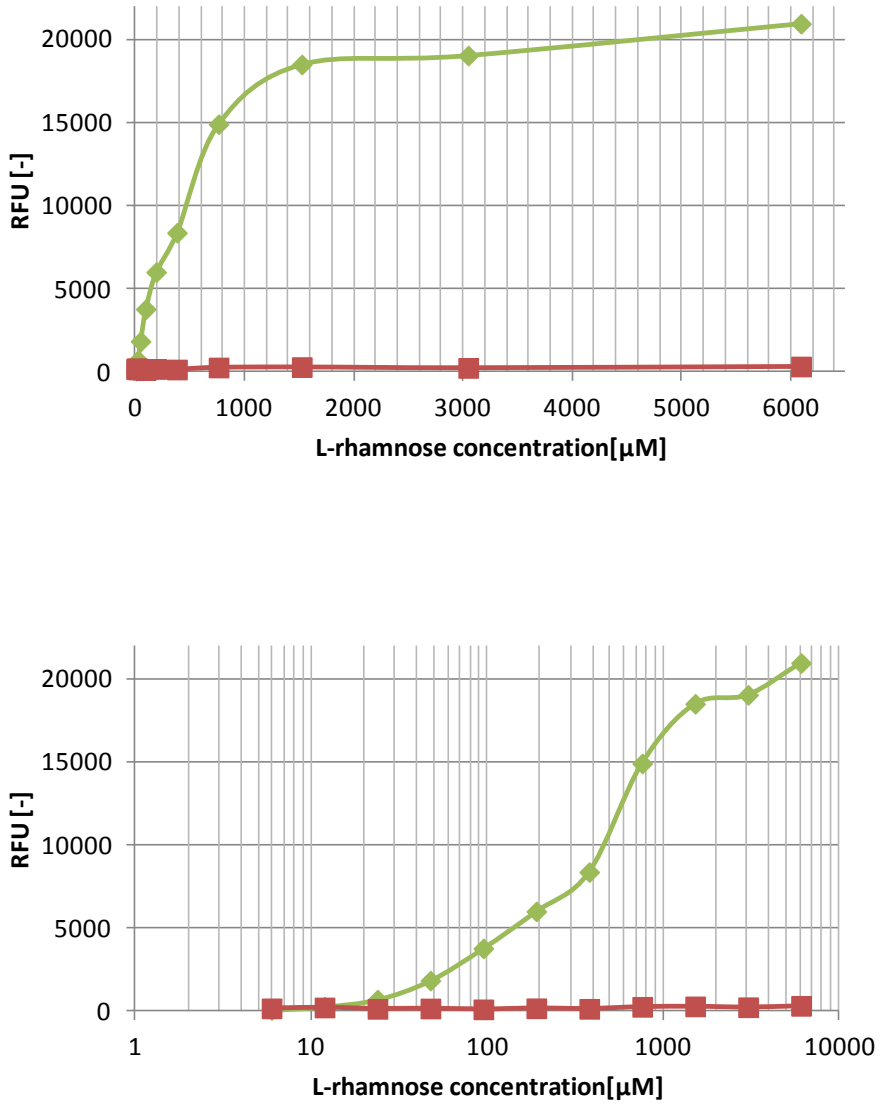

Figure 5.6 Fluorescence of CS1:RFP/CS2:EGFP-transformed cells after 12 h incubation with different concentration of L-rhamnose. Green line denotes EGFP (λ_{exc} = 488 nm ; λ_{em} = 509 nm) and red line corresponds to RFP (λ_{exc} = 584 nm ; λ_{em} = 607 nm). First chart - linear scale; second chart – logarithmic scale.

5.2 Method application

5.2.1 Thermotolerance and halotolerance of *E. coli* DH5α

To demonstrate the usefulness of Multi-Genius, it was decided that the method will be used to identify novel genes encoding for thermotolerance and halotolerance, in order to develop these traits in *E. coli* DH5α.

In order to identify improved variants and choose the right selective conditions, intrinsic resistance of *E. coli* DH5α to high temperatures and high sodium chloride concentration was investigated. To do so, *E. coli* was cultured in a form of spot plates at different temperatures and sodium chloride concentrations. The results of these experiments are presented in Figure 5.7 and 5.8.

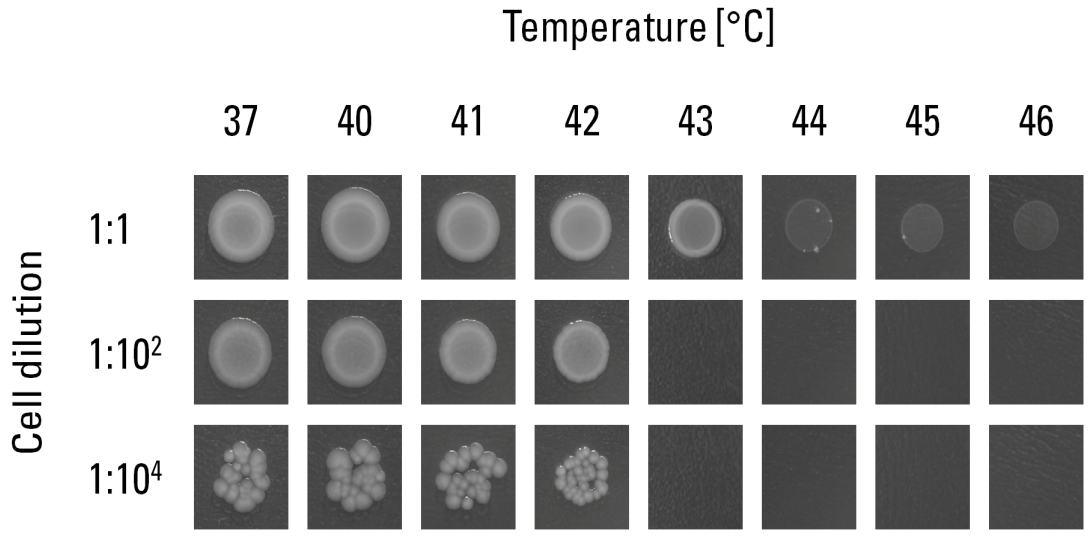

Figure 5.7 Natural thermotolerance of *E. coli* DH5α - the photographs of *E. coli* spots incubated on LB agar plates for 24 h at different temperatures. Each spot was made with 2.5 µl of cell suspension of *E. coli* cells grown to OD_{600} of 0.5 or its appropriate dilution.

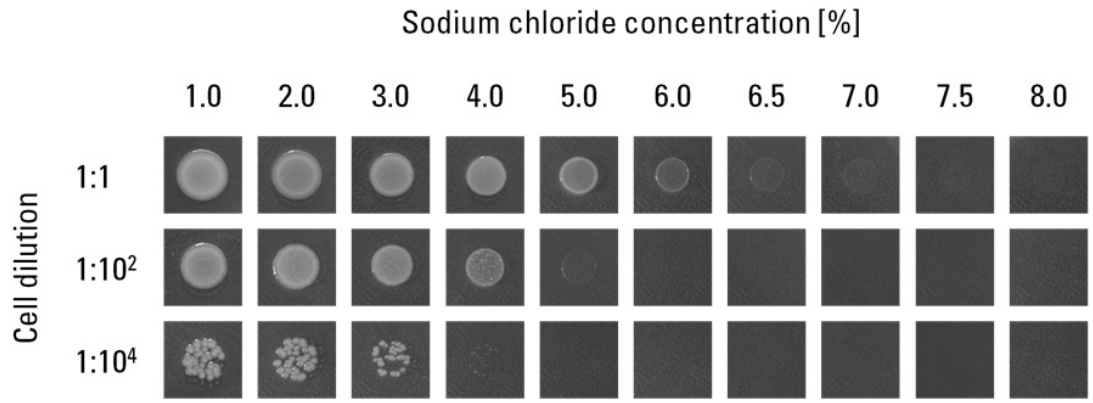

Figure 5.8 Natural halotolerance of *E. coli* DH5α - the photographs of *E. coli* spots incubated for 24 h on LB agar plates supplemented with different concentrations of sodium chloride. Each spot was made with 2.5 µl of cell suspension of *E. coli* cells grown to OD_{600} of 0.5 or its appropriate dilution.

It should be noted that that for both high sodium chloride concentration and high temperature there is no clear cut-off point after crossing of which all the cells suddenly stop propagating. It is rather the case that due to a natural ability of microorganism to evolve, the survivability of *E. coli* DH5α gradually decreases with higher temperature and higher sodium chloride concentration.

It is worth noting that after several days of storing the plates at room temperature, two previously not visible colonies grew on the 8% NaCl agar plate. It was later confirmed that these two colonies are not caused by a contamination and the obtained result is reproducible (when high number of cells is plated on the 8% NaCl agar plate). This observation shows the importance of designing and implementing proper controls

during the subsequent screening experiments (due to a possibility of obtaining false positives).

5.2.2 Application

To identify novel genes encoding for thermotolerance and halotolerance, three organisms were chosen as donors of the genomic DNA and a suitable target of the method. It was ensured that genomes of the selected organisms had been sequenced and the associated information is publicly available.

The first organism is *Thermus thermophilus*, a Gram-negative bacteria and a model organism for genetic engineering, The microorganisms was originally isolated from thermal vents and grows optimally at 65°C. One concern regarding this organism, is that its genes might be difficult to express in *E. coli* due to a significant differences between optimal growth temperature of the two organisms.

As a result, the second selected organism is *Bacillus coagulans,* a well-studied Gram-positive bacteria and obligate thermophile – its optimal growth temperature is 50°C, however, the organisms is capable of tolerating temperatures in the range of 30-55°C.

To help evolve halotolerance in *E. coli*, the genomic DNA from *Chromohalobacter salexigens* was utilised. This gram-negative bacterium is capable of surviving a wide range of salt concentrations.

Genomic DNA from the three organisms was ordered from German Collection of Microorganisms and Cell Cultures GmbH, DSMZ.

Including the genomic DNA extracted from *E. coli* DH5α, this constitutes four sources of genomic libraries – three for evolving thermotolerance and two for halotolerance (genomic DNA extracted from *E. coli* DH5α can be used for both experiments).

The details of the conducted experiment are presented in Table 5.3 It should be noted that all the individual screenings presented in Table 5.3 were conducted concurrently.

Table 5.3 The outline of the experiment using Multi-Genius to identify novel genes encoding for thermotolerance and halotolerance. CS1 denotes the genomic library integrated into the first cloning site (under the control of T_5 promoter) and CS2 denotes genomic library integrated into the second cloning site (rhamnose-inducible). In the case of *C. salexigens* and *E. coli* the integration of the fragmented genomic library was performed twice resulting in two unique set of libraries per organism.

Experiment no.	CS1	CS2	Selective condition
1	*T. thermophilus*	*B. coagulans*	46°C
2	*B. coagulans*	*B. coagulans*	46°C
3	*T. thermophilus*	*T. thermophilus*	46°C
4,5	*E. coli* DH5α	*E. coli* DH5α	46°C
6,7	*C. salexigens*	*C. salexigens*	6.5 and 8% NaCL
8.9	*E. coli* DH5α	*E. coli* DH5α	6.5 and 8% NaCL
Control	*egfp*	*rfp*	All of the above

After integrating the genomic library into the second cloning site and performing electroporation, the transformants were grown on LB plates supplemented with kanamycin, IPTG and rhamnose, and subjected to selective conditions described in Table 5.3. The results of the experiment are summarised in Table 5.4.

Table 5.4 The results of the experiment of using Multi-Genius to identify novel genes encoding for thermotolerance and halotolerance. Library size is the same as transformation efficiency for a given experiment.

Experiment no.	Library size	Observed colonies
1	$3 \cdot 10^6$	254
2	$7 \cdot 10^5$	58
3	$2 \cdot 10^6$	0
4	$9 \cdot 10^4$	24
5	$2 \cdot 10^5$	12
6	$2 \cdot 10^6$	0
7	$2 \cdot 10^5$	0
8	$9 \cdot 10^4$	0
9	$2 \cdot 10^5$	0
Control	-	0

In short, no colonies were observed on agar plates supplemented with increased salt concentration, making the attempt of isolating halotolerant variant unsuccessful.

Also, the cells containing the plasmid with genomic libraries from *T. thermophilus* resulted in no colonies. As it was already discussed, the potential cause might be poor protein expression.

All in all, however, the usefulness of the method was clearly demonstrated by isolating thermotolerant variants containing genomic fragments from *E. coli*, *T. thermophilus* and *B. coagulans*. In addition, a significantly more colonies were observed on the plates associated with *B. coagulans* than *E. coli*. This proves the hidden potential of screening organisms other than *E. coli*, even when the ultimate goal is evolve a phenotype in this organism.

5.2.3 Further characterisation

60 bacterial variants exhibiting an improved thermotolerance were picked from the four plates exhibiting an observable colony growth at 46°C (Experiments 1, 2, 4 and 5). To further characterise them, their growth rate was measured in a 96-well plate and compared to cells containing an empty Multi-Genius vector, which act as a negative control (for more details on the experimental procedure, see Section 5.4.8).

The initial experimental design involved growing the cells in a media supplemented with different combinations of the two inducers at 46°C (the four combinations being: no inducer, IPTG only, L-rhamnose only and both inducers) and comparing the bacterial growth under these conditions. Such experimental setup should not only confirm the improvement in thermotolerance but also determine, for each isolated variant, which genomic insert (or both) is responsible for the observed phenotype (depending on which combination of the two inducers results in the improved thermotolerance). In this way, the usefulness of modular expression employed in Multi-Genius would be clearly demonstrated.

Unfortunately, temporal and spatial variations of the temperature within the shaking incubators (deriving from the limitations of the equipment) interfered with an accurate measurement of bacterial growth rate at 46°C. At this temperature, even small variations in the temperature (<0.5°C) lead to relatively large differences in the growth rate. As no statistically-significant difference between the four conditions could be readily observed in this scenario, the final results are presented as a mean value of the optical density measured for the four conditions. As each condition was performed in two replicates,

the optical density, presented in Table 5.5, is a mean value of eight data points. By averaging individual values, the effect of spatial variation of the temperature (within two incubators used for the experiment) was minimised and the accuracy of the final results was significantly increased.

Table 5.5 Mean optical density of the 60 isolated thermotolerant variants measured after 8 h of exponential growth at 46°C, presented as a ratio of the optical density to the value obtained for *E. coli* cells containing an empty Multi-Genius vector (negative control). The results can be interpreted as a fold improvement in OD_{600} after 8 h exponential growth at 46°C. OD_{600} measured for the negative control was equal to 0.0085. Rows and columns describe a location within the 96 well-plate.

		Column									
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Row	**B**	*15*	*25*	*22*	*11*	*9*	*20*	*17*	*8*	*23*	*23*
	C	*2*	*25*	*17*	*8*	*8*	*18*	*15*	*16*	*20*	*19*
	D	*34*	*12*	*14*	*10*	*24*	*8*	*19*	*12*	*25*	*19*
	E	*19*	*18*	*10*	*8*	*6*	*19*	*22*	*22*	*14*	*18*
	F	*21*	*22*	*8*	*19*	*12*	*20*	*23*	*14*	*20*	*19*
	G	*27*	*21*	*11*	*18*	*17*	*8*	*16*	*20*	*17*	*17*

The presented results clearly demonstrate a marked improvement in the growth rate of the isolated variants, as compared to the *E. coli* cells containing the empty Multi-Genius vector. After 8h of exponential growth, most of the isolated variants show at least 10-fold improvement in the final optical density.

To confirm that the observed phenotype is really conferred by the inserts within the Multi-Genius vector and was not caused by a random mutation of endogenous genomic DNA during the selection process, the plasmids from 10 best-performing variants were isolated and used for transformation of native *E. coli* DH5α cells.

The growth rate of these variants was measured in a similar fashion to the previous experiment (for more details on the experimental procedure, see Section 5.4.8). Once again, no statistically-significant difference between the four combinations of the inducer concentration could be readily observed. As such, the final results (shown in Table 5.6) are presented as an average value of the optical density measured for the four conditions.

Table 5.6 Mean optical density of the *E. coli* DH5α transformed with the plasmids purified from the Multi-Genius experiment, measured after 8 h of exponential growth at 46°C. The results are presented as a ratio of the optical density to the value obtained for *E. coli* cells containing an empty Multi-Genius vector (negative control). The results can be interpreted as a fold improvement in OD_{600} after 8 h exponential growth at 46°C. OD_{600} measured for the negative control was equal to 0.0076.

Variant	OD_{600}/OD_{600}^{-ve} [-]
B2	26
D2	29
G2	30
B3	29
C3	26
B4	28
D6	27
B10	27
D10	28
B11	27

Similarly to the previous experiment, the isolated variants showed a marked improvement in the growth rate at 46°C, as compared to the *E. coli* cells containing an empty Multi-Genius vector. Based on the presented results, it can be claimed with a high dose of confidence that the observed improvement in the thermotolerance is conferred by the inserts within the Multi-Genius vector.

To further elucidate the mechanism behind the observed improvement in themotolerance, the inserts from the ten isolated plasmids were sequenced. The results are presented in Table 5.7.

The sequencing results are somehow unexpected but, at the same time, consistent with the results of the previous experiments. All vectors corresponding to genomic DNA from *B. coagulans* contain the same gene (*thyA*), encoding for an enzyme thymidylate synthase. It should be noted that apart from variants B4 and D6, which contain identical inserts, the remaining variants appear to be unique library members. The fact that all of the selected variants feature the same gene, seems to suggest that it is integral to the investigated phenotype. What is more, *thyA* is also present in the Multi-Genius vectors constructed from genomic DNA of *E. coli* DH5α, further proving the significance of this gene for enhancing thermotolerance. Considering that the improved thermotolerance is observed even when the insert is integrated in a reverse orientation, it can be hypothesised that the expression of the gene relies on its native, constitutive promoter, which is recognised by *E. coli* DH5α transcription machinery. This observation explains the fact that during the previously-discussed bacterial growth

studies, no statistically-significant differences between the four combinations of the inducer concentration could be readily observed.

Table 5.7 Sequencing results of the ten plasmids, isolated from Multi-Genius experiment, including the position of the insert within a bacterial chromosome (GenBank: CP025268.1 for *E. coli* and CP009709.1 for *B. coagulans)* and complete genes present within the fragment.

Variant	Cloning site	gDNA source	Start	End	Complete genes present on the insert
B2	1	*E. coli* DH5α	1,752,276	1,751,639	-
	2	*E. coli* DH5α	2,967,500	2,965,533	*thyA, ppdA*
D2	1	*E. coli* DH5α	Unreadable sequence		
	2	*E. coli* DH5α	Unreadable sequence		
G2	1	*E. coli* DH5α	2,967,473	2,966,501	*thyA*
	2	*E. coli* DH5α	2,697,068	2,698,441	none
B3	1	*B. coagulans*	No insert		-
	2	*B. coagulans*	273,837	272,408	*thyA*
C3	1	*B. coagulans*	454,191	453,077	BF29_449, BF29_450
	2	*B. coagulans*	273,786	272,770	*thyA*
B4	1	*B. coagulans*	No insert		-
	2	*B. coagulans*	273,784	272,578	*thyA*
D6	1	*B. coagulans*	No insert		-
	2	*B. coagulans*	273,784	272,578	*thyA*
B10	1	*T. thermophilus*	No insert		-
	2	*B. coagulans*	274,139	272,788	*thyA, dfrA*
D10	1	*T. thermophilus*	No insert		-
	2	*B. coagulans*	274,154	272,795	*thyA, dfrA*
B11	1	*T. thermophilus*	No insert		-
	2	*B. coagulans*	272,213	274,002	*thyA*

It should be noted that many of the sequenced plasmids contained only one insert. Considering that the observed improvement in the phenotype seems to be conferred by only one gene and the sequenced plasmids come from the variants exhibiting the highest growth rate at 46°C, it is not surprising that such selection pressure would favour the cells with only one functional insert. In other words, if the investigated phenotype is encoded by only one gene (within the space of investigated genomic DNA), the presence of additional insert (that does not confer the phenotype) places an extra metabolic burden on the cell. As such, most of the members of the starting library contain two inserts (as determined in Section 4.6) and only during the subsequent selection process library members with only one insert are preferentially enriched when a phenotype is encoded just by one gene.

As *thyA* features in all nine variants that provided a readable DNA sequence, it can be said with a high dose of confidence that this gene is responsible for observed thermotolerance. Inside a cell, thymidylate synthase catalyses the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate and, as such, is integral to DNA replication and overall cell growth. According to Stout *et al.,* thymidylate synthase from *E. coli* functions as a homodimer with a melting temperature of 45.7°C (Stout et al., 1998).

In light of the presented results and information, it can be hypothesised that thymidylate synthase constitutes a major bottleneck to thermotolerance of *E. coli* and its poor growth at 46°C can be explained by denaturation of a protein integral to DNA replication. This detrimental effect can be alleviated by either increasing the expression

level of the enzyme (as observed in isolated variants B2 and G2) or providing a more thermostable version of the enzyme (as observed in the seven isolated variants containing genomic DNA from *B. coagulans*). It would worthwhile to verify this hypothesis by cloning *thyA* into a common expression vector and conducting further characterisation studies.

5.3 Conclusion

Multi-Genius expands upon the concept of genomic libraries to tap into naturally-existing diversity and expedite identification of genes encoding for useful phenotypes. The method addresses significant limitations and lacking features in the area of identifying genes encoding for useful phenotypes. Multi-Genius expands the concept of screening multiple genes (in search of DNA fragments encoding for complex phenotypes) to genomic elements originating from exogenous genomic DNA.

The potential of the method is demonstrated by isolating thermotolerant variants of *Escherichia coli* DH5α. The applicability of the method to acquisition of new biological parts is shown by identifying the major role of *thyA* (encoding for thymidylate synthase) in enhacing thermotolerance. Based on the presented results, it is hypothesised that natural thermotolerance of *E. coli* can be increased by either increasing the expression of the gene or substituting the enzyme with a more thermostable homologue.

What is more, Multi-Genius utilises the concept of modular expression to address the problem of biological complexity. The use of two orthogonal, inducible promoters allow for studying the interplay between individual genetic elements and optimising the performance of the resultant biological system by fine-tuning the individual gene expression. Due to the intrinsic nature of the investigated system, the usefulness of this concept could not be properly showcased. However, it is believed that this integral feature will prove invaluable when applying Multi-Genius to a different system (namely, a different phenotype or source of genomic DNA is investigated).

All in all, Multi-Genius ties together different concepts and techniques associated with molecular biology to offer a streamlined protocol for uncovering new biological parts and studying interactions between distant genes.

5.4 Methods

5.4.1 Materials

All enzymes, deoxyribonucleotides and DNA ladders were purchased from New England Biolabs (Ipswich, USA).

5.4.2 Primers and genomic DNA

Primers used in this study were synthesized by Eurofins Genomics (Ebersberg, Germany). Melting temperatures of oligonucleotides were determined using the New England Biolabs T_m Calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator).

To linearise Multi-Genius expression vector, the following four HPLC-purified primers were used:

Int_CS1_a_HPLC (5'- CATTTTTTACCTCCTTAAAAG-3'; 21 bp)

Int_CS1_b_HPLC (5'- TAAGTAAGTAAGCTATGGAG-3'; 20 bp)

Int_CS2_a_HPLC (5'- CATATGTATATCTCCTTCTTATAGTTAAAC-3'; 30 bp)

Int2_CS2_b_HPLC (5'- TAATTAGCTGAAGGGAAAGC-3'; 57 bp)

For sequencing of the two cloning sites of the Multi-Genius expression vector, the following four primers were used:

CS1_Seq_Fwd (5'- GAGCGGATAACAATTACGAGC-3'; 21 bp)

CS1_Seq_Rev (5'- AATCCAGATGGAGTTCTGAGG-3'; 21 bp)

CS2_Seq_Fwd (5'- GTCAGTAACGAGAAGGTCG-3'; 19 bp)

CS1_Seq_Rev (5'- CAAATAAAACGAAAGGCTCAGTCG-3'; 24 bp)

Genomic DNA from the following bacterial strains - *Bacillus coagulans* ATCC 7050 (DSM no. 1), *Thermus thermophilus* HB8 (DSM no. 579) and *Chromohalobacter salexigens* 1H11 (DSM no: 3043) – was purchased from German Collection of Microorganisms and Cell Cultures GmbH (DSMZ).

5.4.3 Initial expression studies

50 µl aliquots of overnight culture from both CS1:EGFP/CS2:RFP and CS1:RFP/CS2:EGFP-containing cells were used to inoculate 5 ml aliquots of LB media supplemented with: (1) kanamycin only (50 µg/ml), (2) kanamycin and 2% (w/v) glucose, (3) kanamycin and 1 mM IPTG, (4) kanamycin and 0.2% (w/v) L-rhamnose and (5) kanamycin, 1 mM IPTG and 0.2% (w/v) L-rhamnose. The cells were grown for 12 h at 37°C with shaking (500 rpm).

3 ml aliquots of cell cultures were pelleted by centrifugation in 1.5 ml microcentrifuge tubes (supernatant was carefully removed by pipetting). The photographs of the pellets were taken under visible or UV light.

200 µl aliquots of cell suspension were used for fluorescent measurements using SpectraMax M2e Multi-Mode Microplate Reader (Molecular Devices) using the following settings:

- EGFP – excitation wavelength: 488 nm; emission wavelength: 509 nm
- RFP – excitation wavelength: 584 nm; emission wavelength: 607 nm

Cell suspension of wild-type DH5α cells (containing no plasmid) were used as a blank solution.

5.4.4 Further expression studies

5 µl aliquots of overnight culture from both CS1:EGFP/CS2:RFP and CS1:RFP/CS2:EGFP-containing cells were used to inoculate 0.5 ml aliquots of LB media (supplemented with 50 µg/ml kanamycin and varying concentration of the two inducers) present in individual wells of a 96-deep-well plate. CS1:EGFP/CS2:RFP-containing cells were used to inoculate: (1) 12 wells containing 1:2 serial dilutions of 1mM IPTG (1 mM, 0.5 mM, 0.25 mM and so on), (2) 12 wells containing 1:2 serial dilutions of 0.2% (w/v) L-rhamnose, and (3) 12 wells containing 1 mM of IPTG and 1:2 serial dilutions of 0.2% (w/v) L-rhamnose. CS1:RFP/CS2:EGFP-containing cells were used to inoculate: (1) 12 wells containing 1:2 serial dilutions of 1mM IPTG, (2) 12 wells containing 1:2 serial dilutions of 0.2% (w/v) L-rhamnose and (3) 12 wells containing 0.2% (w/v) L-rhamnose and 1:2 serial dilutions of 1 mM IPTG. 6 additional wells (containing LB media supplemented with 50 µg/ml kanamycin) were inoculated with cells bearing an empty vector.

The cells were grown for 12 h at 37°C with shaking (1050 rpm). 200 µl aliquots of cell suspension were used for fluorescent measurements using SpectraMax M2e Multi-Mode Microplate Reader (Molecular Devices) using the following settings:

- EGFP – excitation wavelength: 488 nm; emission wavelength: 509 nm
- RFP – excitation wavelength: 584 nm; emission wavelength: 607 nm

5.4.5 Investigation of halotolerance of *E. coli* DH5α

5 ml LB media (supplemented with 50 µg/ml kanamycin) was inoculated with 100 µl overnight culture of *E. coli* DH5α transformed with Multi-Genius expression vector (containing T5 and rhamnose promoters). When OD600 of the bacterial culture reached 0.5, three dilutions (1:100, 1:1000 and $1:10^6$) of the cell suspension were prepared. 4 spots of each dilution (1:1, 1:100, 1:1000 and $1:10^6$; 16 spots in total for each plate) were made by carefully transferring 2.5 µl aliquots of cell suspension onto the surface of pre-warmed LB agar plates (supplemented with different concentration of NaCl - 1, 2, 3, 4, 5, 6, 6.5, 7, 7.5 and 8% w/v) and allowing the excess liquid to dry. After 20 h of incubation at 37°C, the photographs of individual spots were taken.

5.4.6 Investigation of thermotolerance of *E. coli* DH5α

5 ml LB media (supplemented with 50 µg/ml kanamycin) was inoculated with 100 µl overnight culture of *E. coli* DH5α transformed with Multi-Genius expression vector (containing T5 and rhamnose promoters). When OD600 of the bacterial culture reached 0.5, three dilutions (1:100, 1:1000 and $1:10^6$) of the cell suspension were prepared. 4 spots of each dilution (1:1, 1:100, 1:1000 and $1:10^6$; 16 spots in total) were made by carefully transferring 2.5 µl aliquots of cell suspension onto the surface of pre-warmed LB agar plates and allowing the excess liquid to dry.

After 20 h of incubation at different temperatures (37°C, 41°C, 42°C, 43°C, 44°C, 45°C and 46°C), the photographs of the individual spots were taken.

5.4.7 Multi-Genius - application

Genomic DNA was extracted from *E. coli* DH5α using Bacterial DNA Kit from OMEGA Bio-Tek, in accordance with manufacturer's instructions. Genomic DNA from *E. coli* DH5α, *B. coagulans*, *T. thermophilus* and *C. salexigens* was amplified and fragmented as described in sections 4.7.4 and 4.7.5. Genomic DNA fragments in the range of 1.5-2 kb were isolated via gel extraction method, using NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel) according to manufacturer's instructions.

DNA Polymerase I, Large (Klenow) Fragment (NEB) was used to blunt the ends of the isolated DNA fragments. DNA fragments from each organism were separately dissolved in 1x NEBuffer and supplemented with 33 μM each dNTP. 1 U of Klenow fragment was added to each reaction mixture. After 15 min incubation at 25°C, the reaction was stopped by heating the reaction mixtures for 20 min at 70°C.

Multi-Genius expression vector was linearised at the cloning site 1 and purified as described in section 4.7.7. Eight ligation mixtures were prepared in parallel to join the linearised Multi-Genius vector with the following inserts: (1,2) *E. coli* DH5α genomic DNA fragments, (3) *B. coagulans* genomic DNA fragments, (4,5) *T. thermophilus* genomic DNA fragments, (6,7) *C. salexigens* genomic DNA fragments, and (8) EGFP gene. The ligation and subsequent electroporation was carried out as described in section 4.7.7. 1/1000 of the electroporation outgrowth was plated on LB agar plates supplemented with 50 μg/ml kanamycin (and incubated overnight at 37°C) to estimate the cloning efficiency. The remaining transformed cells (from each electroporation) were transferred separately into 100 ml LB media supplemented with 50 μg/ml kanamycin

and 2% (w/v) glucose (instead of being plated on agar plates). The cells were grown at 37°C with shaking (250 rpm). When OD600 of the eight bacterial cultures reached 1 (after about 8 h), the plasmids (containing the previously mentioned inserts integrated into the cloning site 1) were isolated from each culture using E.Z.N.A. Plasmid Mini Kit I (OMEGA Bio-Tek), in accordance with manufacturer's instructions.

The eight plasmid mixtures were linearised at the cloning site 2 and purified as described in section 4.7.7. Eight ligation mixtures were prepared in parallel to join the linearised plasmids with the following inserts: (1,2) *E. coli* DH5α genomic DNA fragments, (3,4) *B. coagulans* genomic DNA fragments, (5) *T. thermophilus* genomic DNA fragments, (6,7) *C. salexigens* genomic DNA fragments, and (8) RFP gene. The aim of this ligation step was to create the following constructs:

(i) CS1: *E. coli* DH5α gDNA; CS2: *E. coli* DH5α gDNA

(ii) CS1: *E. coli* DH5α gDNA; CS2: *E. coli* DH5α gDNA

(iii) CS1: *B. coagulans* gDNA; CS2: *B. coagulans* gDNA

(iv) CS1: *T. thermophilus* gDNA; CS2: *B. coagulans* gDNA

(v) CS1: *T. thermophilus* gDNA; CS2: *T. thermophilus* gDNA

(vi) CS1: *C. salexigens* gDNA; CS2: *C. salexigens* gDNA

(vii) CS1: *C. salexigens* gDNA; CS2: *C. salexigens* gDNA

(viii) CS1: *egfp*; CS2: *rfp*

The ligation and subsequent electroporation was carried out as described in section 4.7.7. 1/1000 of the electroporation outgrowth was plated on LB agar plates supplemented with 50 μg/ml kanamycin (and incubated overnight at 37°C) to estimate

the cloning efficiency. The remaining transformed cells (from each electroporation) were again transferred separately into 100 ml LB media supplemented with 50 μg/ml kanamycin and 2% (w/v) glucose. The cells were grown at 37°C with shaking (250 rpm).

When OD600 of the eight bacterial cultures reached 0.5 (after about 7-8 h), aliquots of the cell cultures were plated on LB agar plates supplemented with 50 μg/ml kanamycin, 16 nM IPTG and 0.003% (w/v) L-rhamnose (corresponding to a quarter of maximum expression for the two promoters, as determined for RFP and EGFP via the experiment described in section 5.3.4). 10^6 cells were plated on each agar plate for constructs (i)-(iii) and 10^7 cells were used for the remaining constructs (iv)-(viii). As a result, the following multiplier of each library (library sizes are presented in Table 5.4) were plated for each construct: (i) 11, (ii) 5, (iii) 1.4, (iv) 3, (v) 5, (vi) 5 and (vii) 20. Libraries (i), (ii), (iii), (iv) and (v) plated were plated on the agar plates supplemented with the two inducers and incubated overnight at 46°C. Libraries (i), (ii), (vi) and (vii) were plated on agar plates supplemented with both inducers and either 6.5 or 8% (w/v) NaCl and incubated overnight at 37°C. In addition, 10^7 cells containing construct (viii) were plated for each condition to act as a negative control.

Glycerol stocks were prepared from colonies exhibiting improved thermotolerance and stored in a microtiter plate at -80°C for further analysis – 30 colonies were randomly picked from the plate corresponding to Experiment 1 (described in Tables 5.3 and 5.4), 24 colonies from Experiment 2, 2 colonies from Experiment 4 and 4 colonies from Experiment 5 (60 samples in total).

5.4.8 Further characterisation

The glycerol stocks (prepared as described in Section 5.3.7) were used to inoculate a 96-well plate containing LB media supplemented with 50 μg/ml kanamycin and 2% (w/v) glucose, to suppress the expression from the Multi-Genius vector. The resultant plate contained the isolated thermotolerant bacteria in wells B2-K7 (60 samples in total). 16 wells in the corners of the plate (4 per each corner) contained media only. The remaining 20 wells were inoculated with *E. coli* cells containing the empty Multi-Genius expression vector, to act as a negative control. The plate was incubated overnight at 37°C with shaking (1050 rpm). Subsequently, the plate was briefly centrifuged to spin down the cells and the supernatant was removed with a multichannel pipette. The fresh LB media supplemented with kanamycin only was added to each well (250 μl media per well) and the cells were gently resuspended. Using a multichannel pipette and 50 μl of cell suspension, the master plate was replicated into four plates containing LB media supplemented with either: (1) kanamycin, (2) kanamycin and 16 nM IPTG, (3) kanamycin and 0.003% (w/v) L-rhamnose, and (4) kanamycin, 16 nM IPTG and 0.003% (w/v) L-rhamnose. The four plates were incubated with shaking at 37°C for 2h to induce expression from the Multi-Genius vector. Subsequently, each plate was manually replicated (using a standard pin replicator) into two deep-well plates containing LB media (1.2 ml per well) with the same inducer concentrations, preequilibrated at 46°C. This resulted in eight plates in total – two sets of plates with four different inducer concentrations. Each set of plate was quickly transferred to a separate incubator, preequilibrated at 46°C, and incubated with shaking for 8h. After

the incubation, OD600 was measured using SpectraMax M2e Multi-Mode Microplate Reader.

Plasmids from 10 best-performing variants were purified using QIAprep Spin Miniprep Kit (Qiagen). The two inserts per plasmid were sequenced by Eurofins Genomics (Ebersberg, Germany) using sequencing primers listed in Section 5.3.2.

E. coli DH5α was transformed with the isolated plasmids using a standard chemical transformation protocol (Minagawa et al., 2007). Glycerol stocks were prepared from the transformed cells and stored at -80°C for further analysis.

The glycerol stocks were used to inoculate a 96-well plate containing LB media supplemented with 50 μg/ml kanamycin and 2% (w/v) glucose to suppress the expression from the Multi-Genius vector. Columns 2-11 were used to inoculate eight replicates of the ten variants (one replicate per row, for rows E-H the ten variants were replicated in reverse – with the first variant in column 11). The columns 1 and 12 were inoculated with *E. coli* cells containing the empty Multi-Genius expression vector, to act as a negative control. The plate was incubated overnight at 37°C with shaking (1050 rpm). Subsequently, the plate was briefly centrifuged to spin down the cells and the supernatant was removed with a multichannel pipette. 250 μl fresh LB media was added to each well supplemented with (1) kanamycin only for rows A and H, (2) kanamycin and 16 nM IPTG for rows B and G, (3) kanamycin and 0.003% (w/v) L-rhamnose for rows C and F, and (4) kanamycin, 16 nM IPTG and 0.003% (w/v) L-rhamnose for rows D and E. Immediately afterwards, the cells were gently resuspended. The plate was incubated with shaking at 37°C for 2h to induce expression from the

Multi-Genius vector. Subsequently, the master plate was manually replicated (using a standard pin replicator) into eight deep-well plates containing LB media (1.2 ml per well) with the same inducer concentrations, preequilibrated at 46°C. The plates were quickly transferred to a separate incubator, preequilibrated at 46°C, and incubated with shaking for 8h. After the incubation, OD600 was measured using SpectraMax M2e Multi-Mode Microplate Reader.

Chapter 6
Conclusions and future work

6.1 Conclusions

The aim of the presented research was to expand molecular toolbox for synthetic biology allowing a more efficient access to largely untapped potential of biological diversity. This was achieved via the development of two molecular methods: QuickStep-Cloning and Multi-Genius.

QuickStep-Cloning utilises recent advances in megaprimer-based cloning to allow for seamless integration of a DNA fragment of interest into a plasmid in less than 6 hours – the result that could not be reproduced using traditional cloning or state-of-the-art methods such as restriction-free (RF) cloning. Owing to its rapid protocol and general simplicity, QuickStep-Cloning also compares favourably with four recently-reported strategies of exponential megaprimer-based cloning (ABI-REC, RAM cloning, EMP cloning and IFPC).

The applicability of QuickStep-Cloning is certainly not limited to standard cloning experiments, involving transfer of a gene sequence from a donor vector to a recipient plasmid. The developed method could be especially useful for protein tagging or, potentially, cloning DNA fragments directly from genomic DNA. In general, the method should constitute a useful tool for facilitating the vital process of constructing and testing new biological circuits by allowing faster integration of genetic elements. As such, QuickStep-Cloning should find its applications in the developing fields of protein engineering, metabolic engineering and synthetic biology.

The new improved version of the method, QuickStepS-Cloning, uses phosphorothioate oligonucleotides to not only simplify the overall procedure but also significantly increase

its cloning efficiency. It also shown that incorporating random mutagenesis into the method allows for streamlining directed evolution experiments. It is shown that QuickStepS-Cloning cloning can compete with even the most efficient methods of incorporating combinatorial libraries, such as MEGAWHOP. All in all, the developed method constitute an excellent addition to a general molecular toolbox. QuickStepS-Cloning should be of great interest to synthetic biology because not only it facilitates the construction of biological circuits (expanding the capabilities and the throughput of the previous method) but also, when used in combination with directed evolution, it can expedite the process of acquiring new biological parts.

Whereas the potential applications of QuickStep-Cloning revolve around artificially-induced diversity, Multi-Genius expands upon the concept of genomic libraries to tap into naturally-existing diversity and expedite identification of genes encoding for useful phenotypes. The method addresses significant limitations and lacking features in the area of identifying genes encoding for useful phenotypes, including the recently-reported strategy of Coexisting/Coexpressing Genomic Libraries, CoGeLs. Among other things, Multi-Genius expands the concept of screening multiple genes (in search of DNA fragments encoding for complex phenotypes) to genomic elements originating from exogenous genomic DNA. As such, the method broadens the repertoire of molecular tools for acquiring new biological parts. What is more, Multi-Genius utilises the concept of modular expression to address the problem of biological complexity. The use of two orthogonal, inducible promoters allow for studying the interplay between individual genetic elements and optimise the performance of resultant biological systems by fine-

tuning the expression of individual genes. All in all, the method ties together different concepts and techniques associated with molecular biology to offer a streamlined protocol for uncovering and studying interactions between distant genes. The potential of the method is demonstrated by isolating thermotolerant variants of *E. coli* DH5α. Even though, the usefulness of modular expression could not be showcased in this series of experiments, the applicability of the method to synthetic biology and other the field is clearly shown.

It is worth noting that the two methods, QuickStep-Cloning and Multi-Genius, clearly complement themselves. Multi-Genius allows for tapping into the natural diversity and helps to identify new genes encoding for useful phenotypes. QuickStep-Cloning, on the other hand, expedites the way in which known genes are being engineered - e.g., through the process of directed evolution - to create fit-for-purpose genetic parts.

All things considered, both QuickStep-Cloning and Multi-Genius expand the molecular toolbox available to synthetic biologists and enhance the way biodiversity is being studied and utilised.

6.2 Future work

In the short term, the role of *thyA* (encoding for thymidylate synthase) for improving thermotolerance in *Escherichia coli* DH5α should be further investigated. This should be achieved by:

- cloning the gene from *B. coagulans* into a common expression vector
- comparing the growth rate of *E. coli* cells transformed with the constructed plasmid with those that contain the empty expression vector (acting as a negative control) at elevated temperatures - e.g., 46°C

If the increase in the thermotolerance is very significant, the identified genes can be integrated into the genome of *E. coli*, to create a new, thermotolerant strain that might be of some industrial interest. There is also a potential for employing protein engineering to investigate whether the thermotolerance of *E. coli* cells can be further enhanced by increasing the thermostability of the enzyme. QuickStepS-Cloning would prove invaluable in such an experiment.

In the long term, it would be desirable to utilise Multi-Genius in conjunction with various different genomic libraries to identify new genes encoding for other useful phenotypes. It would also be desirable to showcase the usefulness of modular expression featured in Multi-Genius. This should be achieved by:

- isolating a bacterial colony exhibiting a particular phenotype in which the phenotype is encoded by a two-gene combination

- conducting sequencing of the extracted plasmid to identify genes encoding for the phenotype and elucidate the underlined mechanism

- transforming native *E. coli* DH5α with the isolated plasmid and demonstrating that the observed change in the phenotype comes solely from the plasmid

- utilising microtiter plates or agar plates with varying concentrations of the inducer to identify the optimal expression levels of the two genes encoding for the phenotype

It is also envisaged that the efficiency of QuickStepS-Cloning could be further increased through further experiments and general optimisation.

Ultimately, Multi-Genius and QuickStepS-Cloning could be combined together to provide a comprehensive protocol for streamlined acquisition of fit-for-purpose genetic parts by (1) tapping into natural diversity and (2) further engineering the identified genes via directed evolution.

REFERENCES

1989. Bacteria in the Oil Field: Bad News, Good News. The Technical Review 37, 48-53.

Ajikumar, P.K., Xiao, W.H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330, 70-74.

Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., 2005. Tuning genetic control through promoter engineering. Proceedings of the National Academy of Sciences of the United States of America 102, 12678-12683.

Alper, H., Stephanopoulos, G., 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metabolic engineering 9, 258-267.

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological reviews 59, 143-169.

Andrianantoandro, E., Basu, S., Karig, D.K., Weiss, R., 2006. Synthetic biology: new engineering rules for an emerging discipline. Molecular systems biology 2, 2006 0028.

Anthony, J.R., Anthony, L.C., Nowroozi, F., Kwon, G., Newman, J.D., Keasling, J.D., 2009. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. Metabolic engineering 11, 13-19.

Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H., Ikehara, M., 1983. Essential structure of E. coli promoter: effect of spacer length between the two consensus sequences on promoter function. Nucleic acids research 11, 5855-5864.

ATUM, 2017. https://www.atum.bio/products/expression-vectors/bacterial#2.

Auguet, J.C., Barberan, A., Casamayor, E.O., 2010. Global ecological patterns in uncultured Archaea. The ISME journal 4, 182-190.

Baker, B.J., Banfield, J.F., 2003. Microbial communities in acid mine drainage. FEMS microbiology ecology 44, 139-152.

Balleza, E., Lopez-Bojorquez, L.N., Martinez-Antonio, A., Resendis-Antonio, O., Lozada-Chavez, I., Balderas-Martinez, Y.I., Encarnacion, S., Collado-Vides, J., 2009. Regulation by transcription factors in bacteria: beyond description. FEMS microbiology reviews 33, 133-151.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., Valla, S., 2013. A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. Microbial cell factories 12, 26.

Bass, S.H., Yansura, D.G., 2000. Application of the E. coli trp promoter. Molecular biotechnology 16, 253-260.

Beloqui, A., Nechitaylo, T.Y., Lopez-Cortes, N., Ghazi, A., Guazzaroni, M.E., Polaina, J., Strittmatter, A.W., Reva, O., Waliczek, A., Yakimov, M.M., Golyshina, O.V., Ferrer, M., Golyshin, P.N., 2010. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Applied and environmental microbiology 76, 5934-5946.

Ben-Dov, E., Kramarsky-Winter, E., Kushmaro, A., 2009. An in situ method for cultivating microorganisms using a double encapsulation technique. FEMS microbiology ecology 68, 363-371.

Berlec, A., Strukelj, B., 2013. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. Journal of industrial microbiology & biotechnology 40, 257-274.

Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A.D., Magdassi, S., Tawfik, D.S., 2004. In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. Analytical biochemistry 325, 151-157.

Bevan, M.W., Franssen, M.C., 2006. Investing in green and white biotech. Nature biotechnology 24, 765-767.

Bi, Y., Qiao, X., Hua, Z., Zhang, L., Liu, X., Li, L., Hua, W., Xiao, H., Zhou, J., Wei, Q., Zheng, X., 2012. An asymmetric PCR-based, reliable and rapid single-tube native DNA engineering strategy. BMC biotechnology 12, 39.

Biedendieck, R., Gamer, M., Jaensch, L., Meyer, S., Rohde, M., Deckwer, W.D., Jahn, D., 2007. A sucrose-inducible promoter system for the intra- and extracellular protein production in Bacillus megaterium. Journal of biotechnology 132, 426-430.

Blainey, P.C., 2013. The future is now: single-cell genomics of bacteria and archaea. FEMS microbiology reviews 37, 407-427.

Blank, C.E., Cady, S.L., Pace, N.R., 2002. Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. Applied and environmental microbiology 68, 5123-5135.

Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., Schwaneberg, U., 2010. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. Analytical biochemistry 406, 141-146.

Blazeck, J., Alper, H.S., 2013. Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnology journal 8, 46-58.

Bongers, R.S., Veening, J.W., Van Wieringen, M., Kuipers, O.P., Kleerebezem, M., 2005. Development and characterization of a subtilin-regulated expression system in Bacillus subtilis: strict control of gene expression by addition of subtilin. Applied and environmental microbiology 71, 8818-8824.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K., 2012. Engineering the third wave of biocatalysis. Nature 485, 185-194.

Brosius, J., Erfle, M., Storella, J., 1985. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. The Journal of biological chemistry 260, 3539-3541.

Brown, M.V., Bowman, J.P., 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS microbiology ecology 35, 267-275.

Browning, D.F., Busby, S.J., 2004. The regulation of bacterial transcription initiation. Nature reviews. Microbiology 2, 57-65.

Bryksin, A.V., Matsumura, I., 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. BioTechniques 48, 463-465.

Bunterngsook, B., Kanokratana, P., Thongaram, T., Tanapongpipat, S., Uengwetwanit, T., Rachdawong, S., Vichitsoonthonkul, T., Eurwilaichitr, L., 2010. Identification and

characterization of lipolytic enzymes from a peat-swamp forest soil metagenome. Bioscience, biotechnology, and biochemistry 74, 1848-1854.

Busby, S., Ebright, R.H., 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79, 743-746.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A., Tsien, R.Y., 2002. A monomeric red fluorescent protein. Proceedings of the National Academy of Sciences of the United States of America 99, 7877-7882.

Carlson, E.D., Gan, R., Hodgman, C.E., Jewett, M.C., 2012. Cell-free protein synthesis: applications come of age. Biotechnology advances 30, 1185-1194.

Carrier, T.A., Keasling, J.D., 1997. Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. Biotechnology progress 13, 699-708.

Carrier, T.A., Keasling, J.D., 1999. Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. Biotechnology progress 15, 58-64.

Chandran, D., Bergmann, F.T., Sauro, H.M., 2009. TinkerCell: modular CAD tool for synthetic biology. Journal of biological engineering 3, 19.

Chen, G.J., Qiu, N., Karrer, C., Caspers, P., Page, M.G., 2000. Restriction site-free insertion of PCR products directionally into vectors. BioTechniques 28, 498-500, 504-495.

Chen, I.C., Lin, W.D., Hsu, S.K., Thiruvengadam, V., Hsu, W.H., 2009. Isolation and characterization of a novel lysine racemase from a soil metagenomic library. Applied and environmental microbiology 75, 5161-5166.

Chen, R., 2012. Bacterial expression systems for recombinant protein production: E. coli and beyond. Biotechnology advances 30, 1102-1107.

Choffnes, E.R., Relman, D.A., Pray, L., 2011. The Science and Applications of Synthetic and Systems Biology: Workshop Summary.

Choi, Y.J., Morel, L., Le Francois, T., Bourque, D., Bourget, L., Groleau, D., Massie, B., Miguez, C.B., 2010. Novel, versatile, and tightly regulated expression system for Escherichia coli strains. Applied and environmental microbiology 76, 5058-5066.

Collins, F.S., McKusick, V.A., 2001. Implications of the Human Genome Project for medical science. Jama 285, 540-544.

Connon, S.A., Giovannoni, S.J., 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Applied and environmental microbiology 68, 3878-3885.

Cottrell, M.T., Moore, J.A., Kirchman, D.L., 1999. Chitinases from uncultured marine microorganisms. Applied and environmental microbiology 65, 2553-2557.

Court, D.L., Sawitzke, J.A., Thomason, L.C., 2002. Genetic engineering using homologous recombination. Annual review of genetics 36, 361-388.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P., Pernodet, J.L., 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Applied and environmental microbiology 69, 49-55.

Craig, S.P., 3rd, Yuan, L., Kuntz, D.A., McKerrow, J.H., Wang, C.C., 1991. High level expression in Escherichia coli of soluble, enzymatically active schistosomal hypoxanthine/guanine phosphoribosyltransferase and trypanosomal ornithine decarboxylase. Proceedings of the National Academy of Sciences of the United States of America 88, 2500-2504.

Curtis, T.P., Sloan, W.T., Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America 99, 10494-10499.

Czar, M.J., Cai, Y., Peccoud, J., 2009. Writing DNA with GenoCAD. Nucleic acids research 37, W40-47.

D'Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., Lewis, K., 2010. Siderophores from neighboring organisms promote the growth of uncultured bacteria. Chemistry & biology 17, 254-264.

Daniel, R., 2005. The metagenomics of soil. Nature reviews. Microbiology 3, 470-478.

Davis, K.E., Joseph, S.J., Janssen, P.H., 2005. Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Applied and environmental microbiology 71, 826-834.

de Boer, H.A., Comstock, L.J., Vasser, M., 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. Proceedings of the National Academy of Sciences of the United States of America 80, 21-25.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O.P., Kok, J., 2012. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. BMC genomics 13, 299.

Demidov, V.V., Bukanov, N.O., Frank-Kamenetskii, D., 2000. Duplex DNA capture. Current issues in molecular biology 2, 31-35.

Dennig, A., Shivange, A.V., Marienhagen, J., Schwaneberg, U., 2011. OmniChange: the sequence independent method for simultaneous site-saturation of five codons. PloS one 6, e26222.

Deuschle, U., Kammerer, W., Gentz, R., Bujard, H., 1986. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. The EMBO journal 5, 2987-2994.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., Siebert, P.D., 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proceedings of the National Academy of Sciences of the United States of America 93, 6025-6030.

Donahue Jr, R.A., Bebee, R.L., 1999. BL21-SI™ competent cells for protein expression in E. coli. Focus 21, 49-51.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L., Keasling, J.D., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology 27, 753-759.

Dumont, M.G., Murrell, J.C., 2005. Stable isotope probing - linking microbial identity to function. Nature reviews. Microbiology 3, 499-504.

Ebright, R.H., Cossart, P., Gicquel-Sanzey, B., Beckwith, J., 1984. Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of E. coli. Nature 311, 232-235.

Ekkers, D.M., Cretoiu, M.S., Kielak, A.M., Elsas, J.D., 2012. The great screen anomaly - a new frontier in product discovery through functional metagenomics. Applied microbiology and biotechnology 93, 1005-1020.

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335-338.

Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P., Lilley, P.E., Dixon, N.E., 1990. Modified bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. Gene 87, 123-126.

Endy, D., 2005. Foundations for engineering biology. Nature 438, 449-453.

Fan, X., Liu, X., Wang, K., Wang, S., Huang, R., Liu, Y., 2011. Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. J Mol Catal B: Enzym 72, 319-326.

Fang, Z., Li, T., Wang, Q., Zhang, X., Peng, H., Fang, W., Hong, Y., Ge, H., Xiao, Y., 2011. A bacterial laccase from marine microbial metagenome exhibiting chloride tolerance and dye decolorization ability. Applied microbiology and biotechnology 89, 1103-1110.

Farinas, E.T., Schwaneberg, U., Glieder, A., Arnold, F.H., 2001. Directed evolution of a cytochrome P450 monooxygenase for alkane oxidation. Advanced Synthesis & Catalysis 343, 601-606.

Ferrari, B.C., Binnerup, S.J., Gillings, M., 2005. Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. Applied and environmental microbiology 71, 8714-8720.

Franklin, R.B., Mills, A.L., 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS microbiology ecology 44, 335-346.

Frazzetto, G., 2003. White biotechnology. EMBO reports 4, 835-837.

Fujii, R., Kitaoka, M., Hayashi, K., 2004. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic acids research 32, e145.

Fujii, R., Kitaoka, M., Hayashi, K., 2006. Error-prone rolling circle amplification: the simplest random mutagenesis protocol. Nature protocols 1, 2493-2497.

Gabor, E.M., 2004. Harvesting Novel Biocatalysts from the Metagenome. University of Groningen.

Gabor, E.M., de Vries, E.J., Janssen, D.B., 2004. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environmental microbiology 6, 948-958.

Galbraith, E.A., Antonopoulos, D.A., White, B.A., 2004. Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. Environmental microbiology 6, 928-937.

Gamer, M., Frode, D., Biedendieck, R., Stammen, S., Jahn, D., 2009. A T7 RNA polymerase-dependent gene expression system for Bacillus megaterium. Applied microbiology and biotechnology 82, 1195-1203.

Gao, X., Yuan, X.X., Shi, Z.Y., Guo, Y.Y., Shen, X.W., Chen, J.C., Wu, Q., Chen, G.Q., 2012. Production of copolyesters of 3-hydroxybutyrate and medium-chain-length

3-hydroxyalkanoates by E. coli containing an optimized PHA synthase gene. Microbial cell factories 11, 130.

Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in Escherichia coli. Nature 403, 339-342.

Gentz, R., Bujard, H., 1985. Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. Journal of bacteriology 164, 70-77.

Ghadessy, F.J., Ong, J.L., Holliger, P., 2001. Directed evolution of polymerase function by compartmentalized self-replication. Proceedings of the National Academy of Sciences of the United States of America 98, 4552-4557.

Gibson, D.G., 2011. Enzymatic assembly of overlapping DNA fragments. Methods in enzymology 498, 349-361.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., 3rd, Smith, H.O., Venter, J.C., 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329, 52-56.

Gilbert, J.A., Dupont, C.L., 2011. Microbial metagenomics: beyond the genome. Annual review of marine science 3, 347-371.

Girvan, M.S., Bullimore, J., Pretty, J.N., Osborn, A.M., Ball, A.S., 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Applied and environmental microbiology 69, 1800-1809.

Gish, G., Eckstein, F., 1988. DNA and RNA sequence determination based on phosphorothioate chemistry. Science 240, 1520-1522.

Giver, L., Gershenson, A., Freskgard, P.O., Arnold, F.H., 1998. Directed evolution of a thermostable esterase. Proceedings of the National Academy of Sciences of the United States of America 95, 12809-12813.

Gopal, G.J., Kumar, A., 2013. Strategies for the production of recombinant protein in Escherichia coli. The protein journal 32, 419-425.

Gorke, B., Stulke, J., 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nature reviews. Microbiology 6, 613-624.

Griffiths, A.D., Tawfik, D.S., 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. The EMBO journal 22, 24-35.

Griffiths, A.D., Tawfik, D.S., 2006. Miniaturising the laboratory in emulsion droplets. Trends in biotechnology 24, 395-402.

Gronenborn, B., 1976. Overproduction of phage lambda repressor under control of the lac promotor of Escherichia coli. Molecular & general genetics : MGG 148, 243-250.

Guan, L., Liu, Q., Li, C., Zhang, Y., 2013. Development of a Fur-dependent and tightly regulated expression system in Escherichia coli for toxic protein synthesis. BMC biotechnology 13, 25.

Guillerez, J., Lopez, P.J., Proux, F., Launay, H., Dreyfus, M., 2005. A mutation in T7 RNA polymerase that facilitates promoter clearance. Proceedings of the National Academy of Sciences of the United States of America 102, 5958-5963.

Guo, Z., 2016. Artemisinin anti-malarial drugs in China. Acta pharmaceutica Sinica. B 6, 115-124.

Gupta, R., Beg, Q.K., Lorenz, P., 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. Applied microbiology and biotechnology 59, 15-32.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of bacteriology 177, 4121-4130.

Haldimann, A., Daniels, L.L., Wanner, B.L., 1998. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. Journal of bacteriology 180, 1277-1286.

Hammer, K., Mijakovic, I., Jensen, P.R., 2006. Synthetic promoter libraries--tuning of gene expression. Trends in biotechnology 24, 53-55.

Harley, C.B., Reynolds, R.P., 1987. Analysis of E. coli promoter sequences. Nucleic acids research 15, 2343-2361.

Hawley, D.K., McClure, W.R., 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic acids research 11, 2237-2255.

Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., Daniel, R., 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. Applied and environmental microbiology 66, 3113-3116.

Hershberg, R., Bejerano, G., Santos-Zavaleta, A., Margalit, H., 2001. PromEC: An updated database of Escherichia coli mRNA promoters with experimentally identified transcriptional start sites. Nucleic acids research 29, 277.

Hill, A.D., Tomshine, J.R., Weeding, E.M., Sotiropoulos, V., Kaznessis, Y.N., 2008. SynBioSS: the synthetic biology modeling suite. Bioinformatics 24, 2551-2553.

Hu, Y., Fu, C., Huang, Y., Yin, Y., Cheng, G., Lei, F., Lu, N., Li, J., Ashforth, E.J., Zhang, L., Zhu, B., 2010. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS microbiology ecology 72, 228-237.

Huber, J.A., Butterfield, D.A., Baross, J.A., 2002. Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. Applied and environmental microbiology 68, 1585-1594.

Huber, R., Huber, H., Stetter, K.O., 2000. Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS microbiology reviews 24, 615-623.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., Pace, N.R., 1998. Novel division level bacterial diversity in a Yellowstone hot spring. Journal of bacteriology 180, 366-376.

Hughes, D.S., Felbeck, H., Stein, J.L., 1997. A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm Riftia pachyptila. Applied and environmental microbiology 63, 3494-3498.

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., Bohannan, B.J., 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Applied and environmental microbiology 67, 4399-4406.

iGEM, http://igem.org/.

Ishihama, A., 1993. Protein-protein communication within the transcription apparatus. Journal of bacteriology 175, 2483-2489.

Ishii, T., Yoshida, K., Terai, G., Fujita, Y., Nakai, K., 2001. DBTBS: a database of Bacillus subtilis promoters and transcription factors. Nucleic acids research 29, 278-280.

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology 3, 318-356.

Jensen, P.R., Hammer, K., 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Applied and environmental microbiology 64, 82-87.

Jiang, C., Hao, Z.-Y., Jin, K., Li, S.-X., Che, Z.-Q., Ma, G.-F., Wu, B., 2010. Identification of a metagenome-derived β-glucosidase from bioreactor contents. Journal of Molecular Catalysis B: Enzymatic 63, 11-16.

Kaeberlein, T., Lewis, K., Epstein, S.S., 2002. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 296, 1127-1129.

Katzen, F., Chang, G., Kudlicki, W., 2005. The past, present and future of cell-free protein synthesis. Trends in biotechnology 23, 150-156.

Kaznessis, Y.N., 2011. SynBioSS-aided design of synthetic biological constructs. Methods in enzymology 498, 137-152.

Keller, M., Zengler, K., 2004. Tapping into microbial diversity. Nature reviews. Microbiology 2, 141-150.

Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D., Endy, D., 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. Journal of biological engineering 3, 4.

Knietsch, A., Bowien, S., Whited, G., Gottschalk, G., Daniel, R., 2003a. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Applied and environmental microbiology 69, 3048-3060.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., Daniel, R., 2003b. Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on Escherichia coli. Applied and environmental microbiology 69, 1408-1416.

Knight, T., 2003. Idempotent vector design for standard assembly of biobricks, In MIT Synthetic Biology Working Group Technical Reports. , MIT Synthetic Biology Working Group Technical Reports. .

Konig, H., Frank, D., Heil, R., Coenen, C., 2013. Synthetic genomics and synthetic biology applications between hopes and concerns. Current genomics 14, 11-24.

Kopke, B., Wilms, R., Engelen, B., Cypionka, H., Sass, H., 2005. Microbial diversity in coastal subsurface sediments: a cultivation approach using various electron acceptors and substrate gradients. Applied and environmental microbiology 71, 7819-7830.

Kvist, T., Ahring, B.K., Lasken, R.S., Westermann, P., 2007. Specific single-cell isolation and genomic amplification of uncultured microorganisms. Applied microbiology and biotechnology 74, 926-935.

Kwok, R., 2010. Five hard truths for synthetic biology. Nature 463, 288-290.

Leadbetter, J.R., Schmidt, T.M., Graber, J.R., Breznak, J.A., 1999. Acetogenesis from H2 plus CO2 by spirochetes from termite guts. Science 283, 686-689.

Lee, S.K., Keasling, J.D., 2005. A propionate-inducible expression system for enteric bacteria. Applied and environmental microbiology 71, 6856-6862.

Levy, M., Griswold, K.E., Ellington, A.D., 2005. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. Rna 11, 1555-1562.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., Lu, P., 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 271, 1247-1254.

Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., Jaeger, K.E., 2000. Directed evolution of an enantioselective lipase. Chemistry & biology 7, 709-718.

Life Technologies, 2017. http://www.lifetechnologies.com/uk/en/home/life-science/cloning/gene-synthesis/directed-evolution.html.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. Journal of biomedicine & biotechnology 2012, 251364.

Liu, L., Yang, H., Shin, H.D., Chen, R.R., Li, J., Du, G., Chen, J., 2013. How to achieve high-level expression of microbial enzymes: strategies and perspectives. Bioengineered 4, 212-223.

Lorenz, P., Eck, J., 2005. Metagenomics and industrial applications. Nature reviews. Microbiology 3, 510-516.

Lowman, H.B., Bina, M., 1990. Temperature-mediated regulation and downstream inducible selection for controlling gene expression from the bacteriophage lambda pL promoter. Gene 96, 133-136.

Lu, Q., 2005. Seamless cloning and gene fusion. Trends in biotechnology 23, 199-207.

Lucigen, 2016. http://www.lucigen.com/Expresso-Rhamnose-Cloning-and-Protein-Expression-System/.

Lutz, S., 2010. Beyond directed evolution - semi-rational protein engineering and design. Current opinion in biotechnology 21, 734-743.

Makita, Y., Nakao, M., Ogasawara, N., Nakai, K., 2004. DBTBS: database of transcriptional regulation in Bacillus subtilis and its contribution to comparative genomics. Nucleic acids research 32, D75-77.

Marcy, Y., Ouverney, C., Bik, E.M., Losekann, T., Ivanova, N., Martin, H.G., Szeto, E., Platt, D., Hugenholtz, P., Relman, D.A., Quake, S.R., 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proceedings of the National Academy of Sciences of the United States of America 104, 11889-11894.

Mathieu, J., Alvarez, E., Alvarez, P.J.J., 2014. Recombination-assisted megaprimer (RAM) cloning. MethodsX 1, 23-29.

Mavromatis, K., Ivanova, N., Barry, K., Shapiro, H., Goltsman, E., McHardy, A.C., Rigoutsos, I., Salamov, A., Korzeniewski, F., Land, M., Lapidus, A., Grigoriev, I.,

Richardson, P., Hugenholtz, P., Kyrpides, N.C., 2007. Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. Nature methods 4, 495-500.

McCaig, A.E., Grayston, S.J., Prosser, J.I., Glover, L.A., 2001. Impact of cultivation on characterisation of species composition of soil bacterial communities. FEMS microbiology ecology 35, 37-48.

Mijakovic, I., Petranovic, D., Jensen, P.R., 2005. Tunable promoters in systems biology. Current opinion in biotechnology 16, 329-335.

Miller, O.J., Bernath, K., Agresti, J.J., Amitai, G., Kelly, B.T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D.S., Griffiths, A.D., 2006. Directed evolution by in vitro compartmentalization. Nature methods 3, 561-570.

Minagawa, H., Yoshida, Y., Kenmochi, N., Furuichi, M., Shimada, J., Kaneko, H., 2007. Improving the thermal stability of lactate oxidase by directed evolution. Cellular and molecular life sciences : CMLS 64, 77-81.

Ming-Ming, Y., Wei-Wei, Z., Xi-Feng, Z., Pei-Lin, C., 2006. Construction and characterization of a novel maltose inducible expression vector in Bacillus subtilis. Biotechnology letters 28, 1713-1718.

Ming, Y.M., Wei, Z.W., Lin, C.Y., Sheng, G.Y., 2010. Development of a Bacillus subtilis expression system using the improved Pglv promoter. Microbial cell factories 9, 55.

Miyazaki, K., 2011. MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. Methods in enzymology 498, 399-406.

Miyazaki, K., Takenouchi, M., 2002. Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. BioTechniques 33, 1033-1034, 1036-1038.

Mocali, S., Benedetti, A., 2010. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. Research in microbiology 161, 497-505.

Monsalve, M., Calles, B., Mencia, M., Rojo, F., Salas, M., 1998. Binding of phage phi29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase. Journal of molecular biology 283, 559-569.

Morohoshi, T., Oikawa, M., Sato, S., Kikuchi, N., Kato, N., Ikeda, T., 2011. Isolation and characterization of novel lipases from a metagenomic library of the microbial community in the pitcher fluid of the carnivorous plant Nepenthes hybrida. Journal of bioscience and bioengineering 112, 315-320.

Morris, J.J., Kirkegaard, R., Szul, M.J., Johnson, Z.I., Zinser, E.R., 2008. Facilitation of robust growth of Prochlorococcus colonies and dilute liquid cultures by "helper" heterotrophic bacteria. Applied and environmental microbiology 74, 4530-4534.

Mulligan, M.E., Brosius, J., McClure, W.R., 1985. Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. The Journal of biological chemistry 260, 3529-3538.

Nandagopal, N., Elowitz, M.B., 2011. Synthetic biology: integrated gene circuits. Science 333, 1244-1248.

National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications, 2007. In The New Science of Metagenomics: Revealing the

Secrets of Our Microbial Planet, The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. National Academies Press (US).

Nercessian, O., Reysenbach, A.L., Prieur, D., Jeanthon, C., 2003. Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). Environmental microbiology 5, 492-502.

Neubauer, P., Winter, J., 2001 Expression and fermentation strategies for recombinant protein production in Escherichia coli, In Recombinant Protein Production with Prokaryotic and Eukaryotic Cells. A Comparative View on Host Physiology (Merten, O.W., Mattanovich, D., Lang, C., Larsson, G., Neubauer, P., Porro, D., Postma, P., Teixeira de Mattos, J., Cole, J.A. eds.), Recombinant Protein Production with Prokaryotic and Eukaryotic Cells. A Comparative View on Host Physiology. Kluwer Academic Publishers, pp. 195-258.

Neufeld, J.D., Engel, K., Cheng, J., Moreno-Hagelsieb, G., Rose, D.R., Charles, T.C., 2011. Open resource metagenomics: a model for sharing metagenomic libraries. Standards in genomic sciences 5, 203-210.

NIC_Wiki, 2017. http://nic.ucsf.edu/dokuwiki/doku.php?id=fluorescent_proteins.

Nicolaou, S.A., Gaida, S.M., Papoutsakis, E.T., 2011. Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes. Nucleic acids research 39, e152.

Nocadello, S., Swennen, E.F., 2012. The new pLAI (lux regulon based auto-inducible) expression system for recombinant protein production in Escherichia coli. Microbial cell factories 11, 3.

Novagen, 2011. Duet Vectors; User Protocol TB340.

O'Hara, A.M., Shanahan, F., 2006. The gut flora as a forgotten organ. EMBO reports 7, 688-693.

Ochsenreiter, T., Pfeifer, F., Schleper, C., 2002. Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. Extremophiles : life under extreme conditions 6, 267-274.

Oliphant, A.R., Struhl, K., 1988. Defining the consensus sequences of E.coli promoter elements by random selection. Nucleic acids research 16, 7673-7683.

Paddon, C.J., Keasling, J.D., 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nature reviews. Microbiology 12, 355-367.

Pan, S.H., Malcolm, B.A., 2000. Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). BioTechniques 29, 1234-1238.

Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L., Feng, J.X., 2009. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Current microbiology 58, 404-408.

Paul, S., Stang, A., Lennartz, K., Tenbusch, M., Uberla, K., 2013. Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution. Nucleic acids research 41, e29.

Peplow, M., 2016. Synthetic biology's first malaria drug meets market resistance. Nature 530, 389-390.

Pfleger, B.F., Pitera, D.J., Smolke, C.D., Keasling, J.D., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature biotechnology 24, 1027-1032.

Pham, V.H., Kim, J., 2012. Cultivation of unculturable soil bacteria. Trends in biotechnology 30, 475-484.

Pitera, D.J., Paddon, C.J., Newman, J.D., Keasling, J.D., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metabolic engineering 9, 193-207.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M., Mattanovich, D., 2011. Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? Applied microbiology and biotechnology 89, 939-948.

Qiu, D., Damron, F.H., Mima, T., Schweizer, H.P., Yu, H.D., 2008. PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in Pseudomonas and Burkholderia spp. and other bacteria. Applied and environmental microbiology 74, 7422-7426.

Radajewski, S., Ineson, P., Parekh, N.R., Murrell, J.C., 2000. Stable-isotope probing as a tool in microbial ecology. Nature 403, 646-649.

Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Durr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S., Mascher, T., 2013. The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. Journal of biological engineering 7, 29.

Rappe, M.S., Connon, S.A., Vergin, K.L., Giovannoni, S.J., 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418, 630-633.

Rees, H.C., Grant, W.D., Jones, B.E., Heaphy, S., 2004. Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. Extremophiles : life under extreme conditions 8, 63-71.

Registry of Standard Biological Parts, http://parts.igem.org/.

Retallack, D.M., Jin, H., Chew, L., 2012. Reliable protein production in a Pseudomonas fluorescens expression system. Protein expression and purification 81, 157-165.

Riemann, L., Farnelid, H., Steward, G.F., 2010. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. Aquatic Microbial Ecology 61, 235-247.

Riesenfeld, C.S., Schloss, P.D., Handelsman, J., 2004. Metagenomics: genomic analysis of microbial communities. Annual review of genetics 38, 525-552.

Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-943.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., Goodman, R.M., 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied and environmental microbiology 66, 2541-2547.

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in microbiology 5, 172.

Rud, I., Jensen, P.R., Naterstad, K., Axelsson, L., 2006. A synthetic promoter library for constitutive gene expression in Lactobacillus plantarum. Microbiology 152, 1011-1019.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS biology 5, e77.

Saida, F., Uzan, M., Odaert, B., Bontems, F., 2006. Expression of highly toxic genes in E. coli: special strategies and genetic tools. Current protein & peptide science 7, 47-56.

Sait, M., Davis, K.E., Janssen, P.H., 2006. Effect of pH on isolation and distribution of members of subdivision 1 of the phylum Acidobacteria occurring in soil. Applied and environmental microbiology 72, 1852-1857.

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nature biotechnology 27, 946-950.

Samuelson, J., 2011. Bacterial Systems, In Production of Membrane Proteins (Robinson, A.S. ed.), Production of Membrane Proteins. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 11-35.

Schmeisser, C., Steele, H., Streit, W.R., 2007. Metagenomics, biotechnology with non-culturable microbes. Applied microbiology and biotechnology 75, 955-962.

Schwarz, D., Dotsch, V., Bernhard, F., 2008. Production of membrane proteins using cell-free expression systems. Proteomics 8, 3933-3946.

Sepp, A., Choo, Y., 2005. Cell-free selection of zinc finger DNA-binding proteins using in vitro compartmentalization. Journal of molecular biology 354, 212-219.

Shetty, R.P., Endy, D., Knight, T.F., Jr., 2008. Engineering BioBrick vectors from BioBrick parts. Journal of biological engineering 2, 5.

Shinomiya, K., Mori, T., Aoyama, Y., Sera, T., 2011. Unidirectional cloning by cleaving heterogeneous sites with a single sandwiched zinc finger nuclease. Biochemical and biophysical research communications 414, 733-736.

Siegele, D.A., Hu, J.C., 1997. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proceedings of the National Academy of Sciences of the United States of America 94, 8168-8172.

Sierro, N., Makita, Y., de Hoon, M., Nakai, K., 2008. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic acids research 36, D93-96.

Silverstone, A.E., Arditti, R.R., Magasanik, B., 1970. Catabolite-insensitive revertants of lac promoter mutants. Proceedings of the National Academy of Sciences of the United States of America 66, 773-779.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., Bali, V., Batra, N., 2009. Metagenomics: Concept, methodology, ecological inference and recent advances. Biotechnology journal 4, 480-494.

Siuti, P., Yazbek, J., Lu, T.K., 2013. Synthetic circuits integrating logic and memory in living cells. Nature biotechnology 31, 448-452.

Skerra, A., 1994. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. Gene 151, 131-135.

Sleator, R.D., Shortall, C., Hill, C., 2008. Metagenomics. Letters in applied microbiology 47, 361-366.

Smolke, C.D., Carrier, T.A., Keasling, J.D., 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Applied and environmental microbiology 66, 5399-5405.

Somerville, R.L., 1988. The trp promoter of Escherichia coli and its exploitation in the design of efficient protein production systems. Biotechnology & genetic engineering reviews 6, 1-41.

Song, J., Oh, H.M., Cho, J.C., 2009. Improved culturability of SAR11 strains in dilution-to-extinction culturing from the East Sea, West Pacific Ocean. FEMS microbiology letters 295, 141-147.

Sorensen, H.P., Mortensen, K.K., 2005. Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of biotechnology 115, 113-128.

Stevenson, B.S., Eichorst, S.A., Wertz, J.T., Schmidt, T.M., Breznak, J.A., 2004. New strategies for cultivation and detection of previously uncultured microbes. Applied and environmental microbiology 70, 4748-4755.

Stewart, E.J., 2012. Growing unculturable bacteria. Journal of bacteriology 194, 4151-4160.

Stougaard, P., Jorgensen, F., Johnsen, M.G., Hansen, O.C., 2002. Microbial diversity in ikaite tufa columns: an alkaline, cold ecological niche in Greenland. Environmental microbiology 4, 487-493.

Stout, T.J., Schellenberger, U., Santi, D.V., Stroud, R.M., 1998. Crystal structures of a unique thermal-stable thymidylate synthase from Bacillus subtilis. Biochemistry 37, 14736-14747.

Studier, F.W., 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. Journal of molecular biology 219, 37-44.

Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. Protein expression and purification 41, 207-234.

Studier, F.W., 2014. Stable expression clones and auto-induction for protein production in E. coli. Methods in molecular biology 1091, 17-32.

Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of molecular biology 189, 113-130.

Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M., Elowitz, M.B., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545-550.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome research 20, 1605-1612.

Taupp, M., Mewis, K., Hallam, S.J., 2011. The art and design of functional metagenomic screens. Current opinion in biotechnology 22, 465-472.

Tawfik, D.S., Griffiths, A.D., 1998. Man-made cell-like compartments for molecular evolution. Nature biotechnology 16, 652-656.

Tee, K.L., Wong, T.S., 2013. Polishing the craft of genetic diversity creation in directed evolution. Biotechnology advances 31, 1707-1721.

Tee, K.L., Wong, T.S., 2014. Directed evolution: A powerful algorithm for advancing synthetic biology, In Applied Synthetic Biology (Singh, V. ed.), Applied Synthetic Biology. Gyan Books Pvt. Ltd., pp. 465-497.

Tegel, H., Ottosson, J., Hober, S., 2011. Enhancing the protein production levels in Escherichia coli with a strong promoter. The FEBS journal 278, 729-739.

Terpe, K., 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology 72, 211-222.

The Royal Society, 2007. https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/synthetic-biology/CallForViews.pdf.

Thomas, D.N., Dieckmann, G.S., 2002. Antarctic Sea ice - a habitat for extremophiles. Science 295, 641-644.

Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. Microbial informatics and experimentation 2, 3.

Thuy Le, A.T., Schumann, W., 2007. A novel cold-inducible expression system for Bacillus subtilis. Protein expression and purification 53, 264-269.

Topp, S., Reynoso, C.M., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., Scott, J.R., Gallivan, J.P., 2010. Synthetic riboswitches that induce gene expression in diverse bacterial species. Applied and environmental microbiology 76, 7881-7884.

Turner, N.J., 2009. Directed evolution drives the next generation of biocatalysts. Nature chemical biology 5, 567-573.

Tyo, K.E., Ajikumar, P.K., Stephanopoulos, G., 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nature biotechnology 27, 760-765.

Uchiyama, T., Abe, T., Ikemura, T., Watanabe, K., 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nature biotechnology 23, 88-93.

Ulrich, A., Andersen, K.R., Schwartz, T.U., 2012. Exponential megapriming PCR (EMP) cloning--seamless DNA insertion into any target plasmid without sequence constraints. PloS one 7, e53360.

Uphoff, H.U., Felske, A., Fehr, W., Wagner-Dobler, I., 2001. The microbial diversity in picoplankton enrichment cultures: a molecular screening of marine isolates. FEMS microbiology ecology 35, 249-258.

Valdez-Cruz, N.A., Caspeta, L., Perez, N.O., Ramirez, O.T., Trujillo-Roldan, M.A., 2010. Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters. Microbial cell factories 9, 18.

van den Ent, F., Lowe, J., 2006. RF cloning: a restriction-free method for inserting target genes into plasmids. Journal of biochemical and biophysical methods 67, 67-74.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends in genetics : TIG 30, 418-426.

Van Noorden, R., 2010. Demand for malaria drug soars. Nature 466, 672-673.

Vartoukian, S.R., Palmer, R.M., Wade, W.G., 2010. Strategies for culture of 'unculturable' bacteria. FEMS microbiology letters 309, 1-7.

Vishnivetskaya, T., Kathariou, S., McGrath, J., Gilichinsky, D., Tiedje, J.M., 2000. Low-temperature recovery strategies for the isolation of bacteria from ancient permafrost sediments. Extremophiles : life under extreme conditions 4, 165-173.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nature biotechnology 32, 992-1000.

Wang, B.L., Jiao, Y.L., Li, X.X., Zheng, F., Liang, H., Sun, Z.Y., Guo, G., 2009a. A universal method for directional cloning of PCR products based on asymmetric PCR. Biotechnology and applied biochemistry 52, 41-44.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., Church, G.M., 2009b. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894-898.

Wanner, B.L., Kodaira, R., Neidhardt, F.C., 1977. Physiological regulation of a decontrolled lac operon. Journal of bacteriology 130, 212-222.

Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B., Gill, R.T., 2010. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nature biotechnology 28, 856-862.

Waschkowitz, T., Rockstroh, S., Daniel, R., 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. Applied and environmental microbiology 75, 2506-2516.

Weickert, M.J., Doherty, D.H., Best, E.A., Olins, P.O., 1996. Optimization of heterologous protein production in Escherichia coli. Current opinion in biotechnology 7, 494-499.

Wentzel, A., Lewin, A., Cervantes, F.J., Valla, S., Kotlar, H.K., 2013. Deep Subsurface Oil Reservoirs as Poly-extreme Habitats for Microbial Life. A Current Review, In Polyextremophiles, Polyextremophiles. Springer, pp. 439-466.

Westers, L., Westers, H., Quax, W.J., 2004. Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. Biochimica et biophysica acta 1694, 299-310.

Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences of the United States of America 95, 6578-6583.

Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H.K., Handelsman, J., 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Applied and environmental microbiology 71, 6335-6344.

Wilson, M.C., Piel, J., 2013. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chemistry & biology 20, 636-647.

Wong, T.S., Arnold, F.H., Schwaneberg, U., 2004a. Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. Biotechnology and bioengineering 85, 351-358.

Wong, T.S., Tee, K.L., Hauer, B., Schwaneberg, U., 2004b. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic acids research 32, e26.

Wooley, J.C., Godzik, A., Friedberg, I., 2010. A primer on metagenomics. PLoS computational biology 6, e1000667.

Wu, H., Pei, J., Jiang, Y., Song, X., Shao, W., 2010. pHsh vectors, a novel expression system of Escherichia coli for the large-scale production of recombinant enzymes. Biotechnology letters 32, 795-801.

Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M., Zhou, J., 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Applied and environmental microbiology 67, 5780-5790.

Xing, M.N., Zhang, X.Z., Huang, H., 2012. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnology advances 30, 920-929.

Yadav, V.G., De Mey, M., Lim, C.G., Ajikumar, P.K., Stephanopoulos, G., 2012. The future of metabolic engineering and synthetic biology: towards a systematic practice. Metabolic engineering 14, 233-241.

Yansura, D.G., Henner, D.J., 1984. Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America 81, 439-443.

Yao, A.I., Fenton, T.A., Owsley, K., Seitzer, P., Larsen, D.J., Sit, H., Lau, J., Nair, A., Tantiongloc, J., Tagkopoulos, I., Facciotti, M.T., 2013. Promoter element arising from the fusion of standard BioBrick parts. ACS synthetic biology 2, 111-120.

Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., Yanagawa, H., 2003. DNA display for in vitro selection of diverse peptide libraries. Nucleic acids research 31, e118.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C.S., Li, H., Mashiyama, S.T., Joachimiak, M.P., van Belle, C., Chandonia, J.M., Soergel, D.A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B.J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J.E., Taylor, S.S., Strausberg, R.L., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS biology 5, e16.

You, L., Arnold, F.H., 1996. Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. Protein engineering 9, 77-83.

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., Keller, M., 2002. Cultivating the uncultured. Proceedings of the National Academy of Sciences of the United States of America 99, 15681-15686.

Zhang, Y.X., Perry, K., Vinci, V.A., Powell, K., Stemmer, W.P., del Cardayre, S.B., 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415, 644-646.

Zhao, H., Chockalingam, K., Chen, Z., 2002. Directed evolution of enzymes and pathways for industrial biocatalysis. Current opinion in biotechnology 13, 104-110.

Zhao, S., Wang, J., Bu, D., Liu, K., Zhu, Y., Dong, Z., Yu, Z., 2010. Novel glycoside hydrolases identified by screening a Chinese Holstein dairy cow rumen-derived metagenome library. Applied and environmental microbiology 76, 6701-6705.

Zhou, J., Bruns, M.A., Tiedje, J.M., 1996. DNA recovery from soils of diverse composition. Applied and environmental microbiology 62, 316-322.

Zhu, D., Zhong, X., Tan, R., Chen, L., Huang, G., Li, J., Sun, X., Xu, L., Chen, J., Ou, Y., Zhang, T., Yuan, D., Zhang, Z., Shu, W., Ma, L., 2010. High-throughput cloning of human liver complete open reading frames using homologous recombination in Escherichia coli. Analytical biochemistry 397, 162-167.