A Thesis for the Degree of Ph.D. in Engineering

Compartmentalized replication for directed evolution of DNA polymerase and transcriptional regulator

August 2018

Graduate School of Science and Technology

Keio University

Seaim Lwin Aye

Table of Contents

List of Tables

List of Figures

Abbreviations

As(III)	arsenite
As(V)	arsenate
ASV	Anodic stripping voltammetry
B. subtilis	*Bacillus subtilis*
CBB	coomassie brilliant blue
CPR	compartmentalized partnered replication
CSR	compartmentalized self-replication
Cs	cold-sensitive
CSV	Cathodic stripping voltammetry
E. coli	*Escherichia coli*
epPCR	error-prone PCR
Ex	extension
ETAAS	Electrothermal atomic absorption spectroscopy
FACS	fluorescence activated cell sorting
GFP	green fluorescence protein
GC-MS	Gas chromatography-mass spectrometry
ICP-MS	Inductively coupled plasma-mass spectrometry
ICP-OES/AES	Inductively coupled plasma-atomic emission spectroscopy
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITCHY	incremental truncation for the creation of hybrid enzymes
OD	optical density
PCR	polymerase chain reaction
Pfu	*Pyrococcus furiosus*
pol	polymerase
Rs	resistant
SDS-PAGE	sodium lauryl sulfate-polyacrylamide gel electrophoresis
StEP	staggered extension process
Taq	*Thermus aquaticus*
Tth	*Thermus thermophilus*
UV	Ultroviloet
WT	wild-type

Chapter 1
Introduction

1.1 Protein engineering

During biological evolution, organisms have overcome the obstacles and challenges in the nature and adapted to the environment for survival through natural selection. Natural selection is a process in which the individuals with the best genetic trait pass down their genetic material to their offspring. Consequently, these heritable traits that are required to adapt to the natural environment are maintained and enriched over time. For centuries, humans have used artificial selection also known as selective breeding in order to access useful and desired phenotypes more frequently. In artificial selection, the reproductive cycle is intervened by implying a selection pressure for the desired traits such as breeding of crops (Wright et al. 2005) or farming domesticated animals (Driscoll et al. 2009).

Enzymes are the solution to complete chemical processes at a specific time under conditions relevant for cellular life. Enzymes can accelerate chemical reactions by up to seventeen orders of magnitude by making the use of quantum mechanics, thermodynamics and macromolecular dynamics. Scientists have been applying native enzymes as biocatalysts in the chemical laboratory over a century. While some natural enzymes remain relevant until today, there are some limitations since high specificity and selectivity restrict the usage of enzyme beyond its natural substrate. To adapt enzymes for the desired conditions, the ability to remodel the natural existing enzymes or to develop entirely new biocatalysts has become a challenge as well as an opportunity for the scientists.

In general, protein engineering is the process of developing proteins with useful or desired properties. Example of such properties include (i) enhanced stability or function under restricted conditions such as temperature, pH, organic solvent and salt, (ii) the altered specificity to the enzyme substrate or epitope binding properties and (iii) enhanced enzymatic rate. There are two general strategies for protein engineering; rational design and directed evolution (**Figure 1-1**). For one to perform protein engineering by rational design, there are prerequisites to put into consideration. In order to perform rational design of a target protein, one requires to prepare (i) a cloned gene coded for the enzyme, (ii) sequence of the gene, (iii) the role the active site *i.e.*, which amino acids in the sequence are involved in the activity and (iv) the crystal or NMR structure of the enzyme or another protein with high structural homology. Based on the available information, one can have an idea of

which amino acid(s) to mutate so that it has the likelihood of resulting an improved or desired effect.

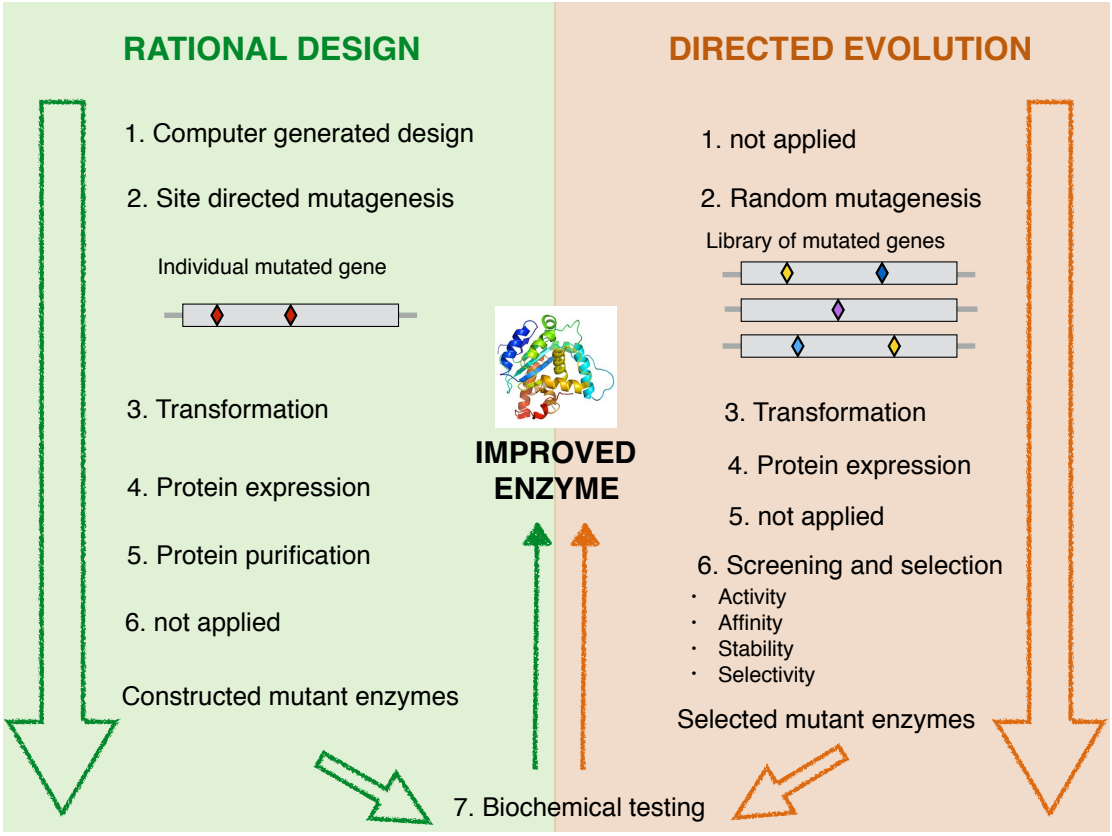


Figure 1-1. Two general strategies for protein engineering; rational design and directed evolution. In rational design, a scientist uses detailed knowledge of the structure and function of a protein to make desired changes. However, its major drawback is that detailed structural knowledge of a protein is often unavailable and even when available, it is very difficult to predict the effects of various mutations. In directed evolution, random mutagenesis, *e.g.* error prone PCR or sequence saturation mutagenesis, is applied to a protein gene of interest, and a selection regime is used to select variants having desired traits mimicking natural evolution, often producing superior results to rational design. On the other hand, directed evolution mimic the natural evolution theory without the knowledge of the important information about the protein such as structure and mechanism of action. Although the potential of computational power in rational design can access to complex designs, initial designs remain suboptimal and require directed evolution to achieve high activity.

1.2 Key steps in directed evolution

Recently, directed evolution in the laboratory has proved to be a promising approach for optimizing the enzymes and proteins via manipulating individual genes and gene products, that are fundamental building blocks in biology. Directed evolution is a technique used in protein engineering that mimics the framework of natural selection in the laboratory with the purpose of optimizing and evolving gene to develop a novel protein with desired properties (Goldsmith and Tawfik 2012; Lutz 2010; Molina-Espeja et al. 2016; Porcar 2010). The work related to directed evolution could be traced back to 70 years ago. Evolution of a single gene in the laboratory requires carefully designed strategies for screening and selection of functional variants in a way that they maintain the genotype to phenotype relationship (**Figure 1-2**). While considering the directed evolution, the key steps in the cycle starts with the generation of diverse library, which is translated into a corresponding library of gene products and screening and selection after iterative rounds of mutagenesis for the functional variants that maintain the correspondence between genes and their functions.

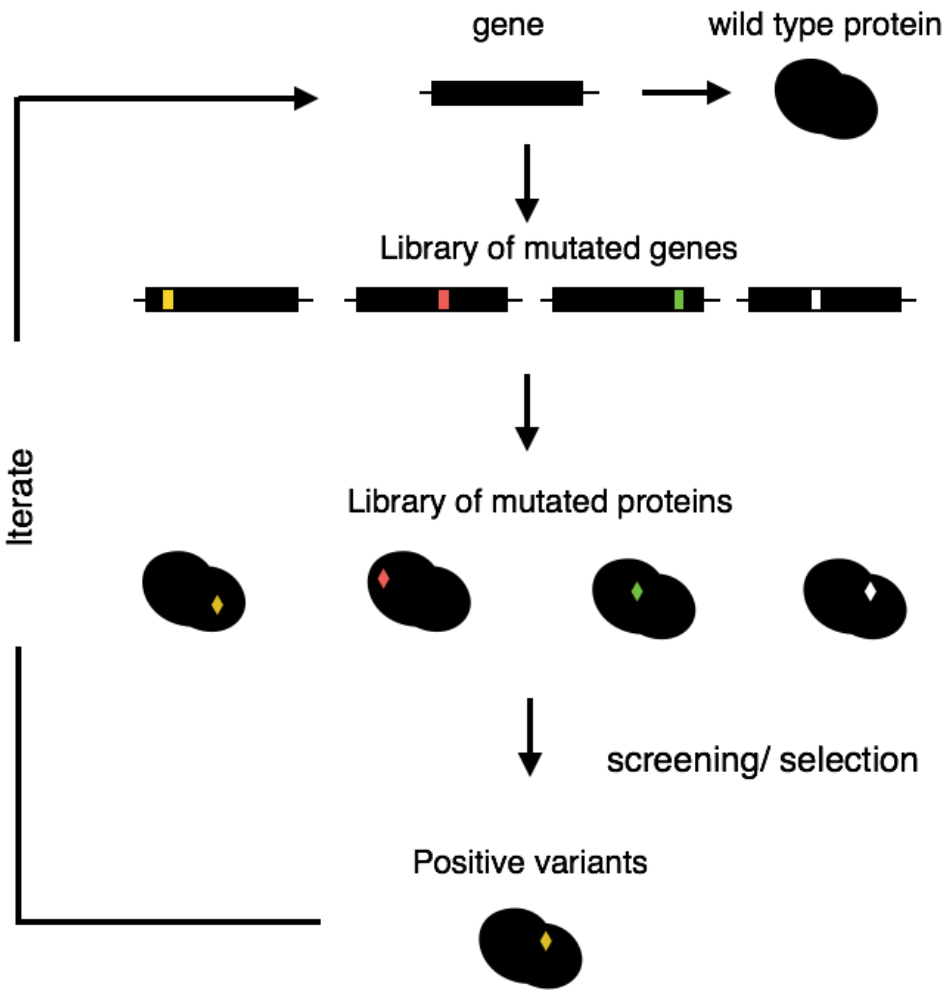


Figure 1-2. Principle of directed evolution of proteins. In the process of directed evolution, a library of genes created by random mutagenesis is translated into a corresponding library of gene products and screened or selected for functional variants in a manner that maintains the linkage between genotype (DNA) and phenotype (protein and their functions). These functional genes are used as starting points for next rounds of diversification and screening or selection.

1.2.1 Diversification step

Random mutagenesis is a strategy of diversification that introduces mutations in an unbiased manner throughout the entire gene. Non-recombinant methods to generate random mutagens include oligonucleotide directed mutagenesis (Sugimoto et al. 1989; Zoller and Smith 1984, 1987), chemical mutagenesis (Kadonaga and Knowles 1985), bacterial mutator strains (Greener et al. 1996; Muteeb and Sen 2010) and error-prone PCR (epPCR) (McCullum et al. 2010). epPCR uses a modified polymerase reaction with a low fidelity DNA polymerase, an increased magnesium concentration and supplementation with manganese. It can increase the mutation rates to 10^{-4} – 10^{-3} per replicated bases (Eckert and Kunkel 1990). The main advantages of epPCR is easy to implement and can provide high mutation rates with relatively even mutation spectrum. However, epPCR can only provide random mutagenesis at the nucleotide level and does not sample amino acid codon space evenly (**Figure 1-3**). The recombinant methods to eliminate neutral and deleterious mutations involve DNA shuffling (Coco et al. 2001), in vivo recombination in yeast (Prado and Aguilera 1994), random priming recombination (Esteban et al. 2003), staggered extension process (StEP) (Zhao et al. 1998), and incremental truncation for the creation of hybrid enzymes (ITCHY) (Patrick and Gerth 2014). Unlike DNA fragmentation-based methods, like epPCR, StEP is a modified PCR protocol where the elongation step is interrupted by heat denaturation (**Figure 1-3**). The technique is a modified polymerase reaction with very short (approximately 10 seconds) cycles in which the elongation of DNA is very quick (a few hundred base pairs). Incomplete extension products are switched between templates by subsequent annealing resulting the recombination of multiple DNA templates into one amplicon. In climbing the fitness landscape of directed evolution (**Figure 1-4**), the optimal diversification approach such as combining the site-directed mutagenesis and the recombinant method can favor in attempt to reach the absolute maximal peak over the course of generations (**Figure 1-3**).

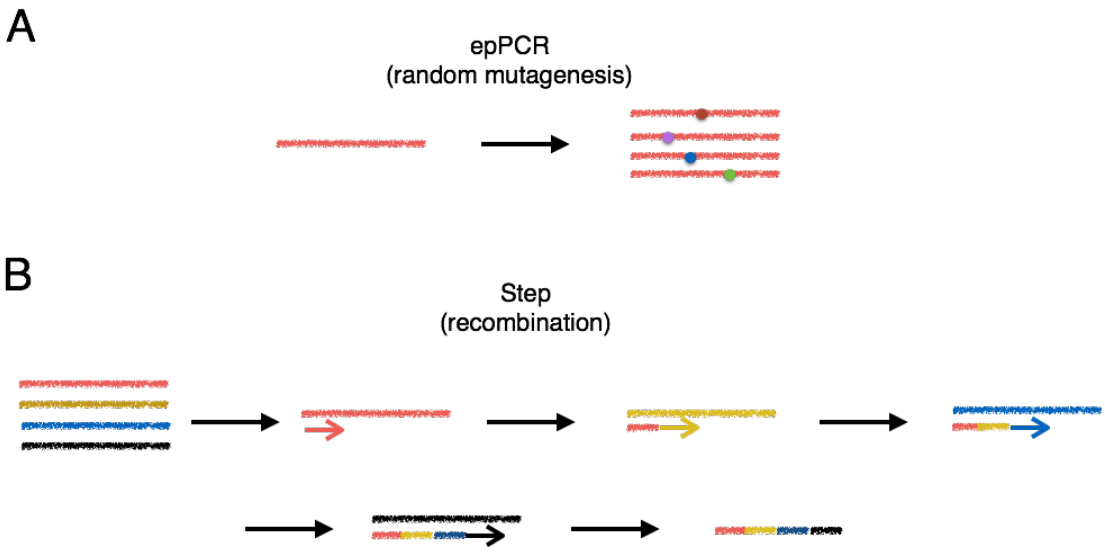


Figure 1-3. Typical approach to library diversification. While random mutagenesis such as epPCR provides mutations at nucleotide level (A), recombinant methods such as StEP (Zhao 2004) can shuffle those nucleotides to sequence level (B). Diversification over the course of generation can be optimized in order to obtain mutant with highest possible activity.

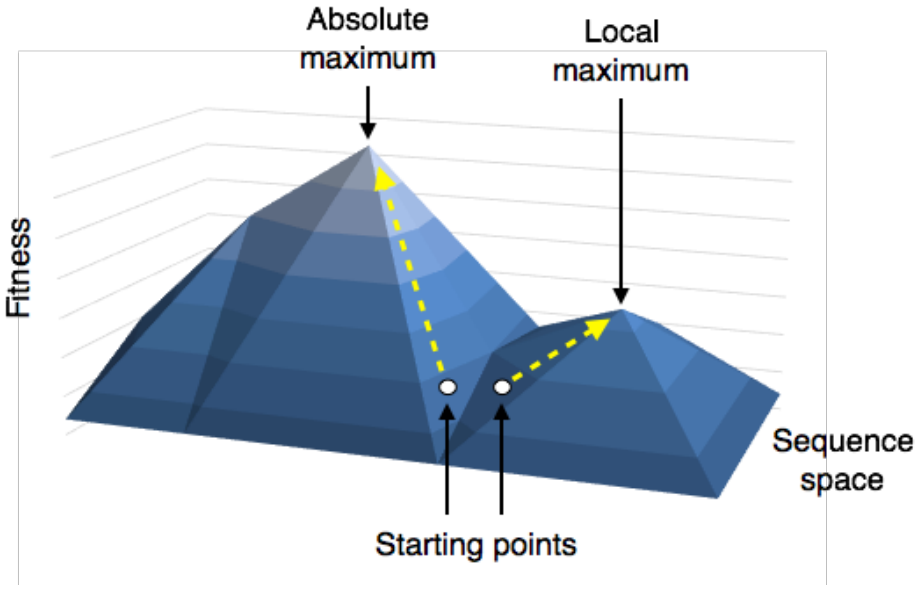


Figure 1-4. Schematic representation of fitness landscape. A series of steps in directed evolution can be visualized as a three-dimensional fitness landscape. Library generation resembles the starting point at the base and selection or screening identifies the function climbing towards peak. Directed evolution can arrive absolute maximum or can be trapped at local maximum where diversification is insufficient to cross fitness valleys and access neighboring peak.

1.2.2 Screening/Selection step

The choice of a screening or selection can be illustrated as decision tree that operates based on the properties of the protein and phenotype to be evolved. Screening is the inspection of individual library for a desired phenotype where organisms with altered phenotypes are characterized by high-throughput DNA sequencing to identify the responsible mutations. On the other hand, selection does not require identifying each library member individually.

Current selection strategies involve selection for binding affinity using phage display (McCafferty et al. 1990), mRNA display (Shiratori et al. 2009), ribosome display (Hanes and Pluckthun 1997) and cell surface display (Bessette et al. 2004). On the other hand, examples of selection for enzyme activity are organismal survival as a basis for selection such as antibiotic resistance (Orencia et al. 2001; Palmer and Kishony 2013), and selection within *in vitro* compartments (Takeuchi et al. 2014; Abil et al. 2017; Ellefson et al. 2014; Ghadessy et al. 2001).

Selections for DNA and RNA polymerases in water-in-oil emulsion droplets are referred as compartmentalized self-replication (CSR) (Ghadessy et al. 2001). In this selection strategy, the polymerases that possess best replicating efficiency encode the most active gene, which are enriched after selection over rounds. The scope of CSR was widened and extends its application to enzymes more than polymerase when CPR was reported by (Ellefson et al. 2014). In CPR, the evolution itself targets another enzyme that controls the expression of DNA polymerase. The more active gene will drive better expression of DNA polymerase leading to enrichment of the target gene over time.

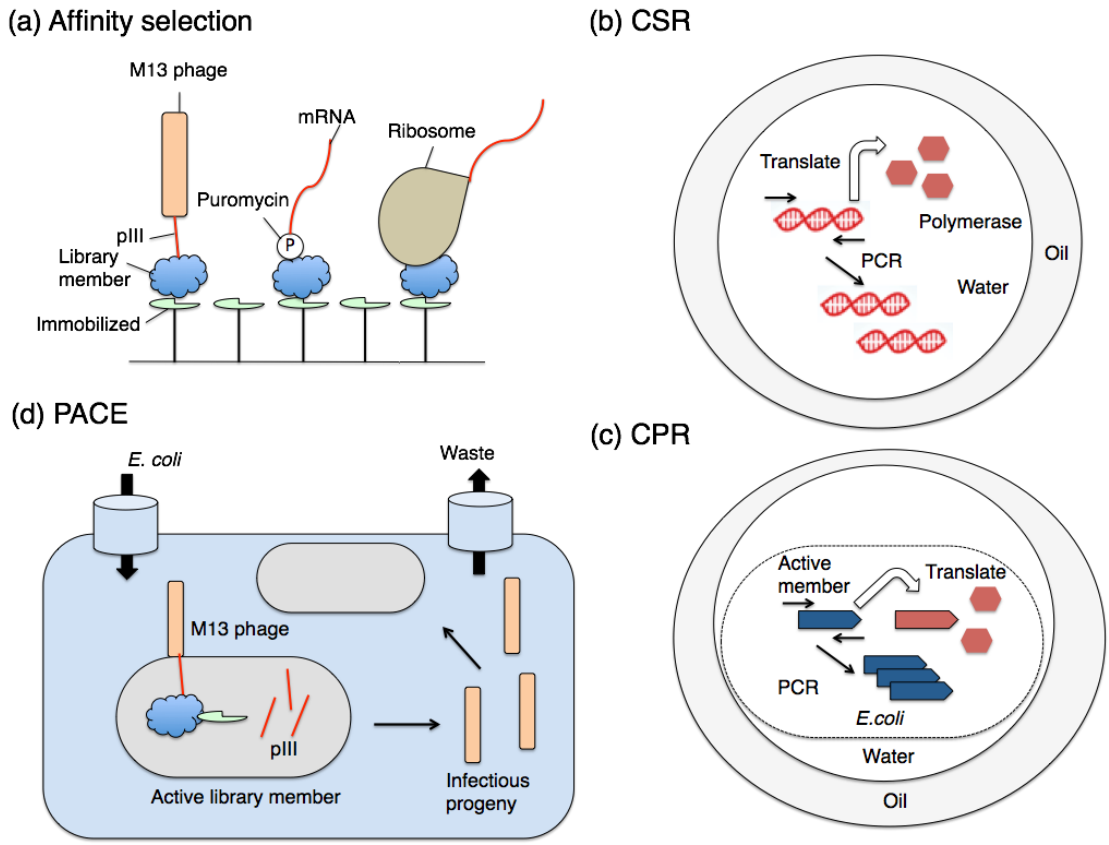


Figure 1-5. Selection methods for evolution of protein (Packer and Liu 2015). (a) Affinity selection methods include display on phage particles via protein pIII (left), mRNA display via puromycin linkage (middle) and ribosome display (right). In this method, the selection indetifies library members that bind to an immobilized target. (b) CSR selects for polymerases thany can amplify their own genes within the water emulsion droplest. (c) In CPR, the evolving activity trigger the polymerase expression. Higher polymerase expressionleads to the better PCR amplification of the active members. (d) In phage-associated continuous evolution (PACE), active variants trigger the expression of host cell protein and inactive variants are diluted out of the vessel.

1.3 Objective of this study

This thesis focuses on developing evolutionary strategies for constructing synthetic genetic circuits, an approach I believe to be generally applicable for optimizing proteins or enzymes. Engineering proteins in cells is plausible to work with challenging circuit designs because of the controllable cell system to evolve the desired outcome. This important feature helps to compensate the designer's lack of knowledge on sequence-function relationships in programming cellular behavior. Different evolutionary strategies can be applied to genetic circuits. Compartmentalized replication is one of the most powerful methods for directed evolution of proteins by using *in vitro* compartmentalization (water-in-oil emulsion) for genotype-phenotype linkage and PCR amplification for selection. This method was first developed for directed evolution of several DNA polymerases by CSR, and then applied to two other proteins T7 RNA polymerase and tRNA-synthetase by CPR (**Figure 1-6**).

However, the utility of CSR and CPR is somewhat limited by low efficiency of emulsion PCR with the wild-type *Taq* DNA polymerase so far used. To overcome this problem, in this study, I first performed engineering of *Tth* DNA polymerase with improved efficiency by using CSR (Chapter 2). To increase the polymerase efficiency, I applied CSR technique to evolve faster-cycling *Tth* mutant that can self-replicate under progressively limited number of PCR cycle. As highlighted in Chapter 2, this work resulted *Tth* mutant with a significantly beneficial attribute to increased polymerization rate. Next I applied this improved DNA polymerase for selection of an arsenic-responsive transcriptional regulator by using CPR (Chapter 3). I employed CPR as the first step in a dual selection of *arsR* mutants along with cell survival selection as the second step. Here I proved the functioning dual selection approach as an evolutionary strategy for optimizing genetic circuits. The design stagey I envision as being most effective for engineering functional circuits was described in this report.

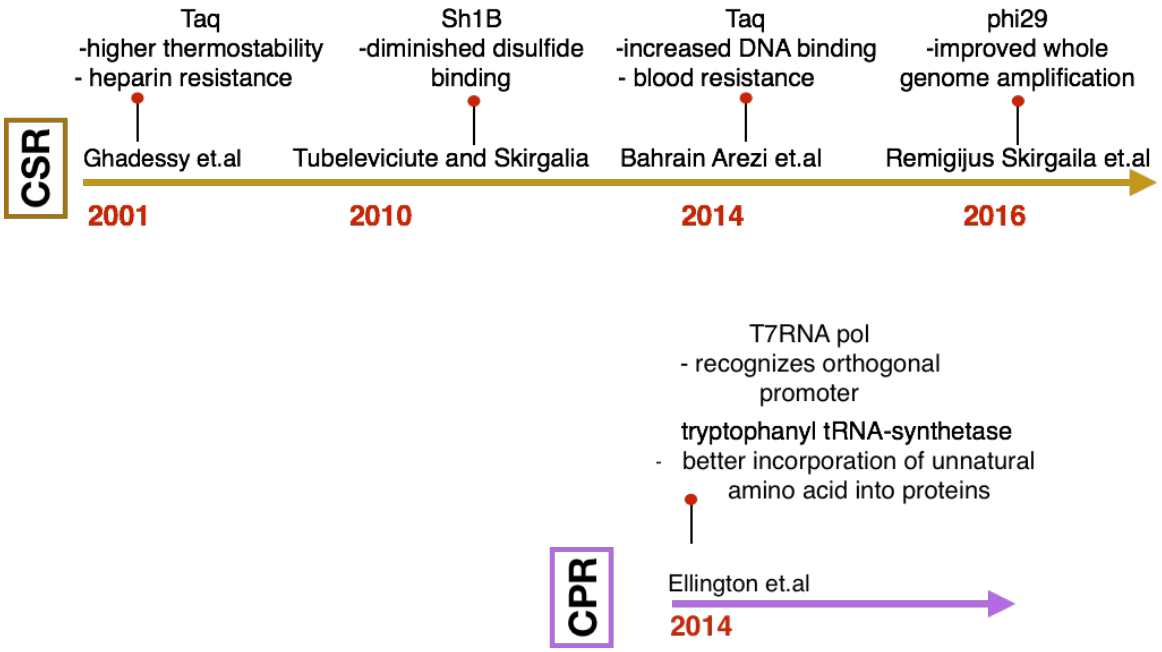


Figure 1-6. Timeline of the development of CSR and CPR techniques.

Chapter 2

Engineering of DNA polymerase I from *Thermus thermophilus* using compartmentalized self-replication

2.1 Background and rationale

Polymerase chain reaction (PCR), a powerful technique for the amplification of DNA, has been elaborated upon in many ways since its introduction in 1985 (Saiki et al. 1985). The applications of PCR range broadly across many research fields, including medicine and forensics (Moore 2005). Since the first description and presentation of PCR at Cold Spring Harbor (Mullis et al. 1986; Mullis and Faloona 1987), scientists have been attempting to improve PCR techniques. A major breakthrough was the replacement of DNA polymerase (pol) with an enzyme from a thermostable species to eliminate the requirement of new enzyme addition to the reaction after each cycle (Saiki et al. 1988). Several thermostable DNA polymerases have been isolated from a handful of thermophilic bacteria, including *Thermus aquaticus* (*Taq*) (Innis et al. 1988), *Thermus thermophilus* (*Tth*) (Carballeira et al. 1990; Glukhov et al. 1990; Myers and Gelfand 1991; Ribble et al. 2015) and *Pyrococcus furiosus* (*Pfu*) (Lundberg et al. 1991). These enzymes are extremely thermostable at a high temperature of 95°C. Among the commercially available enzymes, *Tth* DNA polymerase (*Tth* pol) possesses reverse transcriptase activity in the presence of Mn^{2+} ions and can thus be used in RT-PCR. *Tth* pol shows optimal activity at temperatures between 70°C and 74°C, with an extension rate of 1.5 kb/min (Carballeira et al. 1990). Although its half-life is comparatively shorter (20 min at 95°C), *Tth* pol tolerates higher concentrations of inhibitory components than *Taq* pol (Al-Soud and Radstrom 2001).

Despite the promising performance of PCR, the remaining limitations to PCR are its product yield, the length of DNA that can be amplified, the speed of the polymerase, and the fidelity of the process. Ultimately, PCR run time is solely influenced by the kinetic properties of the PCR enzyme. Currently, faster PCR throughput can be achieved by reducing the time of the denaturation, annealing and extension steps or by using 2-step cycling that combines the annealing and extension steps. However, shorter cycle times yield lower detection signals and higher failure rates. Alternatively, increasing the enzyme concentration, adjusting buffer conditions and reducing the reaction volume can all improve the resultant yields. Another approach to enhance the processivity of polymerase is its fusion to a sequence non-specific dsDNA-binding protein, such as Sso7d from *Sulfolobus solfataricus*, whose

fusion does not affect the catalytic activity and thermal stability of the *Taq* and *Pfu* polymerases but enhances their processivity (Wang et al. 2004). Sso7d has a small mass of 7 kDa and its function in the native organism is the stabilization of the genomic DNA and therefore, facilitation of the action of DNA polymerase.

In this chapter, I attempted to improve the performance of DNA polymerase I from *Tth* HB27 (Henne et al. 2004) for compartmentalized self-replication (CSR) (Ghadessy et al. 2001). CSR is a powerful method for directed evolution of polymerase with desired functions such as enhanced thermostability, increased resistance to inhibitors and increased DNA-binding activity (Arezi et al. 2014; Povilaitis et al. 2016; Tubeleviciute and Skirgaila 2010). CSR is performed by compartmentalizing self-replication reactions into individual non-interactive aqueous compartments of a water-in-oil emulsion to ensure a linkage between genotype and phenotype **(Figure 2-1**). In CSR, the individual polymerases are encapsulated in separate compartments. Each polymerase replicates only its own encoding gene to the exclusion of those in other compartment s in the presence of primers. The reactions take place in individual aqueous compartments of a water-in-oil emulsion that are stable for prolonged periods of times at temperatures exceeding 90°C. Recently, the CSR concept was expanded to compartmentalized partnered replication (CPR) for the directed evolution of genetic circuits that can be linked to the production of *Taq* pol (Ellefson et al. 2014). I expect that the improved *Tth* pol mutants constructed in this chapter will also be useful for the emulsion PCR step of CPR.

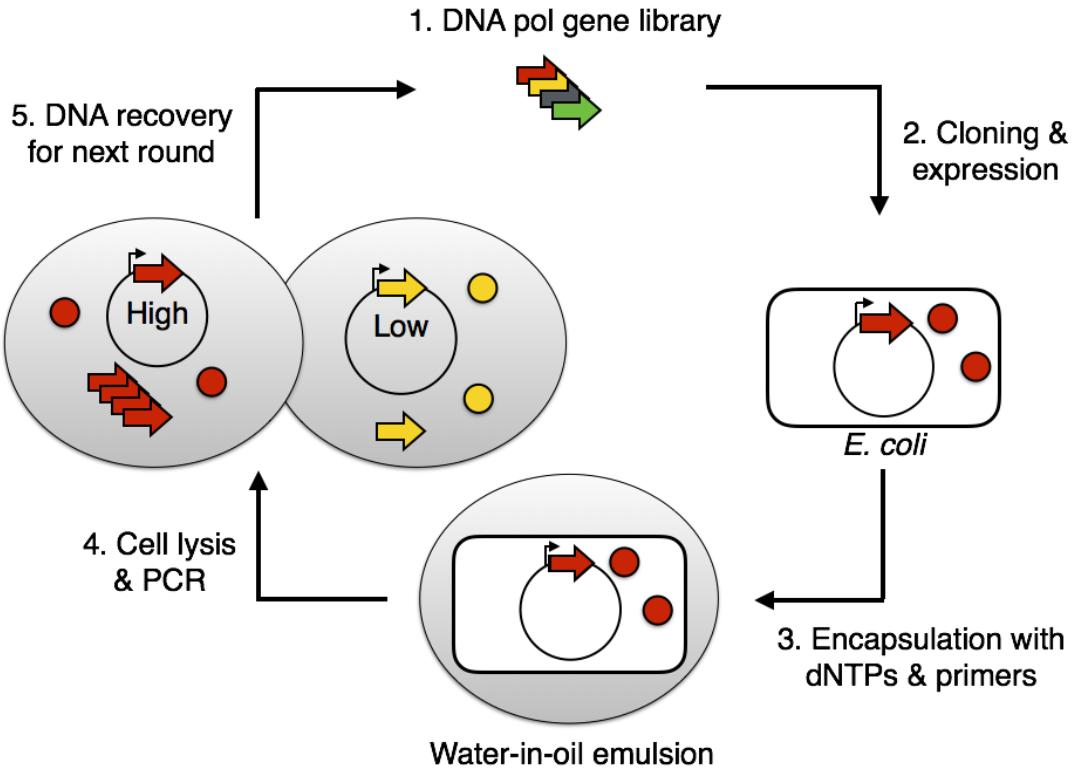


Figure 2-1. Schematic representation of the directed evolution of DNA polymerase (pol) using CSR (Ghadessy et al. 2001). (1) A randomly-mutated DNA pol gene library was prepared by error-prone PCR. (2) The library was cloned and transformed into *E. coli* cells, and the expression of DNA pol was induced. (3) After harvesting, the cells were resuspended in PCR buffer with dNTPs and primers, and then encapsulated in a water-in-oil emulsion. (4) After the heat lysis of each cell in each micelle, genes encoding highly active DNA pol mutants were amplified by their own enzyme. (5) After the emulsion-PCR, the amplified genes were recovered and cloned for further successive selection rounds.

2.2 Results

2.2.1 Strategy to modify the *Tth* DNA polymerase

The wild-type DNA polymerase I gene from *T. thermophilus* HB27 (NCBI-protein ID: AAS81038) is 2505 nucleotides long with a GC content of 67.5% (Henne et al. 2004). Although I first used this polymerase expressed in *E. coli* as a source of enzyme to amplify short and long target DNA sequences containing the polymerase gene itself, no amplification was observed for either condition (data not shown). Because the reason of its failure to self-replicate might be the long and high GC content of the wild-type *Tth* pol gene, I tried to improve the amplification efficiency of *Tth* pol by modification and site-directed mutagenesis. Since the reason of its failure to self-replicate might be the long and high GC content of the wild-type *Tth* pol gene, I tried to improve the amplification efficiency of *Tth* pol by modification and site-directed mutagenesis.

The modification procedure included the following approaches: (i) deletion of the N-terminal domain, (ii) fusion with a DNA-binding protein, (iii) introduction of mutations known to improve *Taq* pol, and (iv) codon optimization to reduce GC content. Previous studies with the deletion of the 5′-3′exonuclease domain of thermostable DNA polymerases confirmed the functional role of truncation in polymerase modification (Lawyer et al. 1993; Lin et al. 1994; Arakawa et al. 1996; Villbrandt et al. 1997; Westberg et al. 1999; Yang et al. 2008). Therefore, I first deleted the N-terminal 280 amino acids of the *Tth* pol. Then, to improve the processivity and DNA binding activity of the enzyme, the resultant truncated *ΔTth* was fused with the DNA binding protein Sso7d from *Sulfolobus solfataricus* (Wang et al. 2004). Sso7d binds to dsDNA as a monomer without marked sequence preference. In theory, by linking to DNA polymerase, Sso7d could provide additional contact with DNA and and consequently enhance the processivity. The resulting DNA polymerase SΔ*Tth* is a protein of 618 amino acids (**Figure 2-2**).

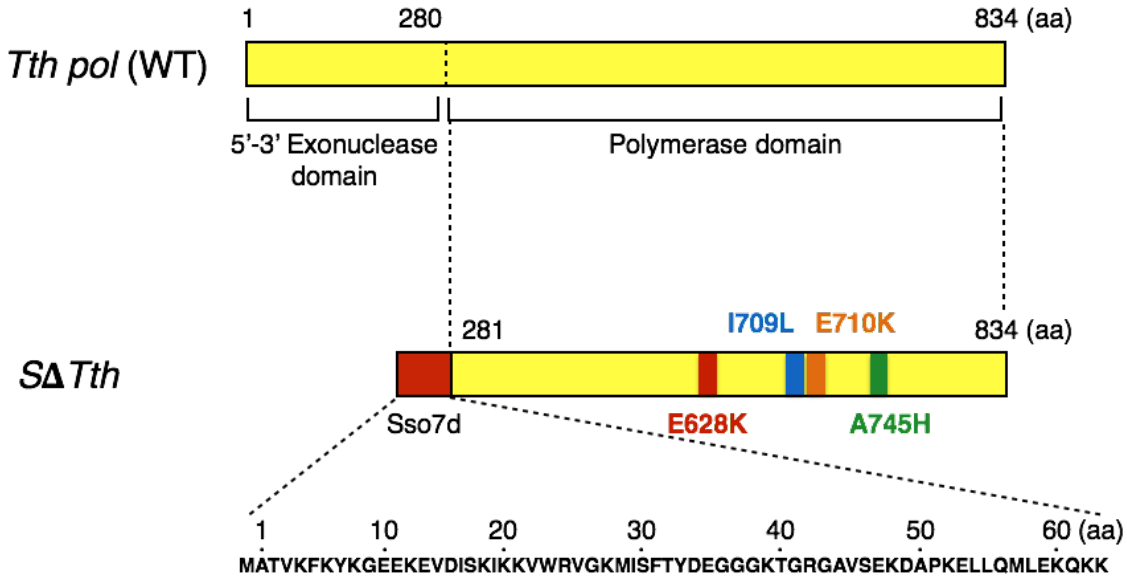


Figure 2-2. Schematic representation of the domain organization of the wild-type (WT) and a mutant SΔ*Tth* pol. In the mutant SΔ*Tth* pol, the N-terminal 280 amino acids of the 2.5-kb long WT *Tth* pol were deleted and the resultant truncated Δ*Tth* pol lacking 5′ to 3′ exonuclease activity was fused to the DNA-binding protein Sso7d from *Sulfolobus solfataricus* (Wang et al. 2004). The full-length amino acid sequence of Sso7d is shown in red box. Four point mutations were also introduced into the Δ*Tth* pol.

Next, I introduced four mutations that are previously reported to improve *Taq* pol into the S*ΔTth* pol at their corresponding locations, *i.e.*, Cs_1, E626K (E628K for *Taq* pol) and Cs_2, I707L (I709L) for cold-sensitivity and applicability in hot start PCR (Kermekchiev et al. 2003); Rs, E708K (E710K) for resistance to PCR inhibitors (Kermekchiev et al. 2009) and Ex, A743H (A745H) for improved elongation activity (Yamagami et al. 2014) (**Figures 2-3 and 2-4**), resulting in S*ΔTth*Cs_{12}RsEx pol. Finally, to reduce GC content and improve its expression levels, I performed the codon optimization of the whole S*ΔTth*Cs_{12}RsEx pol gene sequence (**Figure 2-5**), changing the triplet nucleotides of each amino acid to correspond to *E. coli* codon usage. The resultant DNA sequence encodes the same amino acid sequence (**Figure 2-6**) of the SΔ*Tth*Cs_{12}RsEx pol with a new DNA sequence. The codon optimization reduced the GC content of the SΔ*Tth*Cs_{12}RsEx pol gene from 68% to 53%.

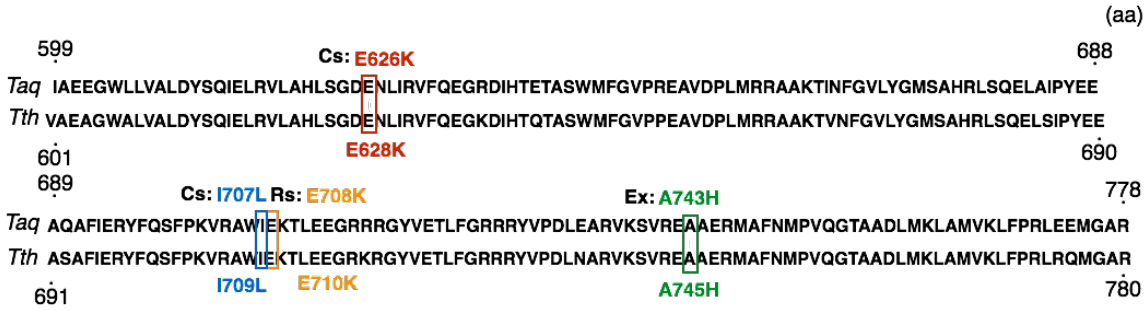


Figure 2-3. Alignment of the amino acid sequences of *Tth* and *Taq* pols. Four mutations with extension (Ex) (Yamagami et al. 2014), resistant (Rs) (Kermekchiev et al. 2009) and cold-sensitive (Cs) (Kermekchiev et al. 2003) activities are shown.

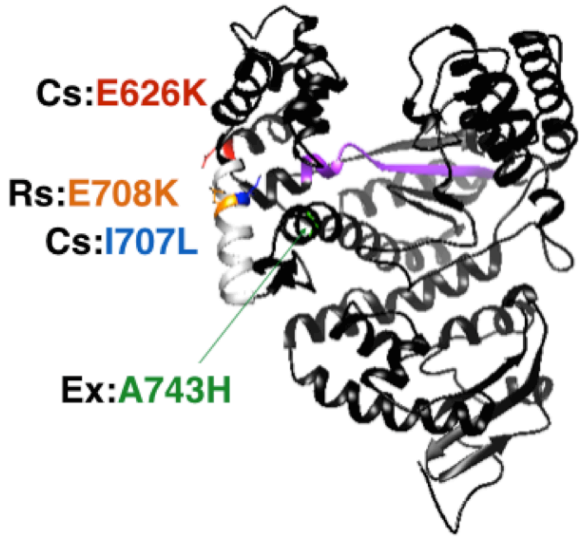


Figure 2-4. The crystal structure of the *Taq* pol (PDBid: 1ktq) (Korolev et al. 1995). It was adapted to show the location of the mutations in S*ΔTth*Cs$_{12}$RsEx pol.

ATGGCCACCGTGAAATTCAAATACAAAGGCGAAGAAAAAGAGGTGGATATCAGCAAAATCAAAAAAG
TTTGGCGTGTGGGCAAAATGATCAGCTTTACCTATGATGAAGGTGGTGGTAAAACCGGTCGTGGTGC
AGTTAGCGAAAAAGATGCACCGAAAGAACTGCTGCAGATGCTGGAAAAACAGAAAAAAGGTAGCCT
GCTGCATGAATTTGGTCTGCTGGAAGCACCGACACCGCTGGAAGAAGCCCCTTGGCCTCCGCCTGA
AGGTGCATTTGTTGGTTTTGTTCTGAGCCGTCCGGAACCGATGTGGGCAGAACTGAAAGCACTGGCA
GCATGTCGTGATGGTCGTGTTCATCGTGCAGAAGATCCGCTGGCAGGTCTGGGTGATCTGGAAGAG
GTTCGTGGCCTGCTGGCAAAAGATCTGGCAGTTCTGGCACTGCGTGAAGGTCTGGATCTGGCACCG
GGTGATGATCCGATGCTGCTGGCATATCTGCTGGATCCGAGCAATACCACACCGGAAGGTGTTGCAC
GTCGTTATGGTGGTGAATGGACCGAAGATGCAGCACATCGTGCACTGCTGAGCGAACGTCTGCATCG
TAATCTGCTGAAACGTCTGGAAGGTGAAGAGAAACTGCTGTGGCTGTATCATGAAGTTGAAAAACCG
CTGAGCCGTGTTCTGGCCCATATGGAAGCAACCGGTGTTCGTCTGGATGTTGCCTATCTGCAGGCAC
TGAGCCTGGAACTGGCAGAAGAAATTCGTCGTCTGGAAGAAGAAGTTTTTCGTCTGGCAGGTCATCC
GTTTAATCTGAATAGCCGTGATCAGCTGGAACGTGTTCTGTTTGATGAACTGCGTCTGCCTGCACTGG
GTAAAACCCAGAAAACCGGTAAACGTAGCACCAGCGCAGCCGTTCTGGAAGCCCTGCGTGAAGCCC
ATCCGATTGTTGAAAAAATCCTGCAGCATCGCGAACTGACCAAACTGAAAAACACCTATGTGGATCCG
CTGCCGAGCCTGGTTCATCCGCGTACCGGTCGCCTGCATACCCGCTTTAATCAGACCGCCACCGCAA
CCGGCCGTCTGAGCAGCAGCGATCCGAACCTGCAGAATATTCCGGTTCGCACCCCGCTGGGTCAGC
GTATCCGTCGCGCCTTTGTTGCCGAAGCGGGTTGGGCGCTGGTTGCCCTGGATTATAGCCAGATTGA
ACTGCGCGTGCTGGCGCATCTGAGCGGTGATAAAAACCTGATTCGTGTTTTTCAGGAAGGCAAAGAT
ATCCATACCCAGACCGCGAGCTGGATGTTTGGTGTGCCGCCGGAAGCGGTTGATCCGCTGATGCGTC
GCGCCGCAAAAACCGTGAATTTTGGTGTTCTGTATGGCATGAGCGCGCATCGTCTGAGCCAGGAACT
GGCCATTCCGTATGAAGAAGCGGTGGCCTTTATCGAACGCTATTTTCAGAGCTTTCCGAAAGTTCGTG
CGTGGTTAAAAAAAACCCTGGAAGAAGGTCGTAAACGTGGCTATGTGGAAACCCTGTTTGGCCGTCG
CCGTTATGTTCCGGATCTGAACGCGCGTGTGAAAAGCGTTCGCGAACACGCCGAACGTATGGCCTTT
AATATGCCGGTGCAGGGTACCGCGGCCGATCTGATGAAACTGGCGATGGTTAAACTGTTTCCGCGTC
TGCGTGAAATGGGTGCACGTATGCTGCTGCAGGTGCATGATGAACTGCTGCTGGAAGCGCCGCAGG
CCCGTGCAGAAGAAGTTGCCGCACTGGCGAAAGAAGCCATGGAAAAAGCCTATCCGCTGGCTGTCC
CGCTGGAAGTGGAAGTTGGCATCGGTGAAGATTGGCTGAGTGCTAAAGGT

Figure 2-5. The nucleotide sequence of the SΔ*Tth*Cs$_{12}$RsEx pol after *E. coli* codon optimization. The portion of Sso7d is shown in red.

MATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTG
RGAVSEKDAPKELLQMLEKQKKGSLLHEFGLLEAPTPLEEAPW
PPPEGAFVGFVLSRPEPMWAELKALAACRDGRVHRAEDPLAG
LGDLEEVRGLLAKDLAVLALREGLDLAPGDDPMLLAYLLDPSN
TTPEGVARRYGGEWTEDAAHRALLSERLHRNLLKRLEGEEKLL
WLYHEVEKPLSRVLAHMEATGVRLDVAYLQALSLELAEEIRRL
EEEVFRLAGHPFNLNSRDQLERVLFDELRLPALGKTQKTGKRS
TSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPR
TGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFV
AEAGWALVALDYSQIELRVLAHLSGDKNLIRVFQEGKDIHTQT
ASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAI
PYEEAVAFIERYFQSFPKVRAWLKKTLEEGRKRGYVETLFGRR
RYVPDLNARVKSVREHAERMAFNMPVQGTAADLMKLAMVKLF
PRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAY
PLAVPLEVEVGIGEDWLSAKG

Figure 2-6. The amino acid sequence of the SΔ*Tth*Cs$_{12}$RsEx pol. Each protein sequence is flanked by the desired tag and stop codon. The portion of Sso7d is shown in red.

2.2.2 Characterization of the modified DNA polymerase

The purpose of DNA polymerase engineering in this chapter is to obtain a polymerase that can be applied to directed evolution technique such as CSR and recently-developed CPR. In these selection methods, DNA polymerase plays a critical role in multiplying the target gene with desired properties or functions. Therefore, it is important for the polymerase to perform well in emulsion PCR with *in vitro* compartmentalization of the polymerase-expressed *E. coli* cells. In CSR, desired polymerase can be selected based on its self-replication activity. Therefore, I cloned and expressed all resulted DNA pol in *E. coli* and determined the self-replication activity in emulsion PCR (**Figure 2-7**). In comparison with the wild-type *Tth* pol, all SΔ*Tth* pol mutants showed higher self-replicating activity. Especially, SΔ*Tth*Cs_{12}RsEx pol with all four mutations showed slightly higher amplification efficiency and thus this mutant was used for further experiments.

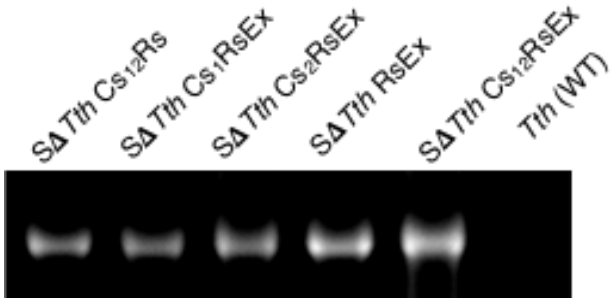


Figure 2-7. *E. coli* cells expressing the series of modified SΔ*Tth* or WT *Tth* pol were subjected to emulsified PCR for self-amplification.

2.2.3 Intermediate length amplification comparison

Next, I tested the WT and modified polymerase activity by the conventional PCR targeting genes with different lengths varying from 0.5 to 7.5 kb. With lack of WT activity prior to modification, the modified *Tth* showed significant improvement in the conventional PCR (**Figure 2-8**). Unfortunately, amplification was not observed for target lengths longer than 7.5 kb for the mutant DNA polymerase (data not shown). Therefore, it can be assumed that the extension efficiency of the SΔ*Tth*Cs_{12}RsEx pol is limited to a 10-kb target DNA length.

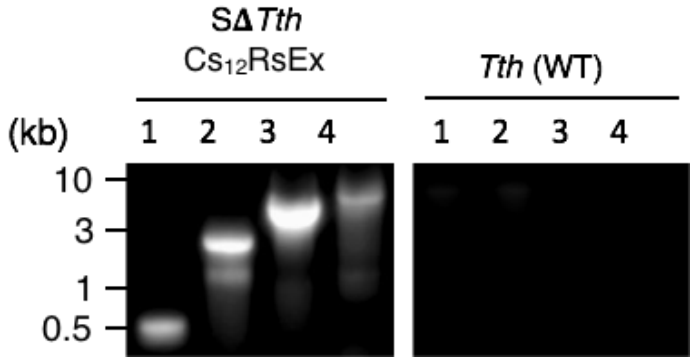


Figure 2-8. Effect of template DNA length on PCR amplification. Template DNA (lane 1, 0.5 kb; lane 2, 2 kb; lane 3, 4.65 kb; lane 4, 7.5 kb) was amplified by 30 cycles of PCR with the mutant SΔ*Tth*Cs_{12}RsEx pol or WT *Tth* pol (elongation time, 1 min/kb).

2.2.4 Fidelity comparison of DNA polymerases

The fidelity of the polymerases was tested using the lacZ-α gene (Cline et al. 1996). Despite lower sensitivity of the white/blue screening to provide fidelity results, the commercial enzymes *ExTaq* DNA polymerase with low fidelity and Phusion DNA polymerase with high fidelity were also tested along with SΔ*Tth*Cs$_{12}$RsEx in order to provide general knowledge between our polymerase and commercial enzymes. Consequently, the SΔ*Tth*Cs$_{12}$RsEx pol showed similar fidelity to that of *ExTaq* pol (**Table 2-1**), suggesting its mutant counterpart is more fitted to apply in PCR with less error-rate requirements.

Table 2-1. Comparison of the fidelity of the modified *Tth*, *ExTaq* and Phusion DNA polymerases

Polymerase	Template doublings[b]	Mutation frequency[a]	Error rate[c] ($\times 10^{-6}$) ± SE
SΔ*Tth*Cs$_{12}$RsE	2.2	0.011	10.56 ± 1.63
ExTaq	3.8	0.029	9.78 ± 1.48
Phusion	2.4	0.0014	0.24 ± 0.26

[a] Mutation frequency is expressed as the percentage of mutant (white) colonies in relation to the total number of colonies.

[b] Template doublings are calculated using the equation 2^d = (amount of PCR product)/(amount of starting target).

[c] Error rate is determined using the equation ER = mf/(bp × *d*), where mf is the mutation frequency. Number of PCRs = 3. Target (ng) = 24. SE, standard error.

2.2.5 Directed evolution of the modified DNA polymerase by CSR

To obtain DNA pol mutants with better self-amplification efficiency in emulsion, I applied CSR to further evolve the modified S*ΔTth*Cs$_{12}$RsEx pol enzyme described above. The detailed steps of CSR were explained in Materials and Methods. Briefly, as the first step of each CSR, a randomly-mutated S*ΔTth*Cs$_{12}$RsEx pol library was generated by error-prone PCR as described in materials and methods. After transformation of the library plasmid into *E. coli*, cultured cells were mixed with aqueous phase (PCR reaction mixture) and subjected to emulsion PCR with an abbreviated number of PCR cycles CSR. The resulted amplified product was retrieved as described in the methods for the further use in the next round. Therefore, the retrieving the template is the final step of each round. In this process, faster replicating mutants will perform better with the fast cycling conditions, and superior mutants are expected to perform robustly. The amplified genes were further selected for successive rounds of CSR selection using progressively lower number of PCR cycles ranging from 15 cycles in round 1 and descended to 10 cycles in round 4. After 4 rounds of CSR, the library was dominated by clones that showed amplification activity with 10 PCR cycles (**Figure 2-9**). It should be noted that, all of the individual lanes in each round represent a single library as a whole and altogether came from separate cultures for a set of single library. Therefore, the amplicon from each round was selected, pooled and subjected to further sequencing.

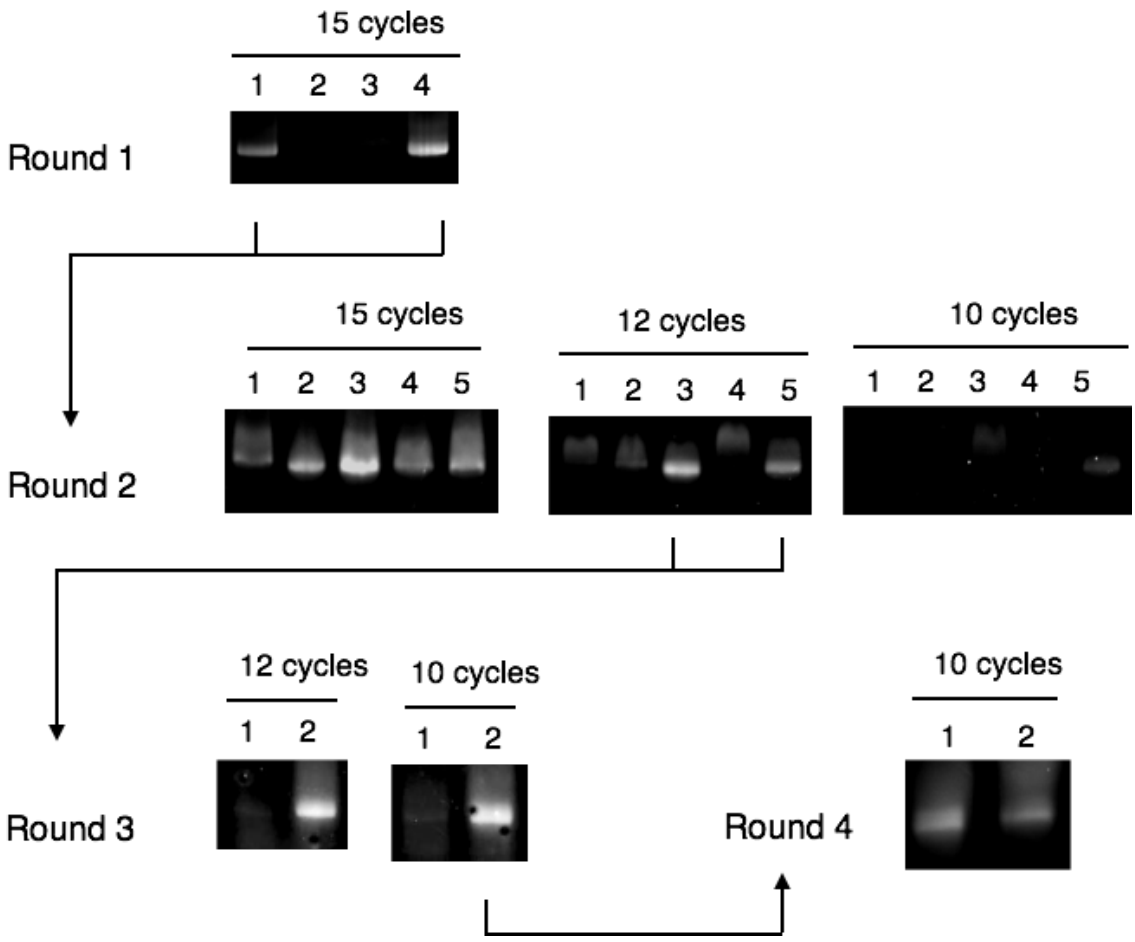


Figure 2-9. CSR selection of randomly-mutated SΔ*Tth*Cs$_{12}$RsEx pol libraries. In each round, several (2 to 5) fractionated *E. coli* libraries were subjected to 10, 12 and 15 cycles of emulsion-PCR. Amplified libraries with the lowest PCR cycle were chosen, mixed and used as the template for the next round of selection.

2.2.6 Further characterization of mutant polymerase

From the sequencing analysis of the selected SΔ*Tth*Cs$_{12}$RsEx pol mutants randomly picked from the 4th round of the CSR selection library, I identified four mutations, K7T and S18T on the Sso7d domain and E338K and E690D on the Δ*Tth* pol domain (**Figure 2-10**), suggesting that one or more of these mutations could contribute to its fast cycling activity. Then I constructed mutants with individual mutations by site-directed mutagenesis and tested the processivity of each mutant in emulsion PCR (**Figure 2-11**). Cell amounts used in both emulsified and non-emulsified PCR were set to be even by adjusting the optical density. The expression level of each polymerase from individual mutant was in turn confirmed by Western blotting (**Figure 2-12**). Consequently, both E338K and E690D amplified the target DNA of 4.65 kb with 10 sec/kb amplification time in both emulsified and non-emulsified PCR for 20 and 30 cycles, respectively, indicating that these mutations contribute the faster cycling activity (**Figure 2-11**).

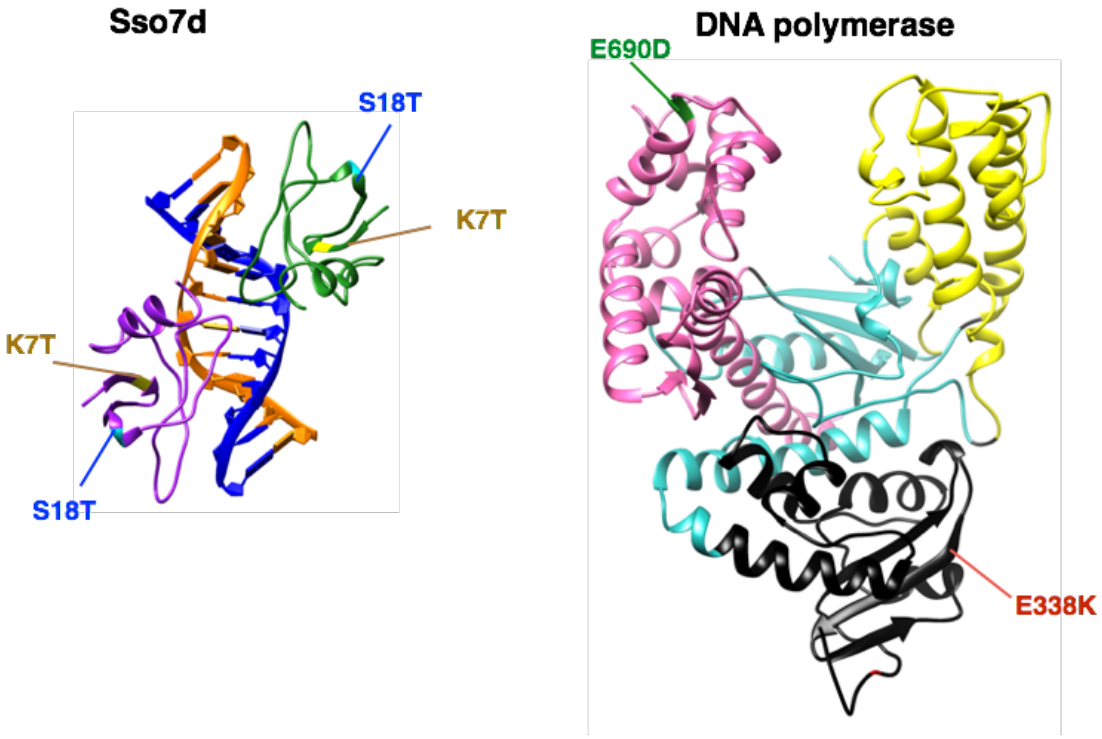


Figure 2-10. Four mutations of SΔ*Tth*Cs$_{12}$RsEx pol obtained by directed evolution. (Left) Locations of the mutations K7T (yellow) and S18T (blue) on the tertiary structure of the homodimeric Sso7d domain complexed with non-specific DNA (PDBid: 1bbx) (Agback et al. 1998). (Right) Predicted locations of the mutations E338K (red) and E690D (green) on the crystal structure of *Taq* pol (PDBid: 1ktq) (Korolev et al. 1995). Finger domain (pink); thumb domain (yellow); palm domain (cyan); and 3′ to 5′ exonuclease domain (black).

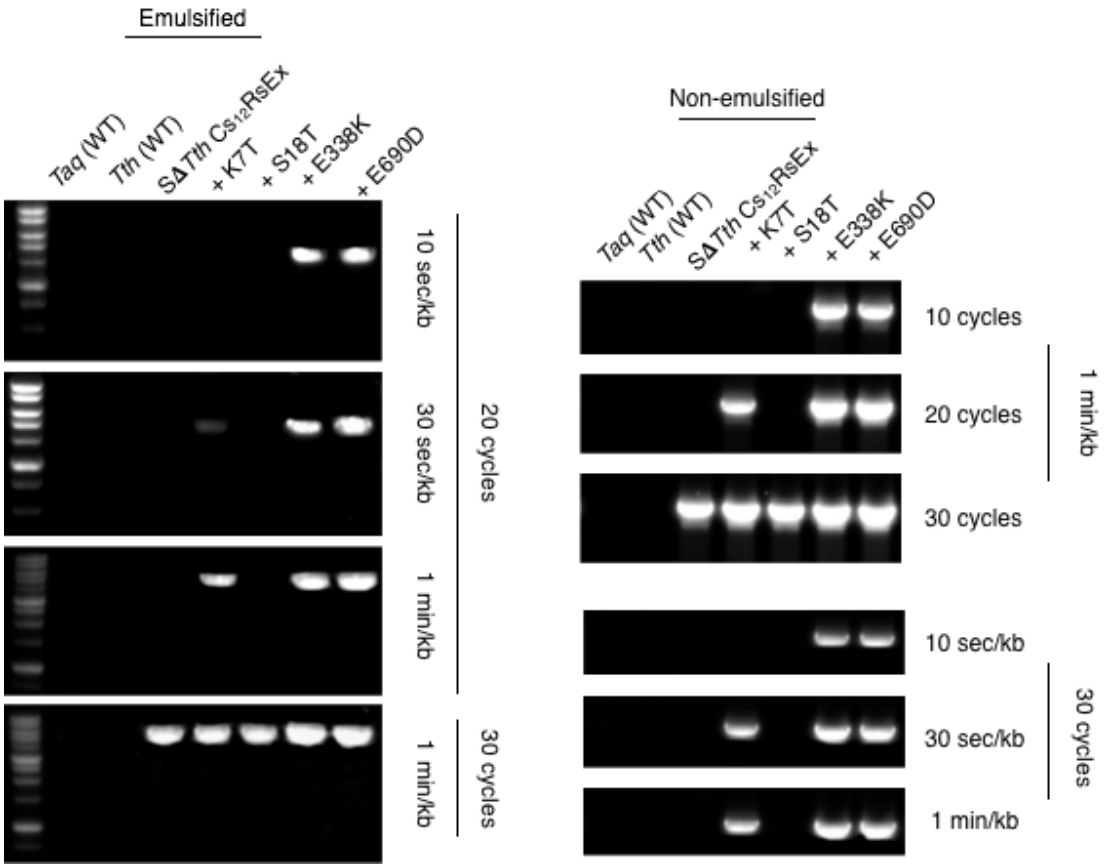


Figure 2-11. *E. coli* cells expressing the WT and modified SΔ*Tth* pols were subjected to emulsified or non-emulsified PCR and analyzed by agarose-gel electrophoresis.

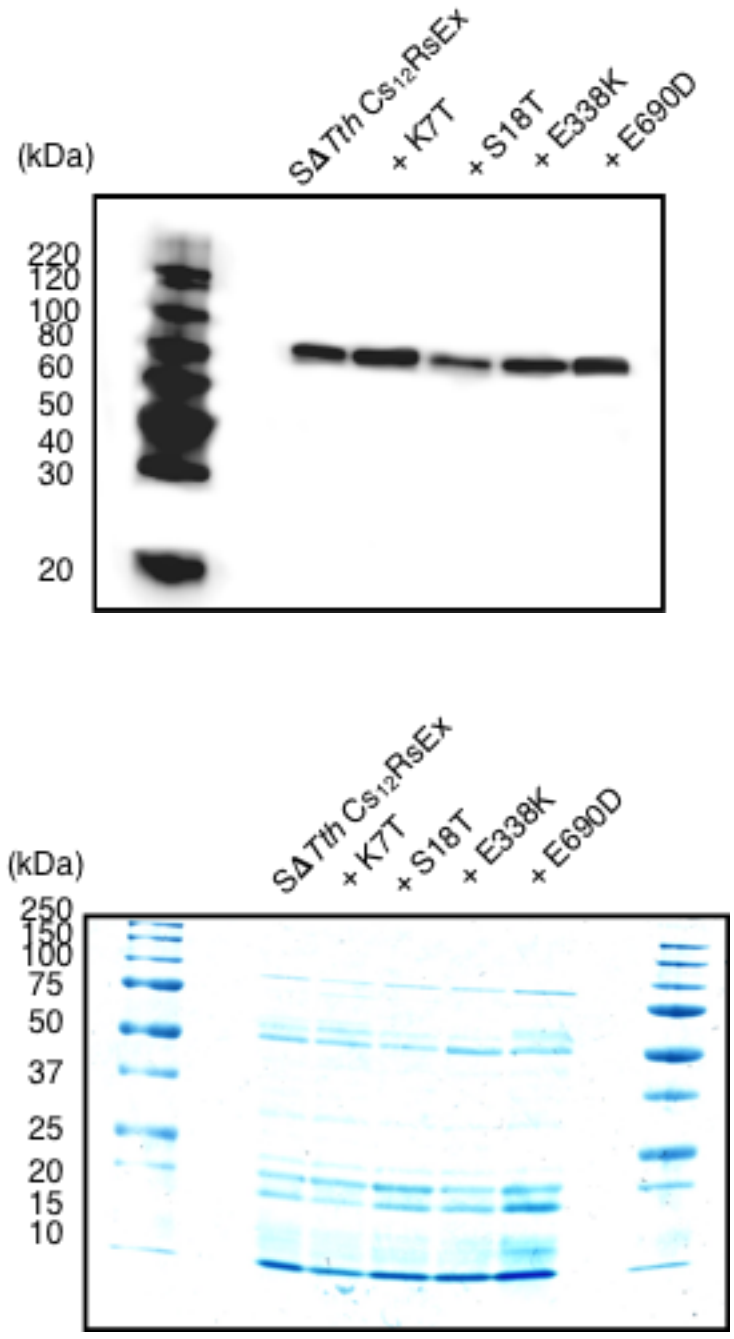


Figure 2-12. Determination of protein expression level. Western blot (top) and CBB staining (control; bottom).

2.3 Discussion

As shown in **Figure 2-10**, the position of Glu338 falls in the 3′ to 5′ exonuclease domain (residues 294-422) and Glu690 is in the finger domain in close proximity to residues that interact with DNA, such as Tyr671 (Li et al. 1998). However, the relevance of the observed genotype changes to the functional improvement of the SΔ*Tth*Cs$_{12}$RsEx pol mutants remains unclear. In addition, amplification with 1 min/kb time at 20 cycles in non-emulsified PCR was observed for not only E338K and E690D but also K7T (**Figure 2-11**), supporting that the improved performance of the mutant K7T. The effect of the K7T mutation on DNA-binding efficiency is unclear. Although a previous study on mutations in Sso7d reported that the DNA-binding capacity of several mutants was lower than the wild-type Sso7d (Consonni et al. 1999) the introduced mutations were located in the hydrophobic core, which plays a crucial role in protein stability. On the other hand, Lys7 is located in the β-sheet (**Figure 2-10**). The formation of a β-bulge involving residues Phe6 and Lys7 on strand I and Lys13 on strand II emphasizes the twist of the strands. Tyr8, the residue next to Lys7, is one of five important amino acid residues (Tyr8, Trp24, Val26, Met29 and Arg43) whose side-chains are in contact with the DNA minor groove surface and help broaden the minor groove (Su et al. 2000). The side-chains of these residues have little variability, while the side-chains that have little contact with DNA are quite variable. The role of Lys7 in Sso7d suggests that the K7T mutation would not have a major impact on DNA-binding activity. Further exploration of these mutations in relation to their improved functionality is required.

In summary, I developed modified *Tth* pol mutants that could amplify its own gene when compartmentalized in a water-in-oil emulsion. In comparison with the conventional *Taq* pol, the mutant *Tth* pol selected by CSR showed higher product yields without decreased fidelity.

In the next chapter 3, this modified *Tth* pol was applied to CPR for directed evolution of a transcriptional regulator.

2.4 Materials and Methods

2.4.1 Bacterial strains and culture condition

Escherichia coli strain XL10-Gold (Agilent Technologies, California, USA) was used for DNA cloning, and *E. coli* strain BL21(DE3)codon-plus RIPL (Stratagene, California, USA) was used for protein expression. Bacterial cultures were grown at 37°C in 2×YT or LB media containing kanamycin (25 μg/ml) (Wako, Japan), ampicillin (100 μg/ml), or both according to the required protocol. IPTG (Sigma Aldrich Japan, Tokyo, Japan) was added to the culture at the required concentration when necessary.

2.4.2. Plasmids, DNA manipulation and transformation

DNA polymerase I (*Tth* pol) gene was amplified from the genome of *Thermus thermophilus* HB27 strain (provided by Professor Toshihiro Ohta at Tokyo University of Pharmacy and Life Sciences). The oligonucleotide sequences of primers used in this study are listed in **Table 2-2**.

I designed a modified *Tth* gene S*ΔTth*, in which a thermostable DNA binding protein Sso7d was fused to the N-terminal truncated fragment of *Tth* pol. Then, after introduction of cold-sensitive mutations E628K (designated as Cs_1) and I790L (Cs_2) and extension mutation A745H (Ex), codon usage of the gene was optimized for *E. coli* codon by ordering synthetic gene to ThermoFisher Scientific (**Figure 2-5**). Furthermore, a resistant mutation E710K (Rs) was into SΔTthCs_{12}Ex as follows. pET29-S*ΔTth*Cs_{12}RsEx was constructed by plasmid PCR using a primers set Sso7dTthE710K-F and -R (**Table 2-2**) and pET29-S*ΔTth*Cs_{12}Ex as DNA template. The PCR product was treated with DpnI at 37°C for 1 hr, and then transformed into XL10-Gold by heat shock. In the same way, each S*ΔTth*Cs_{12}Rs, S*ΔTth*Cs_1RsEx and S*ΔTth*Cs_2RsEx gene was constructed by removal of point mutation from S*ΔTth*Cs_{12}RsEx gene using the primer sets of plasmid PCR, STthCsRs-F and -R, STthExRs2-F and -R, STthExRs1-F and -R, (**Table 2-2**) respectively. Then, *SΔTth*RsEx gene was constructed by restoring Cs_2 point mutation from S*ΔTth*Cs_2RsEx gene using primers STthExRs2-F and -R (**Table 2-2**).

Mutant *Tth* genes were generated by using the PrimeSTAR Max DNA polymerase (Takara, Japan) using 50 pg template DNA. The nucleotide changes corresponding to the amino acid are A19C for K7T, G53C for S18T, G364A for E338K and A1422T for E690D, receptively. Their respective primers are listed in **Table 2-2**. PCR reaction was performed 30 cycles of 98°C for 10 s, 60 °C for 15 s and 72°C for 5s/kb followed by treating with DpnI. After transformation, mutation points were confirmed by sequencing.

For cloning purposes, I applied the Gibson assembly protocol (Gibson et al. 2009) using primers with overlapped regions. In detail, for each ligation reaction, 50 ng of each PCR amplified insert and vector DNA (2 μl) were mixed with an equal volume of 2× Gibson Assembly (GA) mixture. The ligation reaction was kept at 50°C for 15 min and maintained on ice before transformation. Transformation into XL10-Gold cells was performed using a heat shock protocol at 42°C for 30 to 45 sec. After shaking for 1 h at 37°C in SOC, cells were plated on agar plates containing the appropriate antibiotic(s) and were incubated overnight at 37°C.

Sequencing of the plasmids was confirmed by using BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Framingham, MA) and the primers T7promoter-F, OptFwdforSeq, OptRevforSeq and T7terminator-R (**Table 2-2**) with ABI PRISM 3130xl genetic analyzer (Applied Biosystems).

2.4.3 Selection by CSR

CSR was performed essentially as described (Ghadessy et al. 2001). Between CSR rounds, collected genes were re-diversified by staggered extension process (StEP) PCR (Zhao 2004; Zhao et al. 1998). The detail of the CSR procedure was as follows.

The SΔ*Tth* pol gene (**Table 2-2**) was used to generate a mutant library by error-prone PCR in a 100-μl reaction containing 100 ng of template plasmid DNA, 10× *ExTaq* PCR buffer, 4 mM $MgCl_2$, 2 mM dNTPs, 2 μM of each primer, OptFwd and OptRev (Table 2), 0.5 mM $MnCl_2$, and 10 Units of *ExTaq* pol (Takara-Bio, Shiga, Japan). pET29 vector (Novagen, WI, USA) was amplified using the pET29Up and pET29Down primers (**Table 2-2**). An equal molar ratio of PCR product and vector were ligated, transformed and incubated on medium containing kanamycin at 37°C

overnight. The library was collected and shaken in LB broth overnight. On the following day, the pre-culture was diluted to an OD_{600} of 0.1 with 5 mL of fresh 2×YT medium broth containing kanamycin and was then shaken at 37°C. When the OD_{600} reached 0.4, the cells were induced with 0.1 mM IPTG and continued shaking until the OD_{600} reached 1.25. Then, 1 mL of induced culture was harvested by centrifuging at 3000 rpm for 15 min followed by washing with 1/5 volume of 1× PCR buffer [50 mM KCl, 10 mM Tris-HCl (pH 9.0), 1% TritonX100, and 2.5 mM $MgCl_2$]. The cells were slowly resuspended in 200 μL of aqueous CSR reaction mixture containing 1 μM forward and reverse primers, 0.25 mM dNTPs and 1× *ExTaq* PCR buffer. Oil phase CSR was prepared by mixing mineral oil with 4.5% (v/v) Span 80, 0.5% (v/v) Tween 20 and 0.05% (v/v) Triton X100. The aqueous CSR mixture was added dropwise (~20 μl per drop, one drop every 30 sec) to 800 μl of oil-phase in a 2 mL round-bottom Biofreeze vial (Corning, Sigma-Aldrich) under constant stirring at 1000 rpm with a magnetic stir bar at 4°C. After the addition of the last drop, stirring was continued for 5 min. The emulsion mixture was transferred to thin-walled PCR tubes (100 μl/tube) on ice and subjected to thermal cycling.

To select mutants with improved activity, emulsion PCR was performed with the least possible number of cycles, from 10 to 15, with a profile of 94°C (30 sec), 55°C (30 sec), and 72°C (3 min), preceded by an initial 5 min incubation at 94°C to lyse bacterial cells and destroy background enzymatic activity. After thermocycling, the CSR reaction mixture was recovered as follows. First, the reaction was collected in a 1.7 mL microcentrifuge tube from the PCR tubes and was centrifuged at 15,000 rpm for 10 min. The upper oil layer was removed, and the remaining phase was washed with 1 mL of mineral oil. Then, the mixture was centrifuged for 20 min at 15,000 rpm, and 100 μL of the lower aqueous layer was collected and kept at –20°C for further experiments. Next, 20% of the collected sample was used in pull-through PCR reactions with 50 μL of PCR mix containing the *ExTaq* pol. The pull-through PCR was performed for 25 cycles with a profile of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 2 min. These rescued selectants were used for the next round of CSR.

StEP was performed as follows: 20 ng of template DNA was combined with 10 μl of 10× *ExTaq* Buffer, 10 μl of 25 mM dNTPs, 1.5 mM $MgCl_2$, and 30 pmol of each primer and was adjusted with ultrapure water to a 100 μl final volume. The PCR

profile was 80 cycles of 94°C for 30 sec (denaturation) and 55°C for 5 sec (annealing/extension). Aliquots of 10 μl were run on agarose gels. The rest of the reactions was incubated with 1 μl of DpnI at 37°C for 1 h. DpnI digested StEP product (10 μl) was combined with 10 μl of 10× PCR buffer, 10 μl of dNTPs, 6 μl of 25 mM $MgCl_2$, 2.5 μl each of 20 μM primer, and 0.5 μl of *ExTaq* pol and was adjusted to a final volume of 100 μl using ultrapure water to amplify the chimeric full length DNA products. The target sequence was amplified *via* standard PCR with a profile of 96°C for 2 min and 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 1 min/kb, followed by 72°C for 7 min.

2.4.4 Enzyme activity assay

For confirmation of DNA polymerase extension efficiency, a set of primers (**Table 2-2**) was used to amplify DNA fragments 0.5 kb (T7promoter-F and T7terminator-R), 2 kb (OptFwd and OptRev), 4.65 kb (lambda-F and lambda-4.65k-R) and 7.5 kb (lambda-F and lambda-7.5k-R) in size. PCR was performed at 94°C for 2-5 min, followed by 10-30 cycles of 94°C for 15 sec, 60°C for 15 sec and 72°C for 10-60 sec/kb. The PCR reactions were performed in a total volume of 200 μl, and the mixture contained 1× *ExTaq* PCR buffer, 2 mM dNTPs, 0.2 μM of primers, 50 ng of template DNA (lambda-DNA HindIII digest; New England Biolabs) and 0.2 × 10^8 expressed cells induced with 0.1 mM IPTG for 16 h at 30°C. The 200 μl aqueous phase (induced cells resuspended in PCR mixture) was mixed with oil phase (800 μl) in emulsion PCR. Non-emulsified PCR was performed without the emulsification step.

2.4.5 Protein quantitation

Determination of protein expression level in mutant and wild-type DNA polymerase were done as follows. Single colony of each cell was picked and inoculated in LB broth at 37°C overnight. On the following day, full-grown culture was diluted in fresh medium (1/25 volume) and grown at 37°C. Log phase induction of each cell was performed with 0.1 mM IPTG for 16 h at 30°C. After harvesting,

same amount of induced cell were used to obtain crude extract. The lysate was resuspended in 1 g/ 3 ml of lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, 0.05% Trition-X100, 1 mM PMSF (Phenylmethylsulfonyl fluoride)] and incubated at 70°C for 30 min. The lysate was collected by taking the supernatant after centrifugation at 15 000 rpm for 30 min. Same amount of lysate were used to analyze by 10% SDS-PAGE and Western blot analysis in which the membrane was incubated with primary antibody [anti-HA-tag (6E2) Mouse mAb, 1:1000] overnight at 4°C and secondary antibody (anti-mouse IgG HRP, 1:5000) for 1 h incubation at room temperature.

2.4.6 PCR fidelity assay

The fidelity of DNA replication during PCR was measured using a previously described assay (Cline et al. 1996). A 1.3-kb sequence encoding the lacZ-alpha gene was PCR amplified using primers lacF and lacR (**Table 2-2**), which overlap by 20 nucleotides, with linearized pCR2.1-TOPO vector (Thermo Fisher) amplified by Phusion DNA polymerase (Thermo Fisher) using primers TOPOUp and TOPODown (**Table 2-2**). Purified insert and vector were ligated by Gibson assembly as described above. Then, 25, 50 or 100 μl of the transformation mixture was plated on LB plates containing kanamycin, ampicillin, IPTG (0.1 mM) and X-gal (40 μg/mL).

Table 2-2 Oligonucleotides used in the chapter 2.

Name	Sequence (5′ to 3′)
A1422T-F	TATGAAGATGCGGTGGCCTTTATCGAACGC
A1422T-R	GCCACCGCATCTTCATACGGAATGGCCAGT
A19C-F	TGAAATTCCAATACAAAGGCGAAGAA
A19C-R	TTGTATTGGAATTTCACGGTGGCCAT
G364A-F	ATCGTGCAAAAGATCCGCTGGCACCTCTGG
G364A-R	CGGATCTTTTGCACGATGAACACGACCATC
G53C-F	GGATATCACCAAAATCAAAAAAGTTTGGC
G53C-R	TGATTTTGGTGATATCCACCTCTTTTTCTT
lacF	AGCGCCCAATACGCAAACCG
lacR	GAACCTGCGTGCAATCCATC
lambda-F	CCGTTCTTCTTCGTCATAA
lambda-4.65k-R	GATGCCGTTCATGACCTGTAA
lambda-7.5k-R	GCAGCACAAATGCCACAGGTTCAT
OptFwd	GGAGATATACATATGGCCACCGTGAAATTC
OptFwdforSeq	GTAGCCTGCTGCATGAA
OptRev	TGGTGCTCGAGTTAGGATCC
OptRevforSeq	ACGAACTTTCGGAAAGCT
pET29Down	GGATCCTAACTCGAGCACCA
pET29Up	CATATGTATATCTCCTTCTTAAAGTTAAAC
Sso7dTthE710K-F	GCCTGGTTAAAAAAGACCCTGGAGGA
Sso7dTthE710K-R	GTCTTTTTTAACCAGGCCCGCACCTT
STthCsRs-F	GTTCGCGAAGCGGCCGAACGTATGGC
STthCsRs-R	GTTCGGCCGCTTCGCGAACGCTTTTCAC
STthExRs1-F	GAGCGGTGATGAAAACCTGATTCGTG
STthExRs1-R	GGTTTTCATCACCGCTCAGATGCGCC
STthExRs2-F	GCGTGGATTAAAAAAACCCTGGAAGAAGG
STthExRs2-R	GTTTTTTTAATCCACGCACGAACTTTCGG
T7promoter-F	TAATACGACTCACTATAGGG
T7terminator-R	GCTAGTTATTGCTCAGCGG
TOPOUp	CGGTTTGCGTATTGGGCGCT
TOPODown	GATGGATTGCACGCAGGTTC

Chapter 3

A dual system using compartmentalized-partnered replication for selection of arsenic-responsive transcriptional regulator

3.1 Background and rationale

Arsenic contamination in groundwater is a harmful threat to million people worldwide (Erickson and Barnes 2005; Kumar et al. 2016; Nordstrom 2002; Fendorf et al. 2010; Kim et al. 2011; Rahman et al. 2009) (**Figure 3-1**). Arsenic exists under the form of various chemical species in different physicochemical behavior, toxicity, bioavailability and biotransformation. Among various arsenic species, arsenite [As(III)] and arsenate [As(V)] are the most toxic forms and regarded as human carcinogenic substances (Fendorf et al. 2010). The tolerable weekly intake for inorganic arsenic is set at 15 μg/kg of body weight, and the highest acceptable level of arsenic in drinking water is 10 μg/L (World Health 2011).

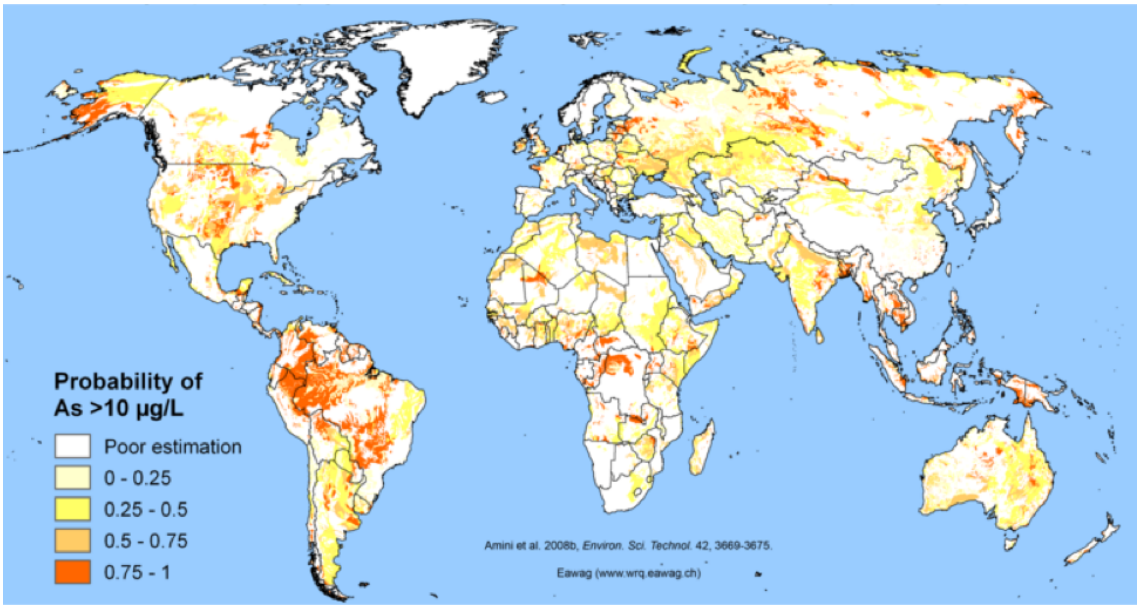


Figure 3-1. Global geogenic arsenic contamination in groundwater.
Source: https://phys.org/news/2016-04-milestone-arsenic-fluoride-contaminated.html

Unfortunately, populations in the most contaminated area have least access to the centralized drinking water system. Moreover, arsenic analysis of environmental samples requires sophisticated chemical equipment and technical expertise that leads to become detecting the level of arsenic in potable water a troublesome in these regions (**Table 3-1**). To address this problem, scientists have been keen on developing biosensors for arsenic as an alternative solution (Fernandez et al. 2016; Hu et al. 2010; Huang et al. 2015; Truffer et al. 2014; Webster et al. 2014).

Due to more simplicity and lower cost in relative to the analytical instruments, biosensors have been paid attention in fieldworks requiring *in situ* detection. However, the deployment of biosensors is limited in term of operational lifetimes and their sensitivity to other environmental factors such as temperature and pH (**Table 3-2**). These problems are the obstacle to use biosensors in place of conventional analytical instruments. Since biosensors are beneficial in measuring the contaminant concentration in places where the access to the instruments and technical experts are out of reach, solution of these problems should be contributed to environmental assessments such as surveillance of arsenic contamination in groundwater.

Biosensors using whole microbial cells to determine a variety of heavy metals using whole microbial cells have been reported (Huang et al. 2015; Liao and Ou 2005; Yu et al. 2017). Over time, bacteria have evolved mechanisms to become resistance to many toxic metals. Many of these heavy metal resistance systems have been studied to their genetic determinants such as mercury (Cai et al. 2018), zinc (Watstein and Styczynski 2018), cobalt (Peca et al. 2008), cadmium (Shetty et al. 2003), lead (Shetty et al. 2003) and arsenic (Liao and Ou 2005; Siegfried et al. 2015; Yu et al. 2017). Arsenic resistance gene in bacteria, *ars* operon, has been studied widely (Silver et al. 1993; Rosen 1995; Oremland and Stolz 2003). The main genes of arsenic resistant system (**Figure 3-2**) include the transcriptional repressor ArsR, the arsenic efflux pump ArsB, and the arsenate reductase ArsC (Xu et al. 1998) The *ars* genes have been found in chromosomal locations or plasmid-encoded in a number of Gram-negative bacteria as well as in Gram-positive. The minimal number of gene in one operon is *arsRBC* in the chromosome of *E. coli* (Carlin et al. 1995).

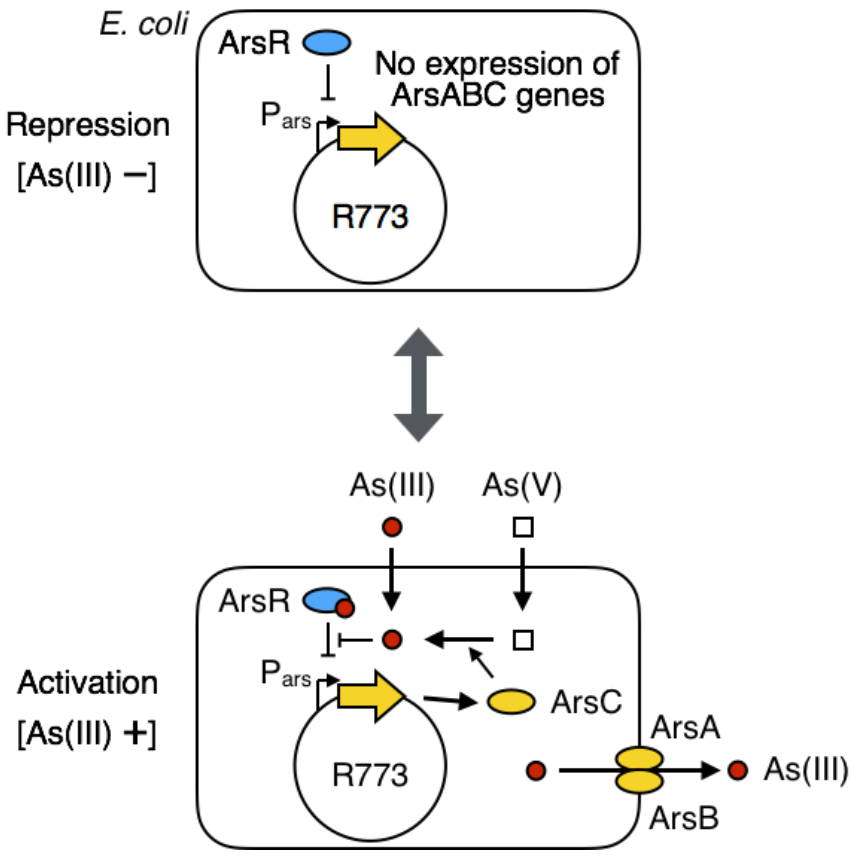


Figure 3-2. Schematic representation of Ars operon in *E. coli* plasmid R773 (Chen and Rosen 2014). Although arsenic enters *E. coli* cells as either arsenite [As(III)] or arsenate [As(V)], As(V) is transformed to As(III) by a reductase ArsC in the cells and As(III) is excluded from the cells by the ArsAB efflux pump. The expression of the Ars genes is regulated by a transcriptional repressor ArsR. In the absence of As(III), ArsR repress the transcription of Ars operon, but in the presence of As(III), As(III) binds to ArsR so that induce the expression of ArsABC.

Table 3-1 Analytical methods used for detection of arsenic in the previous studies

Analysis	Tools	Detection limit	Disadvantages	References
Spectrpscopy	UV-vis	1.0 μg/L	Sample prone to false postive and negative readings, non-specific detection	Tahir et al. (2008)
	ETAAS	0.02 μg/L	Strong interference and costly matrix modifiers Expensive analysis	Anawar (2012)
	ICP-OES/AES	0.05 μg/L	Expensive and technical expertise requirements	Chen et al. (2009)
	ICP-MS	0.4 ηg/L		Petursdottir et al. (2012)
	GC-MS	5.8 ηg/L	Difficult sample preparation	Richter et al. (2012)
Electrochemical voltammetry	ASV	0.8 ηg/L	Expensive, technical expertise requirement and copper intereference	Gao et al. (2013)
	CSV	37.5 ηg/L		Gibbon-Walsh et al. (2010)
Chemical sensors	Thiol based SPR sensor	3.0 ηg/L	Non-specific and redox interference	Forzani et al. (2007)
	Dosimetric fluorescent Probe-ArsenoFlluors	0.14 μg/L	Only on non-aqueous media	Ezeh and Harrop (2013)
	Antipyrine based schiff	225 μg/L	Poor detection and pH dependent	Loharetal (2013)

Table 3-2 Arsenic biosensors in the previous studies

Biosensor	Detection limit (μg/L)	Induction time (h)	Limitations	References
Luciferase-based				
S. cerevisiae pdr5Δ luxAB	0.0007	1	Low genetic stability	Bakhrat et al. (2011)
Photobacterium leiognathi	4000	0.5	Poor detection, non-specificity	Ranjan et al. (2012)
E. coli DH5α (pASPW2-*arsR-luxCDABE*)	0.74	2	Antimony interference, culture maintenance	Sharma et al. (2013)
E. coli arsRp: luc	3.75	2	Cell lysis requirement	Hou et al. (2014)
***LacZ* based**				
E. coli JN 109-pSB1A2-BBa-J15501 (*arsR-lacZ-xylE*)	50	24	Long response, bicarbonate interference, false positive results due to lactose	De Mora et al. (2011)
E. coli (*arsD-lacZ*)	100	0.5	Poor detection limit	Chiou et al. (2011)
E. coli DH5α (pPROBE *arsR-ABS-RBS-lacZ*)	0.8	4.23	Contamination risk of repeated use	Cortes-Salazar et al. (2013)
***gfp* based**				
E. coli (*arsR-egfp*)	10	3.33	Short storage stability	Buffi et al. (2011)
ArsRCis-gfp trans	5	0.5	Non-specificity	Siddiki et al (2011)
E. coli AW3110 with two plasmids	10	5	Long response time	Chen et al. (2012)

In this chapter, I coupled the *arsR* gene with a reporter gene whose expression was controlled by the arsenite-dependent promoter. In principle, in the presence of arsenite, the ArsR repressor binds to the metal leading to the release of the promoter and the consequent expression of a reporter gene. First, I constructed and characterized an arsenite biosensor strain by using green fluorescent protein (GFP) as a reporter. Next, I developed a novel two-step selection system based on compartmentalized partnered replication (CPR) (Abil et al. 2017; Ellefson et al. 2014) and a *sacB*-dependent sucrose sensitivity (Pelicic et al. 1996) for directed evolution of ArsR.

Two major bacterial proteins deal with most arsenic species. One is a pump which is integrated into the bacterial cell wall and removes any arsenite from the interior of the cell to the outside. The other protein is ArsR that is an arsenite sensing protein. It has tow binding capacities. In the absence of arsenite, it has affinity to DNA and represses the gene expression. However, the repression is not complete and small amounts of arsenic sensing elements are always present. In arsenite presence, ArsR binds the arsenite and loses the affinity for the DNA binding site. The dissociation constant of arsenic to ArsR is 0.03 (Ivakhno and French 2006).

For the development of arsenic biosensor, we took advantage of ArsR. Our interest is not to have the defense mechanism against it but to have the protein or enzyme, which can easily be measured. This is where genetic engineering comes into play. I altered the bacteria in a way that they produce a signal such as fluorescence. The question remains how to use a biosensor. In its simple form, liquid culture of biosensor cells are preculture in a liquid medium and mixed with aqueous sample to be measured. Such kind of preculture can be frozen and kept for future use. The volume of the assay mixture can be as low as 200 μl and many assays can be carried out simultaneously in 96 well plates. The assay can be performed within two hours. For unknown samples, different dilutions have to be made in order to perform accurate measurements.

From the initial measurements with wild type ArsR, the response time is optimal at 2 hr and the detection limit is 1 μg/L. The wild type sensor was found no response for some other metals tested in this study confirming the specificity. However, the fluorescence measurements of wild type sensor show poor sensitivity upon induction where it can be obstacles to obtain the accurate measurements of unknown sample.

Therefore, prior to engineering state that will be suitable in field study and practical use, I intend to evolve the *arsR* gene using CPR.

In other words, better sensitivity will lead to better expression of a DNA polymerase as a reporter gene, which will eventually amplify the corresponding *arsR* in each compartment. Over round while decreasing the arsenite concentration as a selection pressure, a library member with higher sensitivity are expected to enrich. However, in this CPR process, inactive *arsR* mutants are also selected in principle. Such false positive clones can be excluded by the second step of OFF selection in the absence of arsenite by replacing the DNA polymerase with *sacB* gene as a reporter. Here I confirmed the two-step selection by proof-of-principle experiments.

3.2 Results

3.2.1 Design of biosensor strain

First I constructed a whole-cell biosensor with GFP as a reporter to measure arsenic in a broth sample. This arsenite [As(III)] biosensor strain contains a double-plasmid system where the regulator *arsR* gene was placed under the control of T7 promoter on pET29 (pBR322 ori; kanamycin resistance) and the reporter *gfp* gene was placed under the control of As(III)-dependent *ars* operator/promoter (O/P) on pACYCduet1 (p15A ori; ampicillin resistance). We utilized BL21(DE3) as host organism with the double-plasmid system, pET-*arsR*/pACYC-*gfp* (**Figure 3-3**). In theory, when there is no As(III), the ArsR repressor inhibits the reporter expression. Upon the arsenic arrival into the cell *via* phosphate channel (Owolabi and Rosen 1990; Rosen 1999; Diorio et al. 1995), ArsR binds to As(III) releasing the promoter and leading downstream reporter expression (**Figure 3-3**).

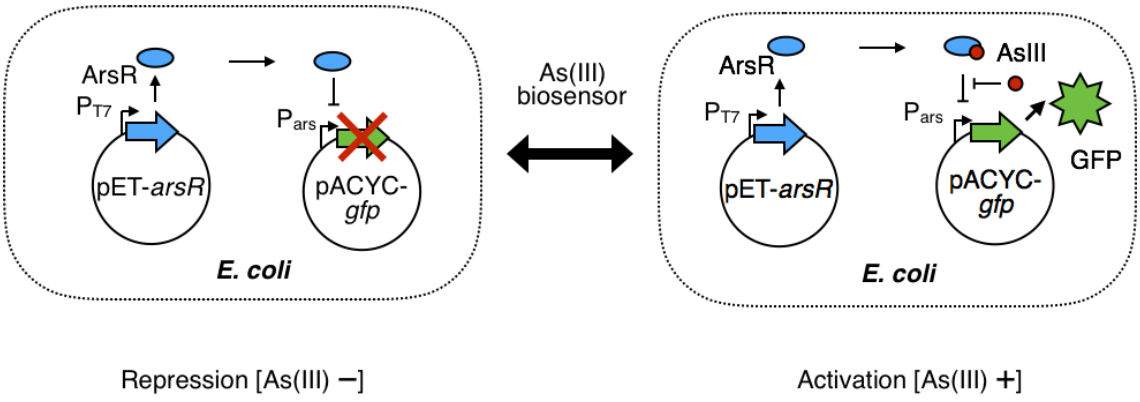
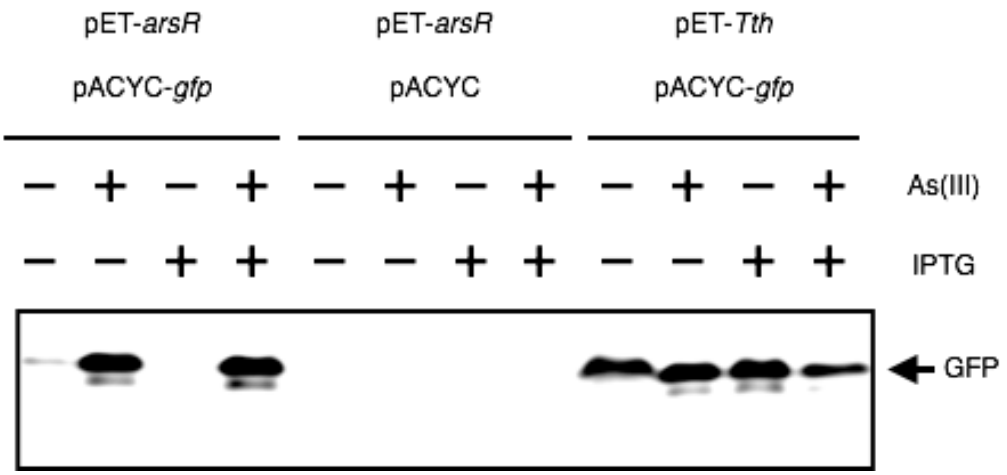


Figure 3-3. Illustration of arsenic dependent repression and activation of gene expression by ArsR. ArsR represses transcription at P_{ars} in the absence of arsenite [As(III)] (left) while activates in the presence of As(III) (right).

3.2.2 Characterization of biosensor strain

The response of the biosensor strain was confirmed in four separate induction conditions with As(III) and IPTG. In term of the above arsenic-responsive strain (pET-*arsR*/pACYC-*gfp*), only the As(III) induction gave the increase in GFP expression and fluorescence as expected (**Figure 3-4**). On the other hand, the IPTG induction showed no effect, suggesting leaky expression of ArsR on the pET vector. Thus, further experiments were performed without IPTG. As negative-control strains, ALWAYS-OFF strain (pET-*arsR*/pACYC; *gfp* was deleted from pACYC) and ALWAYS-ON strain (pET-*Tth*/pACYC-*gfp*; pET29 carry *Tth* instead of *arsR*) were also tested.

(A)



(B)

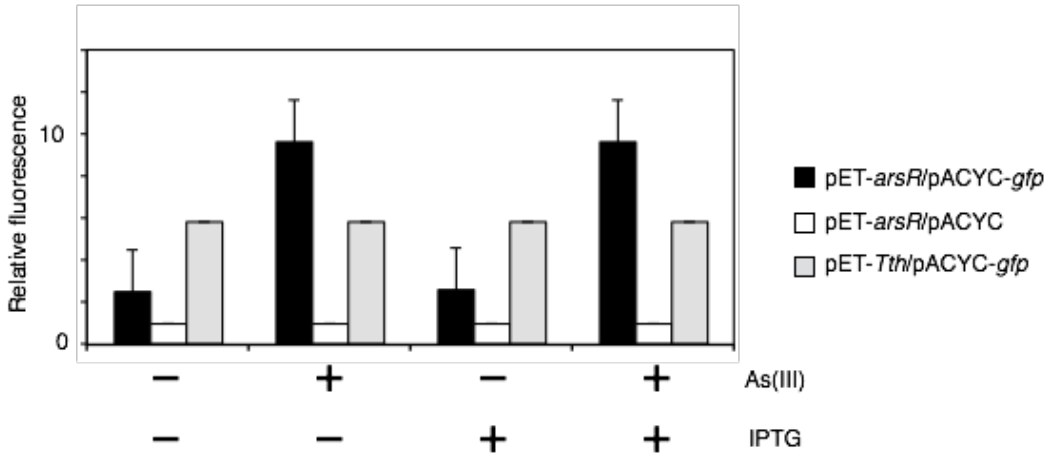


Figure 3-4. The synthetic circuit was confirmed by producing three different strains. The first strain is arsenic dependent and contains pET29-*arsR* and pACYC-*gfp*. ALWAYS-ON strain carries pET29-*Tth* and pACYC-*gfp*. ALWAYS-OFF strain consists of pET29-*arsR* and pACYC. Fluorescence of each strain was monitored after inducing with sodium arsenate and IPTG at log phase. Uninduced culture was used to normalize data. The results were confirmed by (A) fluorescence imager to visualize the GFP bands after SDS-PAGE analysis or (B) by fluorescence microplate reader.

3.2.3 Effects of various induction level on biosensor

I also measured the fluorescence of three strains after induction with different amount of As(III). Upon induction with 10 μg/L arsenite, the arsenic responsive strain reached a little over 2 fold of background expression level (**Figure 3-5**). The fold of GFP intensity reached approximately 5.5 when the arsenic level increased to 100 μg/L. There was no effect of arsenic concentration upon the GFP level with the ALWAYS-ON and ALWAYS-OFF strains. The GFP expression of the arsenic-responsive strain was also confirmed by visualizing the protein bands with fluorescence imager (**Figure 3-5**).

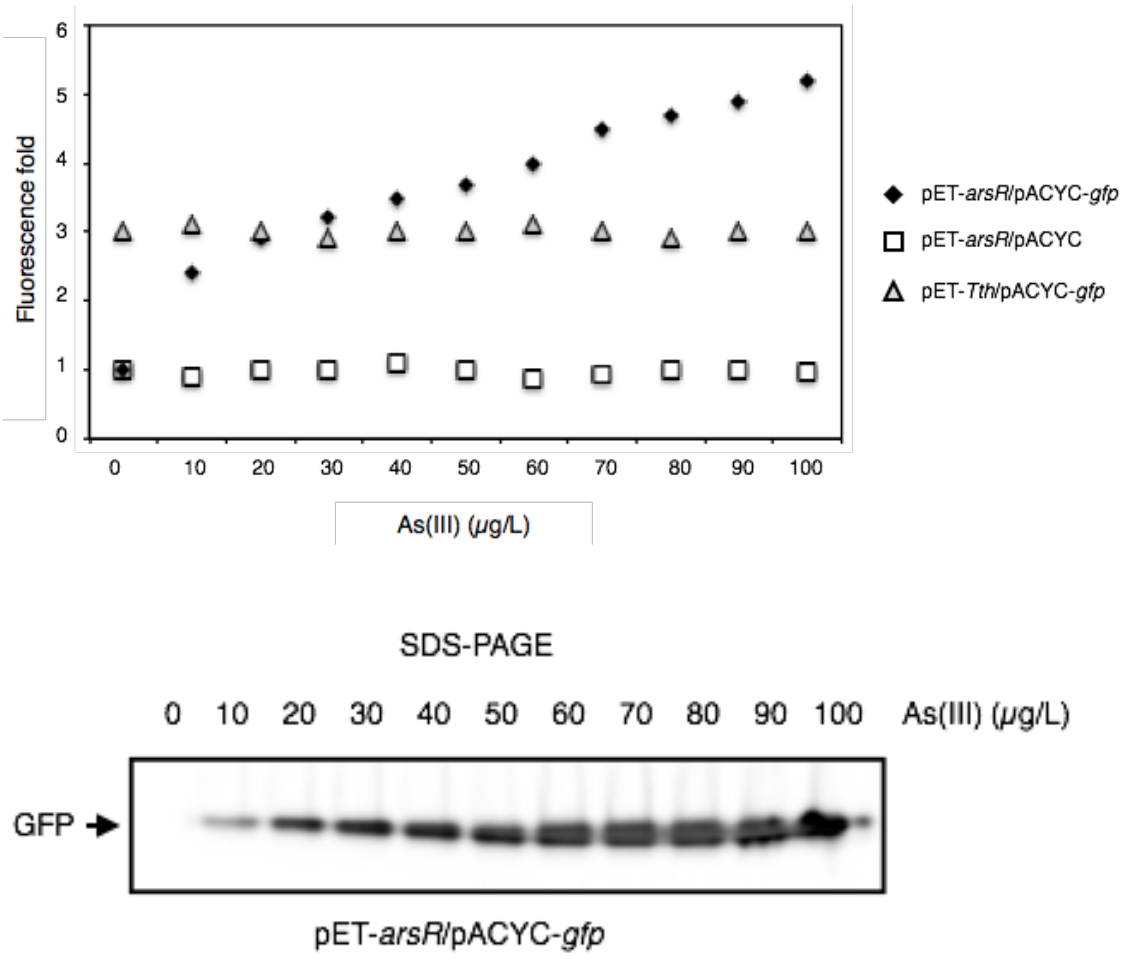


Figure 3-5. Biosensor response. The three strains were treated with serial concentrations of arsenic from 10 to 100 μg/L. (Top) Fluorescence intensity of log-phase induced culture for (2 hrs induction time). (Bottom) Cell lysate was visualized by fluorescence imager.

3.2.4 Time and dose dependent assay

To elucidate the detection limit and time dependent response, the biosensor strain with fluorescence level upon exposure to different concentrations of arsenic level ranging from 0, 5, 10, 25, 50, 75 and 100 μg/L sodium arsenate was determined 1, 2 and 4 h after the addition of sodium arsenate (**Figure 3-6**). Background fluorescence was still observable from the bacterial cell itself without arsenic presence. With the 1-h induction time, the fluorescence increased almost 2 fold while the 2-h induction reached to over 3.5 fold relative to uninduced cell at 10 μg/L. The fluorescence regression was not found to be linear with increased arsenic concentration with the 2-h induction measurements. The fluorescence was found to decrease after 4 h indicating increased toxicity to the cell after long-term exposure to arsenic.

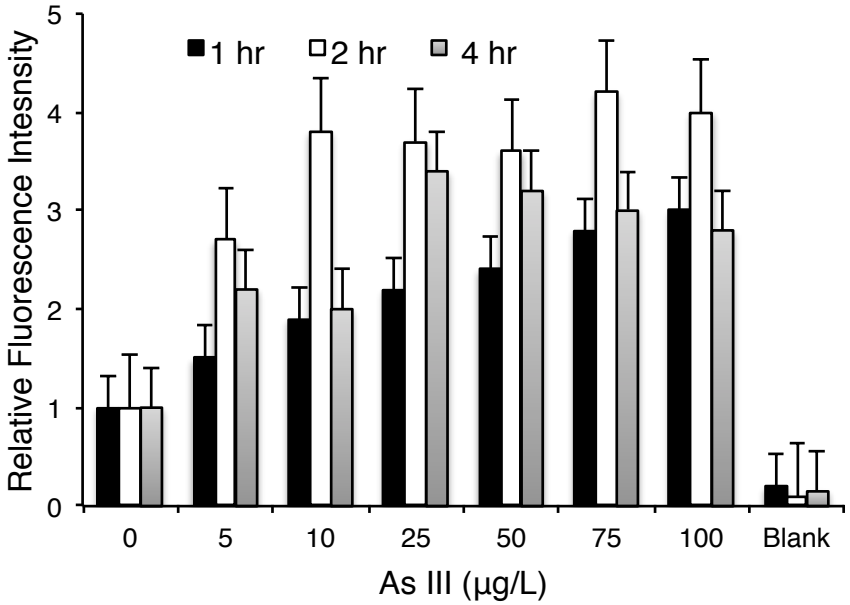


Figure 3-6. Dose and time dependent assay of biosensor strain. Single colony was picked to preculture followed by 1:25 fold dilution on the following day. Log phase induction was done with sodium arsenate 0 to 100 μg/L. Fluorescence was measured using 96 well microtiter plates at 1, 2 and 4 h respectively.

3.2.5 Lifetime performance and specificity assay

One of the concerns in applying whole cell biosensor in the practical field is the lifetime of the bacterial cell. Therefore, to monitor the lifetime of arsenic detecting strain in term of viability, log phase cultures were kept as broth culture and pellet at room temperature or 4°C. The fluorescence of each culture was determined daily for six consecutive days by inducing with 10 μg/L As(III) for 3 h (**Figure 3-7**). From all the measurements, no significant difference was found in the viability between different temperatures or between broth culture and pellet. The cell between first and second day showed similar response. The response significantly dropped on day 3 and cell became non-responsive from day 4 and onwards.

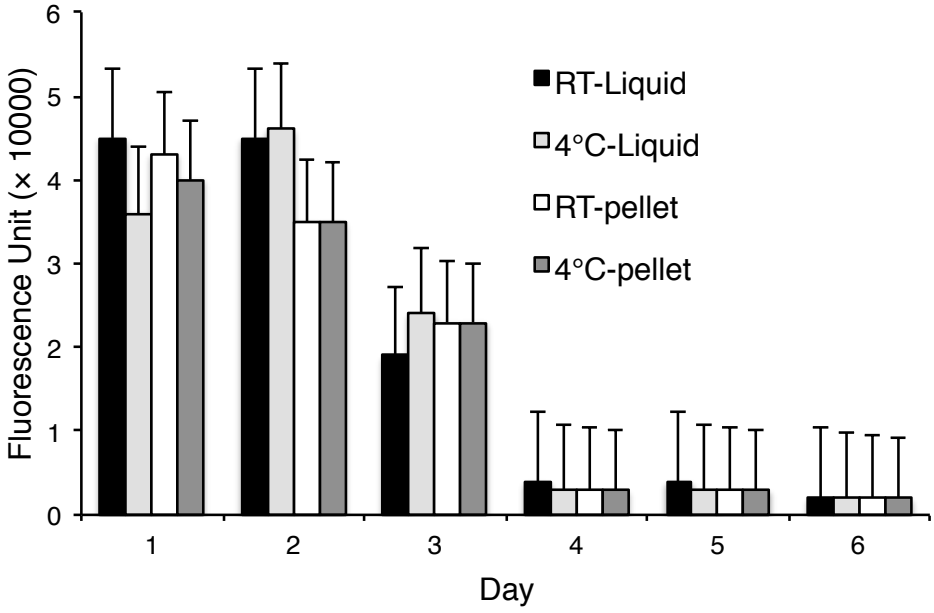


Figure 3-7. Lifetime determination of biosensor strain. Log-phase cells were kept at room temperature or 4°C in broth (liquid) or pellet state. On each day onward, 200 μL of stored cultures were incubated 96 well plates and induced with 10 μg/L arsenic. Fluorescence was measured after 3 h incubation. All incubation was performed at 37°C.

Finally, I examined the response of the cell to some of the most common metals along with arsenic. As shown in **Figure 3-8**, the biosensor showed no response to other metals while showing high affinity to arsenite. This result indicated the specificity of the biosensor.

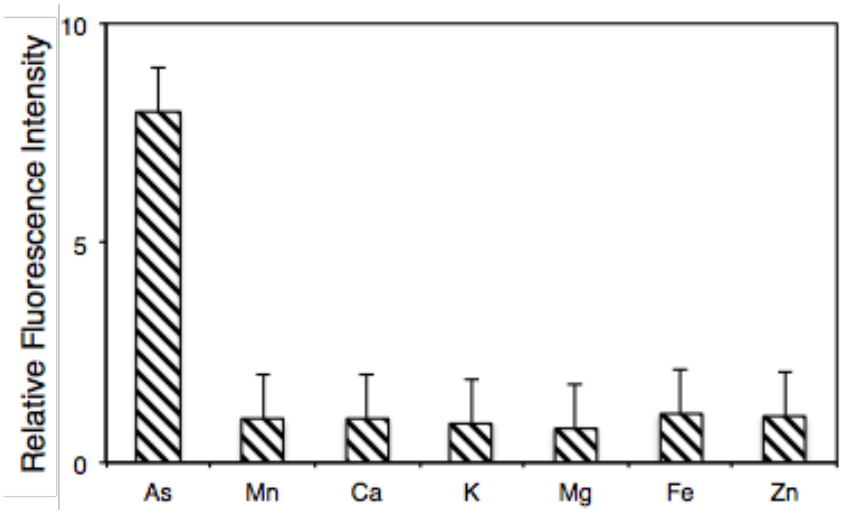


Figure 3-8. Specificity test of arsenite-responsive bacterial biosensor. Metals other than arsenic were used in 5 mg/L while arsenic was used in 50 μg/L. Fluorescence levels were measured after induction with different elements for 3 h at 37°C.

3.2.6 Strategy for 2-step selection of arsenic biosensor

Since the genetic circuit of arsenic regulatory system described above worked well with the *gfp* reporter (pET-*arsR*/pACYC-*gfp*), I moved on to the set-up directed evolution system for *arsR* by compartmentalized partnered replication (CPR) (**Figure 3-9**). I adopted CPR as a positive selection system because it is beneficial for selection under toxic conditions such as arsenite and it does not require expensive apparatus such as flow cytometer for selection (Ellefson et al. 2014). In CPR, separation of single cells by compartmentalization ensures the amplification of a partnered gene that can drive the expression of a thermostable DNA polymerase gene. We constructed a reporter plasmid pACYC-*Tth* using the modified *Thermus thermophilus* DNA polymerase developed in Chapter 2. In this ON selection system, however, not only a mutated *arsR* gene with higher sensitivity against As(III) but also inactive genes can be also selected as false positive (**Figure 3-9**).

In order to exclude false positive clones from CPR, I applied levansucrase *sacB* gene from *Bacillus subtilis* that confers sucrose sensitivity in *E. coli* (Recorbet et al. 1993; Gay et al. 1983). *B. subtilis sacB* gene encodes the secreted enzyme levansucrase (EC 2.4.1.10) that catalyzes hydrolysis of sucrose and synthesis of levans, which are high-molecular weight fructose polymers (Rapoport and Dedonder 1966). SacB expression in the presence of sucrose is lethal in gram-negative bacteria such as *E. coli* (Gay et al. 1985) although the mechanism behind this lethality is not fully understood. I constructed the pET-*arsR*/pACYC-*sacB* and pET-*gfp*/pACYC-*sacB* strains similar to the other two reporter systems. In theory, the *sacB* will not be expressed in *E. coli* containing active *arsR* gene in the absence of As(III). Similarly, *sacB* will not be expressed in cells with inactive *arsR*. Consequently, in the presence of sucrose, false positive clones with inactive gene will be removed and active *arsR* variants can be enriched (**Figure 3-10**).

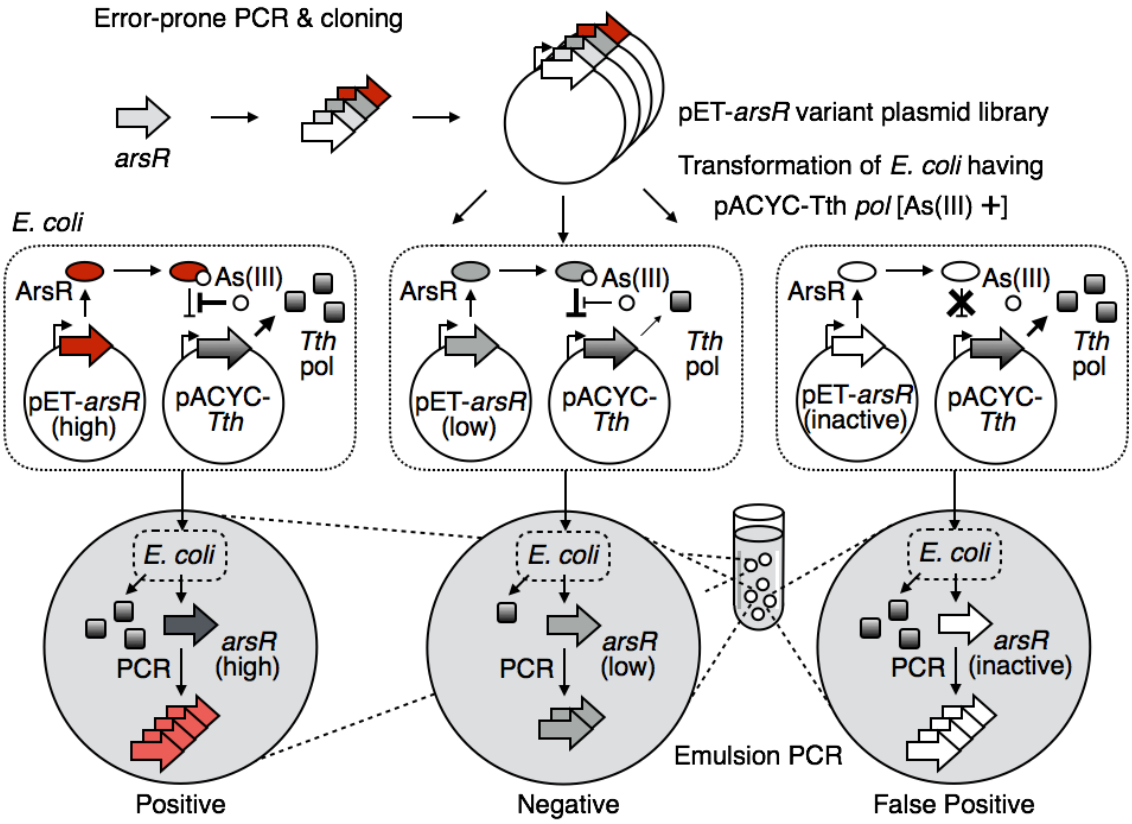


Figure 3-9. ON (CPR) selection of ArsR mutants in the presence of As(III). A pET-*arsR* variant plasmid library is created by error-prone PCR and cloning. The library is transformed into *E. coli* with pACYC-*Tth*. The DNA polymerase expression repressed by ArsR is restored by As(III). When each cell is compartmentalized by water-in-oil emulsion and subjected to emulsion PCR, heat lysis releases *Tth* pol leading to the amplification of the active variant genes. At this time, an *arsR*(high) gene with high sensitivity against As(III) is expected to be amplified more than an *arsR*(low) gene with low sensitivity. However, inactive variants (with stop codon or frameshift mutations) also resulted in *Tth* pol production in *E. coli* and are designated as false positive (FP).

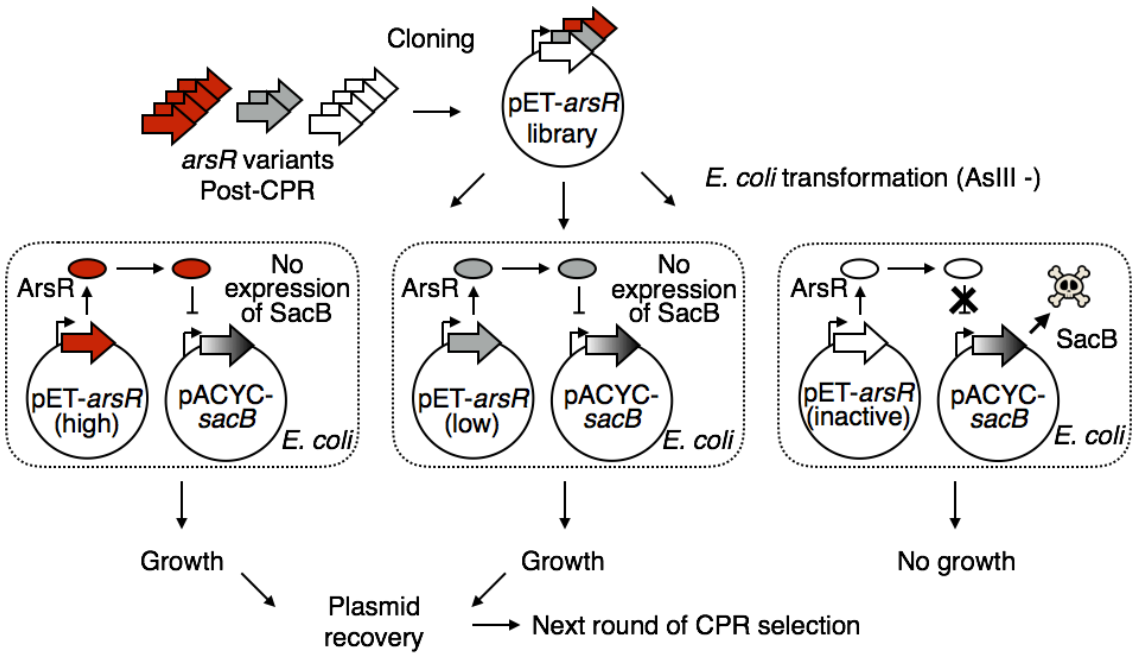


Figure 3-10. OFF selection of ArsR mutants in the absence of As(III) to remove inactive FP variants. The *arsR* variant genes recovered from emulsion PCR is re-cloned into plasmid and is transformed into *E. coli* with pACYC-*sacB*. Since the SacB expression is lethal to *E. coli* in the presence of sucrose, the false positive variant lacks the ability to repress downstream gene expression leading to sucrose sensitivity in the absence of As(III). The active clones resistant to sucrose are taken for the next round of ON selection.

3.2.7 Proof-of-principle experiments for CPR selection

In order to confirm the performance of the strain with dual plasmids pET-*arsR*/pACYC-*Tth* according to arsenic induction (**Figure 3-9**), I performed PCR with arsenite-induced cells. Consequently, amplified product of target gene at 500 bp was observed as expected (**Figure 3-11, bottom**). This result was also confirmed by Western blotting of the HA-tagged DNA polymerase using anti-HA-tag antibody (**Figure 3-11, top**).

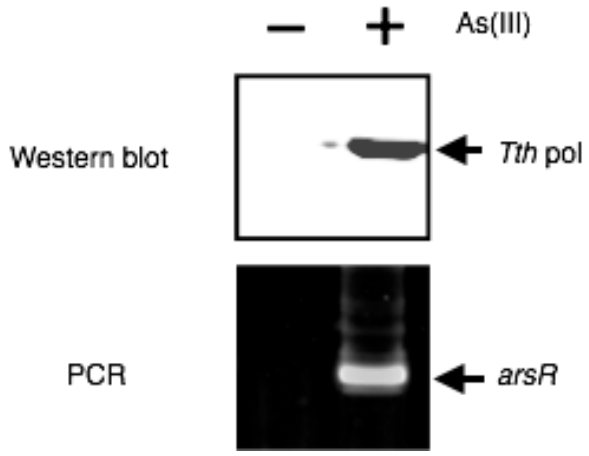


Figure 3-11. Expression and activity of DNA polymerase in *E. coli* containing dual plasmids pET-*arsR*/pACYC-*Tth* in the presence of 10 μg/L of As(III). Western blotting (top) and PCR (bottom).

Next, as a proof-of-principle experiment of CPR selection, I created a model library by mixing the pET-*arsR*/pACYC-*Tth* strain and a strain harboring pET-*gfp* in which *arsR* was replaced with *gfp* at ratio of 1:10. The cell mixture was compartmentalized in water-in-oil emulsion. In theory (**Figure 3-9**), expression of *Tth* pol in the pET-*arsR*/pACYC-*Tth* strain depends on the relation between ArsR binding activity and arsenite concentration. In contrast, the pET-*gfp*/pACYC-*Tth* strain expresses *Tth* pol under any circumstances. Thus, without arsenite in the system, only the pET-*gfp*/pACYC-*Tth* strain produces DNA polymerase leading amplification of the target gene, *i.e.*, *gfp*. As shown in **Figure 3-12**, the stronger amplification of *gfp* gene was observed under the emulsified and no arsenite condition as expected. The band intensity profiles suggested 15-fold enrichment by a single round of CPR selection. These proof-of-principle experiments indicated that *arsR* variant with higher sensitivity on arsenite or inactive *arsR* will be enriched by the positive selection system.

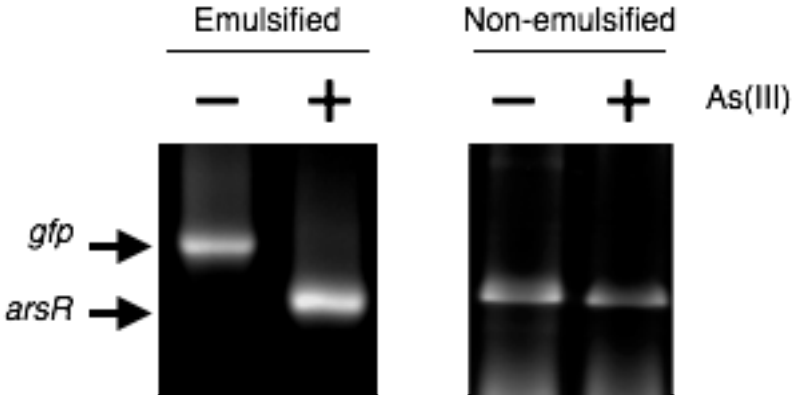


Figure 3-12. *E. coli* cells with pET-*gfp*/pACYC-*Tth* and pET-*arsR*/pACYC-*Tth* were mixed in a ratio of 1:10 and subjected to single round of CPR selection (emulsified). The resulted PCR products were analyzed by agarose gel electrophoresis.

3.2.8 Proof-of-principle experiments for OFF selection

Since the positive selection system developed above also enrich inactive *arsR* variants, I next checked OFF selection system designed above. First, I confirmed the viability of pET-*arsR*/pACYC-*sacB* strain (**Figure 3-10**). When sucrose was not added in medium, arsenite addition did not affect the growth of cells. In contrast, under sucrose medium condition, arsenite addition that activates *sacB* expression induced cell lethality as expected (**Figure 3-13**). These results indicated that *arsR* activity is required for the viability of cells harboring pACYC-*sacB* in the sucrose medium.

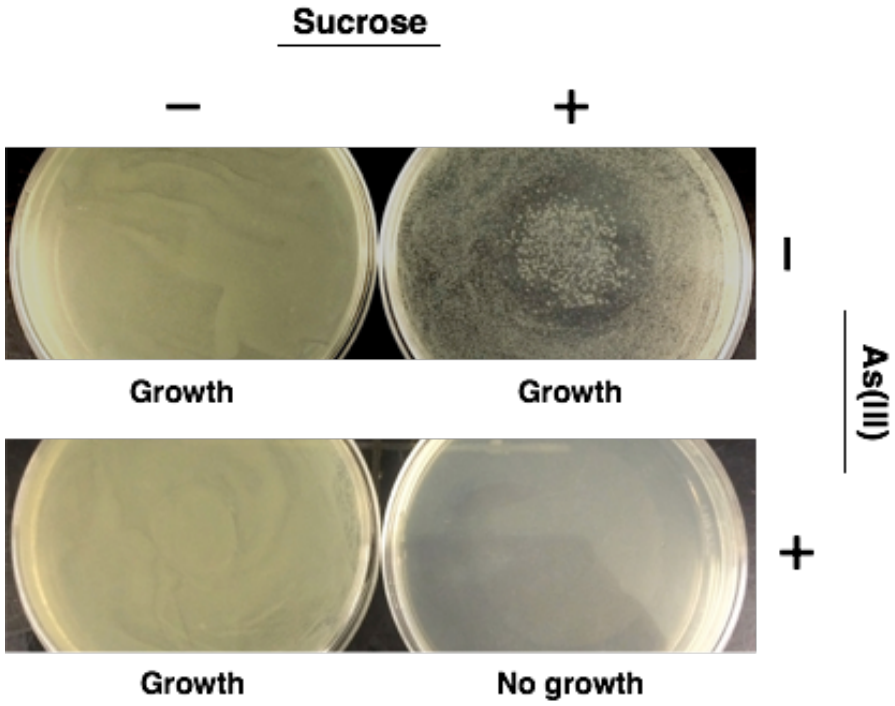
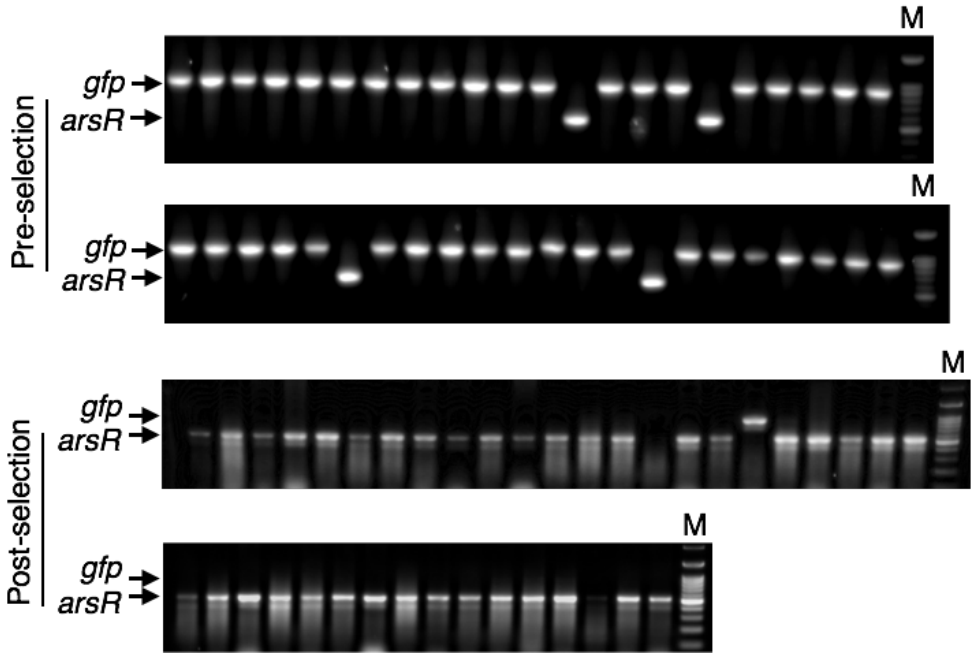


Figure 3-13. Growth of *E. coli* with pET-*arsR*/pACYC-*sacB* induced with or without As(III) was examined on agar plates with or without sucrose.

With this result, I proceeded to the enrichment experiment for the counter selection. Similarly to the positive selection, a model library was prepared by mixing the pET-*arsR*/pACYC-*sacB* and pET-*gfp*/pACYC-*sacB* strains at ratio of 1:10. Colony-direct PCR with common primers for *gfp* and *arsR* showed the model library actually prepared (**Figure 3-14**). Next, the model library cells were plated on sucrose medium, and resultant colonies were randomly picked up to colony-direct PCR. In this case, it was expected that only the strain that repress *sacB* expression under no arsenite conditions formed single colonies. Thus, colonies derived from the pET-*arsR*/pACYC-*sacB* strain was expected to be enriched by the selection. As expected, majorities of colonies after the OFF selection were the pET-*arsR*/pACYC-*sacB* strain (**Figure 3-14A**). As summarized in **Figure 3-14B**, 37-fold enrichment of active *arsR* gene was observed over a single round of the selection. This result strongly supported that inactive *arsR* as the false positives of the CPR selection were efficiently eliminated through the OFF selection developed here.

(A)



(B)

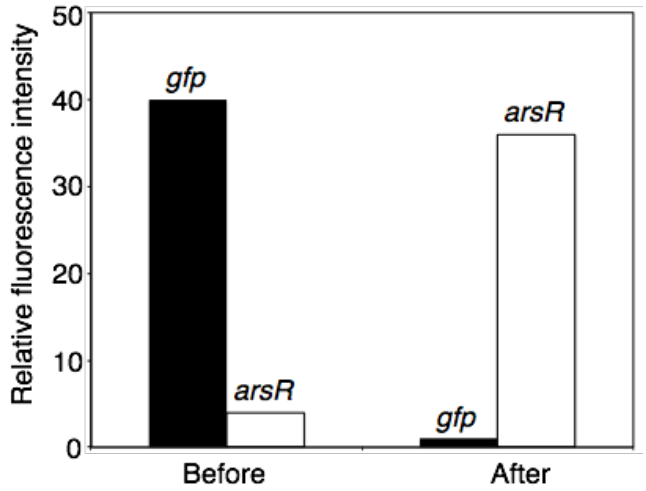


Figure 3-14. *E. coli* cells having pET-*arsR*/pACYC-*sacB* and pET-*gfp*/pACYC-*sacB* were mixed in a ratio of 10:1 and then subjected to single round of OFF selection in the absence of As(III). (A) The *arsR* or *gfp* gene on pET-vector in colonies grown on plates containing no sucrose (Pre-selection) or 5% sucrose (Post-selection) were amplified by colony-direct PCR and analyzed by agarose gel electrophoresis. (B) The number of colonies containing pET-*gfp* (black bar) or pET-*arsR* (white bar) before and after OFF selection was counted.

3.3 Discussion

Here, I developed an arsenite biosensor strain and established a dual-selection system toward its directed evolution. The biosensor can be stored without any temperature control during at least 2 days, which may have sufficient tolerance for daily surveillance of environmental arsenite near human habitations. Selection for cells in the ON state was achieved by coupling the expression of a gene required for active gene amplification and that in the OFF state was coupled to cell viability. Although I only showed that separation of ON/OFF state by the selection system, I believe that *arsR* variants with higher activity can be selected because of the potential of CPR system (Ellefson et al. 2014). Thus, the dual system will be utilized to select *arsR* variants with increased ability that contributes to improvement of the arsenite biosensor in near future.

Although the dual selection system developed here focused on the ArsR system in *E. coli*, the system can be applied to any genetic circuit using pairs of transcriptional regulator and its elements to regulate gene expression levels. The dual selection system will be most useful in the initial stages of constructing complex circuits, especially when there is only a small functional fraction of a large pool of candidates. It will be also possible to select for specific circuit performance by adjusting the selection pressure. The simplicity of the strategy and strength of this selection method should facilitate optimization of genetic switches and circuits in the future.

Recently Li *et al*. reported an evolved arsenic biosensor with improved sensitivity to the arsenic *via* directed evolution using fluorescence activated cell sorting (FACS) (Li et al. 2015). Although FACS is a powerful tool to screen a biosensor cells with GFP as a reporter, the CPR selection system based on DNA polymerase as a reporter have following advantages: (i) it is not need to determine an appropriate threshold to select mutants with a higher sensitivity from a DNA library for each evolutionary cycle and (ii) the highly expensive FACS instrument is not required. These features are very important for applying the dual-selection system to versatile usage as described above.

3.4 Materials and Methods

3.4.1 Bacterial strains and culture condition

Escherichia coli strain XL10-Gold and HST08 (Agilent Technologies) were used for all experiments of constructing plasmids and BL21(DE3) codon-plus RIPL (Stratagene) for protein expression. Bacterial cultures were grown at 37°C in 2×YT or LB media containing kanamycin (25 μg/ml) (Wako), ampicillin (100 μg/ml), or both according to the required condition. Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma Aldrich) and sodium arsenate (Nacalai Tesque) for required concentration were added to the culture for induction purposes.

3.4.2 Construction of plasmids

All primer sequences used in this chapter are listed in **Table 3-1**. All cloning was performed by isothermal assembly (Gibson et al. 2009) after PCR amplification with primers that contain 15-20 bp overlapped regions. For each assembled reaction, equal molar volume of insert and vector PCR products were mixed with 2× Gibson assembly master mix (New England Biolabs) and incubated at 50°C for 15 min. All transformation was performed by heat shock method using 42°C for 45 sec. For dual plasmids transformation to obtain whole-cell biosensor, *E. coli* BL21(DE3) codon-plus RIPL cell carrying the first plasmid was beforehand prepared to be chemically competent followed by transformation with the second plasmid. Positive colonies were checked by colony-direct PCR with primers ACYC-Up1 and T7R targeting both inserts in each plasmid.

For construction of pET-*arsR*, pET29 vector (Novagen) linearized by pET29-Up and -Down was ligated with an *arsR* gene amplified from *E. coli* K12 strain using the primers arsR-F and -R. pET-*Tth* was constructed in the previous Chapter 2.

pACYC-*gfp* was constructed as follows. Ars operator/promoter sequence, a *gfp* gene, and pACYCDuet1 vector (Novagen) were first amplified by arsO/P-F and -R, sfGFP-F and -R, and ACYC-Up1 and -Down1, respectively, and three sequences were ligated by the Gibson assembly. After ligation, the chloramphenicol resistance gene in

pACYC was replaced with the ampicillin resistance gene from pUC19 using primers pUC19-F and -R, and pACYC-Up2 and -Down2. Then, T7 promoter from resulted plasmid was removed by using primers pACYC-Up3 and -Down3.

pACYC-*Tth* and pACYC-*sacB* were constructed using primers pACYC-Up4 and -Down5 for the vector, and Tth-F and -R for a modified *Tth* DNA pol I gene as constructed in the previous Chapter 2 or sacB-F and -R for *sacB* gene from pKM154 (Addgene), respectively. The null plasmid pACYC was constructed by deleting the *gfp* gene from pACYC-*gfp* by primers pACYC-Up3 and -Down4.

3.4.3 Induction of biosensors and fluorescence measurements

Induction condition for both large culture volume (3-5 mL) and small culture volume (200 μL) was as follows. Single colony was picked and incubated overnight at 37°C. Overnight cultures were diluted to OD_{600} (~0.1) and incubated until OD becomes 0.4 and induced with required concentrations of IPTG and/or As(III) ($AsNaO_2$). GFP expression was measured on 10% SDS-PAGE gels with Molecular Imager FX (Bio-Rad) or on 96-well microtiter plates (Thermo Fisher Scientific) with TECAN Safire Fluorescence microplate reader (MTX Lab Systems) at excitation/emission wavelengths of 470 nm/520 nm.

3.4.4 Compartmentalized partnered replication (CPR selection)

CPR was performed essentially as described by Ellefson *et al*. (Ellefson et al. 2014). The pre-culture of the library was prepared in LB broth and incubated at 37°C overnight. Then, 100 μl of overnight culture was diluted in 2 mL of 2× YT and shaken 37°C for 1 h. When the culture entered the log-phase, the culture was induced with required concentration of arsenic and kept shaking for 4 h. Approximately 10^9 cells were mixed with 200 μl PCR reaction mixture that contains PCR buffer, dNTPs and primers targeting the *arsR* gene. The resulted aqueous phase was added into 600 μL of oil mixture at 20 sec interval while rotating at maximum speed in 4°C with additional spinning for 5 min. The resulted water in oil reaction mixture was transferred to PCR tubes and

performed the PCR of 20 cycles (95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min/kb). After PCR, collection was done by centrifugation. In detail, emulsion-PCR products were transferred to 1.7 mL micro centrifuge tube and spun at 15000 rpm for 10 min. Upper layer above white film (oil phase) was discarded and washed with 1 mL mineral oil by repeating the centrifugation process. The remaining aqueous phase was spun for 20 min. Then, the aqueous phase was transferred to a new tube using gel loading pipette tip by penetrating the white film layer.

3.4.5 Levansucrase activity test (OFF selection)

The pET29-derived plasmid library was transformed into the strain having pACYC-*sacB*. The pre-culture of the collected transformants was prepared by incubating at 37°C overnight. This full-grown culture was mixed in 1/100 volume with fresh medium and shaken at 37°C for 6 h until the OD becomes 1. This culture was plated on the LB agar plates with or without 5% sucrose. The colonies grown on sucrose-containing medium were checked by colony-direct PCR with primers T7-promoter and T7-terminator.

Table 3-3. Oligonucleotides used in the chapter 3

Primer	Sequence (5' to 3')
ACYC-Up1	GGATCTCGACGCTCTCCCT
arsO/P-F	CATCACCACAGCCAGGATCCCATTCGTTAAGTCATATATGTTT TTGACTTATCCGCTTCGAAGAGAGA
arsO/P-R	AGTTGTTCTCCTTTACTCATATTGCGCTCCTGATTGTTGCAGG TAGTGTCTCTCTTCGAAGCG
arsR-F	TTTAAGAAGGAGATATACATATGTCATTTCTGTACCCATCC
arsR-R	GTGCTCGAGGGATCCACTGCAAATGTTCTTAC
pACYC-Down1	TCGAGTCTGGTAAAGAAACC
pACYC-Down2	TTTAGCTTCCTTAGCTCCTG
pACYC-Down3	TAGCGACTCCTGCATTAGGAAGGAGATATACCATGGGCA
pACYC-Down4	TTGCGACTCCTGCATTAGGTCGAGTCYGGTAAAGAAACC
pACYC-Down5	TCGAGTCTGGTAAAGAAACC
pACYC-Up1	GGATCCTGGCTGTGGTGATG
pACYC-Up2	TAATTTTTTTAAGGCAGTTA
pACYC-Up3	CCTAATGCAGGAGTCGCATA
pACYC-Up4	ATTGCGTCCTGATTGTTGC
pET29-Up	CATATGTATATCTCCTTCTTAAAGTTAAAC
pET29-Down	GGATCCCTCGAGCACCACC
pUC19-F	TAACTGCCTTAAAAAAATTATTACCAATGCTTAATCAGTG
pUC19-R	CAGGAGCTAAGGAAGCTAAAATGAGTATTCAACATTTCCG
sacB-F	GCAACAATCAGGAGCGCAATATGAACATCAAAAAGTTTGC
sacB-R	GGTTTCTTTACCAGACTCGATTATTTGTTAACTGTTAA
sfGFP-F	ATGAGTAAAGGAGAAGAACT
sfGFP-R	GGTTTCTTTACCAGACTCGATCATTTGTAGAGCTCATCCA
T7-promoter	TAATACGACTCACACTATAGGG
T7-terminator	GCTAGTTATTGCTCAGCGG
Tth-F	GCAACAATCAGGAGCGCAATTTAATACGACTCA
Tth-R	CAAAAAACCCCTCAAGACCC

Chapter 4
Conclusion

4.1 Summary

Although compartmentalized replication such as CSR and CPR is a powerful tool for directed evolution of proteins and gene circuits, limitations remain in the emulsion PCR process with the wild-type *Taq* pol used so far, including long run times, low amounts of product, and false negative results due to inhibitors. In Chapter 2, I developed a high-efficiency mutant of *Tth* pol suited for CSR and CPR. First I modified the wild-type *Tth* pol by (i) deletion of the N-terminal 5' to 3' exonuclease domain, (ii) fusion with the DNA-binding protein Sso7d, (iii) introduction of four known effective point mutations from other DNA polymerase mutants, and (iv) codon optimization to reduce the GC content. Consequently, I obtained a mutant that provides higher product yields than the conventional *Taq* pol without decreased fidelity. Next, I performed four rounds of CSR selection with a randomly mutated library of this modified *Tth* pol and obtained mutants that provide higher product yields in fewer cycles of emulsion PCR than the parent *Tth* pol as well as the conventional *Taq* pol.

Arsenic is a well-known toxin that holds a constant threat to a large population worldwide. My ultimate goal is to develop an arsenite-inducible biosensor with the best possible sensitivity and specificity. In order to evolve the *ars* operon of *E. coli*, I intend to apply CPR, in which the *ars* operon is coupled with expression of *Tth* pol mutant described above. In Chapter 3, I proposed a positive-negative selection system for selecting a transcription factor that activates gene expression in response to arsenic in solution. First, I developed a whole cell biosensor for sensing arsenite in liquid using a regulator (ArsR) and a reporter (GFP), and evaluated its performance. Second, I developed a positive selection system for active ArsR using compartmentalized partnered replication which uses thermostable DNA polymerase as the reporter of activity. Third, I developed a negative selection system using sucrose-induced suicide gene SacB as the reporter for exclusion of inactive ArsR variants. The dual selection developed here will contribute to directed evolution of whole cell biosensors to detect arsenic contamination in groundwater.

4.2 Future perspectives

Directed evolution is an extremely efficient method for optimizing biocatalyst for practical applications. The key to a directed evolution program is the development of a reliable high-throughput screen, coupled with effective methods for generating libraries that are filled with variants of new function. Many enzymes have proven to be remarkably evolvable for the enhancement of solubility, thermostability, catalytic turnover or affinity for substrates. These properties remain important in contemporary directed evolution because increased stability and activity can facilitate the engineering or evolution of other desirable properties. On the other hand, the more moderate successes with engineering transcription factors reflect limitations to their evolvability and consequently to the less effort to engineering them.

In this study, the ease of coupling the function of transcription factors to thermostable DNA polymerase or to cell growth offers the significant advantages for their directed evolution. Biological parts evolved in bacteria can be incorporated directly into pathways and circuits in other organisms such as *E. coli* (Esvelt et al. 2011; McCafferty et al. 1990)) and yeast (Roodveldt et al. 2005) or an artificial cell (Fujii et al. 2015). New selection technique that can carry out more continuous rounds of evolution along with higher throughput can widen the exploration of the fitness landscape. Therefore, novel directed evolution methods will continue to generate proteins with useful new activities and specificities. Such methods as one demonstrated here will expand the scope of protein engineering to larger sets of chemical and biological functions.

The *Tth* pol mutants with resistance to emulsion PCR described in Chapter 2 will be promising tools for emulsion PCR-based biotechnologies, such as (i) CPR for the directed evolution of various genetic circuits (Ellefson et al. 2014) and (ii) the preparation of DNA libraries for *in vitro* selection and high-throughput sequencing (Diehl et al. 2006; Williams et al. 2006; Sumida et al. 2012).

Engineering and design of genetic circuit in living cell is critical in accessing the beneficial application of synthetic biology. Directed evolution can avoid the complicated rational design of such circuit by screening or selecting functional circuit from non-functional one. In Chapter 3, I illustrated a dual selection system based on our modified

Tth DNA polymerase gene for ON selection and a suicide *sacB* gene for OFF selection of an arsenic-responsive whole cell biosensor. The genetic switch in this system can serve generally for engineering transcriptional regulators with ON/OFF responses. The dual selection system will be most useful in the initial stages of constructing complex circuits, especially when there is only a small functional fraction of a large pool of candidates. It is also possible to select for specific circuit performance by adjusting the selection pressure. The simplicity of the strategy and strength of this selection method should facilitate optimization of genetic switches and circuits in the future.

References

Abil, Z., J. W. Ellefson, J. D. Gollihar, E. Watkins, and A. D. Ellington. 2017. Compartmentalized partnered replication for the directed evolution of genetic parts and circuits. *Nat Protoc* 12 (12):2493-2512.

Agback, P., H. Baumann, S. Knapp, R. Ladenstein, and T. Hard. 1998. Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. *Nat Struct Biol* 5 (7):579-584.

Al-Soud, W. A., and P. Radstrom. 2001. Purification and characterization of PCR-inhibitory components in blood cells. *J Clin Microbiol* 39 (2):485-493.

Anawar, H.M., 2012. Arseni speciation in environmental samples by hydride generation and electrothermal atomic absorption spectrometry. *Talanta* 88, 30-42.

Arakawa, T., B. Jongsareejit, Y. Tatsumi, K. Tanaka, K. Ikeda, H. Komatsubara, H. Inoue, B. Kawakami, M. Oka, S. Emi, T. Yomo, Y. Shima, S. Negoro, and I. Urabe. 1996. Application of N-terminally truncated DNA polymerase from Thermus thermophilus (delta Tth polymerase) to DNA sequencing and polymerase chain reactions: comparative study of delta Tth and wild-type Tth polymerases. *DNA Res* 3 (2):87-92.

Arezi, B., N. McKinney, C. Hansen, M. Cayouette, J. Fox, K. Chen, J. Lapira, S. Hamilton, and H. Hogrefe. 2014. Compartmentalized self-replication under fast PCR cycling conditions yields Taq DNA polymerase mutants with increased DNA-binding affinity and blood resistance. *Front Microbiol* 5:408.

Bakhrat, A., Etzov, E., Finkelstein, Y., Marks, R., Raveh, D., 2011. UV and arsenate toxivity: a specific and sensitive yeast bioluminescence assay. *Cell Biol. Toxicol.* 27 (3), 227-236.

Bessette, P. H., J. J. Rice, and P. S. Daugherty. 2004. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng Des Sel* 17 (10):731-739.

Buffi, N., Merulla, D., Beutier, J., Barbaud, F., Beggah, S., van Lintel, H., Renaud, P., van der Meer, J.R., 2011. Development of a microfluidics biosensor for agarose-bead immobilized *Escherichia coli* bioreporter cells for arsenite detection in aqueous samples. *Lab Chip* 11 (14), 2369-2377.

Cai, S., Y. Shen, Y. Zou, P. Sun, W. Wei, J. Zhao, and C. Zhang. 2018. Engineering highly sensitive whole-cell mercury biosensors based on positive feedback loops from quorum-sensing systems. *Analyst* 143 (3):630-634.

Carballeira, N., M. Nazabal, J. Brito, and O. Garcia. 1990. Purification of a thermostable DNA polymerase from Thermus thermophilus HB8, useful in the polymerase chain reaction. *Biotechniques* 9 (3):276-281.

Carlin, A., W. Shi, S. Dey, and B. P. Rosen. 1995. The ars operon of Escherichia coli confers arsenical and antimonial resistance. *J Bacteriol* 177 (4):981-986.

Chen, D., Huang, Ce, M., Hu, B., 2009. Application of rough set on train operation adjustment. *J. Hazard. Mater.* 164 (2-3), 1146-1151.

Chen, J., and B. P. Rosen. 2014. Biosensors for inorganic and organic arsenicals. *Biosensors (Basel)* 4 (4):494-512.

Chen, J., Zhu, Y.G., Rosen, B.P., 2012. A novel biosensor selective for organoarsenicals. *Appl. Environ. Microbiol*. 78 (19), 7145-7147.

Cline, J., J. C. Braman, and H. H. Hogrefe. 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* 24 (18):3546-3551.
Chiou, C.H., Chien, L.J., Chou, T.C., Lin, J.L., Tseng, J.T., 2011. Rapid whole-cell sensing chip for low-level arsenite detection. *Biosens. Bioelectron*. 26 2484-2488.
Coco, W. M., W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires, and D. J. Monticello. 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* 19 (4):354-359.
Consonni, R., L. Santomo, P. Fusi, P. Tortora, and L. Zetta. 1999. A single-point mutation in the extreme heat- and pressure-resistant sso7d protein from sulfolobus solfataricus leads to a major rearrangement of the hydrophobic core. *Biochemistry* 38 (39):12709-12717.
Cortes-Salazar, F., Beggah, S., van der Meer, J.R., Girault, H.H., 2013.Electrochemical As(III) whole-cell based biochip sensor. *Biosens. Bioelectron*. 47, 237-242.
De Mora, K., Joshi, N., Balint, B., Ward, F., Elfick, A., French, C., 2011. A pH-based biosensor for detection of arsenic in drinking water. *Anal. Bioanal. Chem.* 400 (4), 1031-1039.
Diehl, F., M. Li, Y. He, K. W. Kinzler, B. Vogelstein, and D. Dressman. 2006. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* 3 (7):551-559.
Diorio, C., J. Cai, J. Marmor, R. Shinder, and M. S. DuBow. 1995. An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. *J Bacteriol* 177 (8):2050-2056.
Driscoll, C. A., D. W. Macdonald, and S. J. O'Brien. 2009. From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci U S A* 106 Suppl 1:9971-9978.
Eckert, K. A., and T. A. Kunkel. 1990. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res* 18 (13):3739-3744.
Ellefson, J. W., A. J. Meyer, R. A. Hughes, J. R. Cannon, J. S. Brodbelt, and A. D. Ellington. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol* 32 (1):97-101.
Ellington, A.D., Ellefson, J. W., Meyer, A. J. Hughes, R. A., Cannon, J. R., and Brodbelt, J. S. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol* 32 (1): 97-101.
Erickson, M. L., and R. J. Barnes. 2005. Well characteristics influencing arsenic concentrations in ground water. *Water Res* 39 (16):4029-4039.
Esteban, O., R. D. Woodyer, and H. Zhao. 2003. In vitro DNA recombination by random priming. *Methods Mol Biol* 231:99-104.
Esvelt, K. M., J. C. Carlson, and D. R. Liu. 2011. A system for the continuous directed evolution of biomolecules. *Nature* 472 (7344):499-503.
Ezeh, V.C. Harrop, T.C., 2013. Syntesis and properties of arsenic (III) reactive coimarin appended benzothiazolines: A new approach for inorganic arsenic detection. *Inorg. Chem.* 52 (5), 2323-2334.
Fendorf, S., H. A. Michael, and A. van Geen. 2010. Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science* 328 (5982):1123-1127.

Fernandez, M., B. Morel, J. L. Ramos, and T. Krell. 2016. Paralogous Regulators ArsR1 and ArsR2 of Pseudomonas putida KT2440 as a Basis for Arsenic Biosensor Development. *Appl Environ Microbiol* 82 (14):4133-4144.

Forzani, E.S., Foley, K., Westerhoff, P., Tao, N., 2007. Detection of arsenic in groundwater using a surface plasmon resonance sensor. *Sens. Actuators B-Che.* 123, 82-88.

Fujii, S., T. Matsuura, and T. Yomo. 2015. In vitro directed evolution of alpha-hemolysin by liposome display. *Biophysics (Nagoya-shi)* 11:67-72.

Gao, C., Yu, X.Y., Xiong, S.Q., Liu, J.H., Huang, X.J., 2013. Electrochemical detection of arsenic(III) completely free from nobel metal: Fe_3O_4 microspheres-room temperature ionic liquid composite showing better performance than gold. *Anal. Chem*. 85 (5), 2673-2680.

Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* 164 (2):918-921.

Gay, P., D. Le Coq, M. Steinmetz, E. Ferrari, and J. A. Hoch. 1983. Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. *J Bacteriol* 153 (3):1424-1431.

Ghadessy, F. J., J. L. Ong, and P. Holliger. 2001. Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci U S A* 98 (8):4552-4557.

Gibbon-Walsh, K., Green, T., Leskinen, P., Ibask, A., Marks, R., Virta, M., 2004. Arsenic speciation in natural water by cathodic stripping voltammetry. *Anal. Chim. Acta* 662 (1), 1-8.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6 (5):343-345.

Glukhov, A. I., S. A. Gordeev, S. V. Vinogradov, V. I. Kiselev, V. M. Kramarov, O. I. Kiselev, and E. S. Severin. 1990. [Amplification of DNA sequences in Epstein-Barr virus and human immunodeficiency virus using DNA polymerase from Thermus thermophilus]. *Mol Biol (Mosk)* 24 (3):781-787.

Goldsmith, M., and D. S. Tawfik. 2012. Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol* 22 (4):406-412.

Goldsmith, M., and D. S. Tawfik. 2017. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr Opin Struct Biol* 47:140-150.

Greener, A., M. Callahan, and B. Jerpseth. 1996. An efficient random mutagenesis technique using an E. coli mutator strain. *Methods Mol Biol* 57:375-385.

Hanes, J., and A. Pluckthun. 1997. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A* 94 (10):4937-4942.

Henne, A., H. Bruggemann, C. Raasch, A. Wiezer, T. Hartsch, H. Liesegang, A. Johann, T. Lienard, O. Gohl, R. Martinez-Arias, C. Jacobi, V. Starkuviene, S. Schlenczeck, S. Dencker, R. Huber, H. P. Klenk, W. Kramer, R. Merkl, G. Gottschalk, and H. J. Fritz. 2004. The genome sequence of the extreme thermophile Thermus thermophilus. *Nat Biotechnol* 22 (5):547-553.

Hou, Q.H., Ma, A.z., Lv, D., Bai, Z.H., Zhuang, G.Q., 2014. Appl. The impacts of different long-term fertilization regimes on the bioavilability of arsenic in soil:

integrating chemical approach with *Escherichia coli arsRp::luc*-based biosensor. *Microbiol. Biotechnol*. 98 (13), 6137-6146.
Hu, Q., L. Li, Y. Wang, W. Zhao, H. Qi, and G. Zhuang. 2010. Construction of WCB-11: a novel phiYFP arsenic-resistant whole-cell biosensor. *J Environ Sci (China)* 22 (9):1469-1474.
Huang, C. W., C. C. Wei, and V. H. Liao. 2015. A low cost color-based bacterial biosensor for measuring arsenic in groundwater. *Chemosphere* 141:44-49.
Innis, M. A., K. B. Myambo, D. H. Gelfand, and M. A. Brow. 1988. DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc Natl Acad Sci U S A* 85 (24):9436-9440.
Kadonaga, J. T., and J. R. Knowles. 1985. A simple and efficient method for chemical mutagenesis of DNA. *Nucleic Acids Res* 13 (5):1733-1745.
Kermekchiev, M. B., L. I. Kirilova, E. E. Vail, and W. M. Barnes. 2009. Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res* 37 (5):e40.
Kermekchiev, M. B., A. Tzekov, and W. M. Barnes. 2003. Cold-sensitive mutants of Taq DNA polymerase provide a hot start for PCR. *Nucleic Acids Res* 31 (21):6139-6147.
Kim, K. W., P. Chanpiwat, H. T. Hanh, K. Phan, and S. Sthiannopkao. 2011. Arsenic geochemistry of groundwater in Southeast Asia. *Front Med* 5 (4):420-433.
Korolev, S., M. Nayal, W. M. Barnes, E. Di Cera, and G. Waksman. 1995. Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-A resolution: structural basis for thermostability. *Proc Natl Acad Sci U S A* 92 (20):9264-9268.
Kumar, A., M. S. Rahman, M. A. Iqubal, M. Ali, P. K. Niraj, G. Anand, P. Kumar, Abhinav, and A. K. Ghosh. 2016. Ground Water Arsenic Contamination: A Local Survey in India. *Int J Prev Med* 7:100.
Lawyer, F. C., S. Stoffel, R. K. Saiki, S. Y. Chang, P. A. Landre, R. D. Abramson, and D. H. Gelfand. 1993. High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl* 2 (4):275-287.
Li, L., J. Liang, W. Hong, Y. Zhao, S. Sun, X. Yang, A. Xu, H. Hang, L. Wu, and S. Chen. 2015. Evolved bacterial biosensor for arsenite detection in environmental water. *Environ Sci Technol* 49 (10):6149-6155.
Li, Y., S. Korolev, and G. Waksman. 1998. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J* 17 (24):7514-7525.
Liao, V. H., and K. L. Ou. 2005. Development and testing of a green fluorescent protein-based bacterial biosensor for measuring bioavailable arsenic in contaminated groundwater samples. *Environ Toxicol Chem* 24 (7):1624-1631.
Lin, T. C., G. Karam, and W. H. Konigsberg. 1994. Isolation, characterization, and kinetic properties of truncated forms of T4 DNA polymerase that exhibit 3'-5' exonuclease activity. *J Biol Chem* 269 (30):19286-19294.

Lohar, S., Sahana, A., Banerjee, A., Banik, A., Mukhopadhyay, S.K., Sanmartin Matalobos, J.S., Das, D., 2013. Antipyrine based arsenate selective fluorescent probe for living cell imaging. *Anal. Chem*. 85 (3), 1778-1783.

Lubick, N. 2008. Mapping fluoride and arsenic hot spots. *Environ Sci Technol* 42 (10):3483-3484.

Lundberg, K. S., D. D. Shoemaker, M. W. Adams, J. M. Short, J. A. Sorge, and E. J. Mathur. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* 108 (1):1-6.

Lutz, S. 2010. Beyond directed evolution--semi-rational protein engineering and design. *Curr Opin Biotechnol* 21 (6):734-743.

McCafferty, J., A. D. Griffiths, G. Winter, and D. J. Chiswell. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348 (6301):552-554.

McCullum, E. O., B. A. Williams, J. Zhang, and J. C. Chaput. 2010. Random mutagenesis by error-prone PCR. *Methods Mol Biol* 634:103-109.

Molina-Espeja, P., J. Vina-Gonzalez, B. J. Gomez-Fernandez, J. Martin-Diaz, E. Garcia-Ruiz, and M. Alcalde. 2016. Beyond the outer limits of nature by directed evolution. *Biotechnol Adv* 34 (5):754-767.

Moore, P. 2005. PCR: replicating success. *Nature* 435 (7039):235-238.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1:263-273.

Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-350.

Muteeb, G., and R. Sen. 2010. Random mutagenesis using a mutator strain. *Methods Mol Biol* 634:411-419.

Myers, T. W., and D. H. Gelfand. 1991. Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase. *Biochemistry* 30 (31):7661-7666.

Nordstrom, D. K. 2002. Public health. Worldwide occurrences of arsenic in ground water. *Science* 296 (5576):2143-2145.

Oremland, R. S., and J. F. Stolz. 2003. The ecology of arsenic. *Science* 300 (5621):939-944.

Orencia, M. C., J. S. Yoon, J. E. Ness, W. P. Stemmer, and R. C. Stevens. 2001. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat Struct Biol* 8 (3):238-242.

Owolabi, J. B., and B. P. Rosen. 1990. Differential mRNA stability controls relative gene expression within the plasmid-encoded arsenical resistance operon. *J Bacteriol* 172 (5):2367-2371.

Packer, M.S. and Liu, D. R. 2015. Methods for the directed evolution of proteins. *Nat Rev Genet* 16(7):379-94.

Palmer, A. C., and R. Kishony. 2013. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet* 14 (4):243-248.

Patrick, W. M., and M. L. Gerth. 2014. ITCHY: Incremental Truncation for the Creation of Hybrid enzYmes. *Methods Mol Biol* 1179:225-244.

Peca, L., P. B. Kos, Z. Mate, A. Farsang, and I. Vass. 2008. Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc. *FEMS Microbiol Lett* 289 (2):258-264.

Pelicic, V., J. M. Reyrat, and B. Gicquel. 1996. Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. *J Bacteriol* 178 (4):1197-1199.

Petursdottir, A.H., Gunnlaugsdottir, H., Jorundsdottir, H., Mestrot, A., Krupp, E.M., Feldmann, J., 2012. HPLC-HG-ICP-MS: a sensitive and selective method for inorganic arsenic in seafood. *Anal. Bioanal. Chem.* 404 (8), 2185-2191.

Porcar, M. 2010. Beyond directed evolution: Darwinian selection as a tool for synthetic biology. *Syst Synth Biol* 4 (1):1-6.

Povilaitis, T., G. Alzbutas, R. Sukackaite, J. Siurkus, and R. Skirgaila. 2016. In vitro evolution of phi29 DNA polymerase using isothermal compartmentalized self replication technique. *Protein Eng Des Sel*.

Prado, F., and A. Aguilera. 1994. New in-vivo cloning methods by homologous recombination in yeast. *Curr Genet* 25 (2):180-183.

Rahman, M. M., R. Naidu, and P. Bhattacharya. 2009. Arsenic contamination in groundwater in the Southeast Asia region. *Environ Geochem Health* 31 Suppl 1:9-21.

Rangan, R., Rastogi, N.K., Thakur, M.S., 2012. Development of immobilized biophotonic beads consisting of *Photobacterium leiognathi* for the detection of heavy metals and pesticides. *J. Hazard. Mater* 225, 114-123.

Rapoport, G., and R. Dedonder. 1966. [Specificity of the levansucrase activity of Bacillus subtilis: clarification]. *Bull Soc Chim Biol (Paris)* 48 (12):1311-1322.

Recorbet, G., C. Robert, A. Givaudan, B. Kudla, P. Normand, and G. Faurie. 1993. Conditional suicide system of Escherichia coli released into soil that uses the Bacillus subtilis sacB gene. *Appl Environ Microbiol* 59 (5):1361-1366.

Ribble, W., S. D. Kane, and J. M. Bullard. 2015. Long-Range PCR Amplification of DNA by DNA Polymerase III Holoenzyme from Thermus thermophilus. *Enzyme Res* 2015:837842.

Richter, J., Lischka, S., Piechotta, C.2012. Analysis of arsenic species in fish after derivatization by GC-MS. *Talanta* 101, 524-529.

Roodveldt, C., A. Aharoni, and D. S. Tawfik. 2005. Directed evolution of proteins for heterologous expression and stability. *Curr Opin Struct Biol* 15 (1):50-56.

Rosen, B. P. 1995. Resistance mechanisms to arsenicals and antimonials. *J Basic Clin Physiol Pharmacol* 6 (3-4):251-263.

Rosen, B. P. 1999. The role of efflux in bacterial resistance to soft metals and metalloids. *Essays Biochem* 34:1-15.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239 (4839):487-491.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230 (4732):1350-1354.

Sergii, I. and Chris, F. 2006. Arsenic biosensor modelling: simple alternative formulation.

Siddiki, M.S.R., Kawakami, Y., Ueda, S., Maeda, I., 2011. Solid phase biosensors for arsenic or cadmium composed of A *trans* Factor and *cis* element complex. *Sensors* 11 (11), 10063-10073.

Sharma, P., Asad, S., Bhattacharyay, D., Turner, A.O., 2010. Bioluminescent bioreporter for assessment of arsenic contamination in water samples in India. *Ecotoxicol. Environ. Saf*. 73 (6), 1495-1501.

Shetty, R. S., S. K. Deo, P. Shah, Y. Sun, B. P. Rosen, and S. Daunert. 2003. Luminescence-based whole-cell-sensing systems for cadmium and lead using genetically engineered bacteria. *Anal Bioanal Chem* 376 (1):11-17.

Shiratori, M., T. Kobayashi, and T. Shibui. 2009. Identification of amino acids essential for antibody binding by mRNA-display using a random peptide library: an anti-human tumor protein p53 antibody as a model. *Mol Biotechnol* 41 (2):99-105.

Siegfried, K., S. Hahn-Tomer, A. Koelsch, E. Osterwalder, J. Mattusch, H. J. Staerk, J. M. Meichtry, G. E. De Seta, F. D. Reina, C. Panigatti, M. I. Litter, and H. Harms. 2015. Introducing Simple Detection of Bioavailable Arsenic at Rafaela (Santa Fe Province, Argentina) Using the ARSOlux Biosensor. *Int J Environ Res Public Health* 12 (5):5465-5482.

Silver, S., G. Ji, S. Broer, S. Dey, D. Dou, and B. P. Rosen. 1993. Orphan enzyme or patriarch of a new tribe: the arsenic resistance ATPase of bacterial plasmids. *Mol Microbiol* 8 (4):637-642.

Su, S., Y. G. Gao, H. Robinson, Y. C. Liaw, S. P. Edmondson, J. W. Shriver, and A. H. Wang. 2000. Crystal structures of the chromosomal proteins Sso7d/Sac7d bound to DNA containing T-G mismatched base-pairs. *J Mol Biol* 303 (3):395-403.

Sugimoto, M., N. Esaki, H. Tanaka, and K. Soda. 1989. A simple and efficient method for the oligonucleotide-directed mutagenesis using plasmid DNA template and phosphorothioate-modified nucleotide. *Anal Biochem* 179 (2):309-311.

Sumida, T., H. Yanagawa, and N. Doi. 2012. In vitro selection of fab fragments by mRNA display and gene-linking emulsion PCR. *J Nucleic Acids* 2012:371379.

Takeuchi, R., M. Choi, and B. L. Stoddard. 2014. Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proc Natl Acad Sci U S A* 111 (11):4061-4066.

Tahir, M.A., Rasheed, H., Malana, A, 2008. Method development for arsenic analysis by modification in spectrophotometric technique . *Drink Water Eng. Sci. Discuss.* 1, 135-154.

Truffer, F., N. Buffi, D. Merulla, S. Beggah, H. van Lintel, P. Renaud, J. R. van der Meer, and M. Geiser. 2014. Compact portable biosensor for arsenic detection in aqueous samples with Escherichia coli bioreporter cells. *Rev Sci Instrum* 85 (1):015120.

Tubeleviciute, A., and R. Skirgaila. 2010. Compartmentalized self-replication (CSR) selection of Thermococcus litoralis Sh1B DNA polymerase for diminished uracil binding. *Protein Eng Des Sel* 23 (8):589-597.

Villbrandt, B., G. Sagner, and D. Schomburg. 1997. Investigations on the thermostability and function of truncated Thermus aquaticus DNA polymerase fragments. *Protein Eng* 10 (11):1281-1288.

Wang, Y., D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney, and P. B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res* 32 (3):1197-1207.

Watstein, D. M., and M. P. Styczynski. 2018. Development of a Pigment-Based Whole-Cell Zinc Biosensor for Human Serum. *ACS Synth Biol* 7 (1):267-275.

Webster, D. P., M. A. TerAvest, D. F. Doud, A. Chakravorty, E. C. Holmes, C. M. Radens, S. Sureka, J. A. Gralnick, and L. T. Angenent. 2014. An arsenic-specific biosensor with genetically engineered Shewanella oneidensis in a bioelectrochemical system. *Biosens Bioelectron* 62:320-324.

Westberg, J., A. Holmberg, M. Uhlen, and B. Pettersson. 1999. Truncated fragments in polymerase chain reaction-based DNA sequencing. *Electrophoresis* 20 (3):502-510.

Williams, R., S. G. Peisajovich, O. J. Miller, S. Magdassi, D. S. Tawfik, and A. D. Griffiths. 2006. Amplification of complex gene libraries by emulsion PCR. *Nat Methods* 3 (7):545-550.

World Health, O. 2011. Evaluation of certain contaminants in food. *World Health Organ Tech Rep Ser* (959):1-105, back cover.

Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. The effects of artificial selection on the maize genome. *Science* 308 (5726):1310-1314.

Xu, C., T. Zhou, M. Kuroda, and B. P. Rosen. 1998. Metalloid resistance mechanisms in prokaryotes. *J Biochem* 123 (1):16-23.

Yamagami, T., S. Ishino, Y. Kawarabayasi, and Y. Ishino. 2014. Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. *Front Microbiol* 5:461.

Yang, Z., Y. Ding, Y. Zhang, and F. Liu. 2008. Rapid purification of truncated Taq DNA polymerase Stoffel fragment by boiling lysis of bacterial expression cultures. *Biotechnol Appl Biochem* 50 (Pt 2):71-75.

Yu, D., L. Bai, J. Zhai, Y. Wang, and S. Dong. 2017. Toxicity detection in water containing heavy metal ions with a self-powered microbial fuel cell-based biosensor. *Talanta* 168:210-216.

Zhao, H. 2004. Staggered extension process in vitro DNA recombination. *Methods Enzymol* 388:42-49.

Zhao, H., L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16 (3):258-261.

Zoller, M. J., and M. Smith. 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *DNA* 3 (6):479-488.

Zoller, M. J., and M. Smith. 1987. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *Methods Enzymol* 154:329-350.

List of published papers and conference presentations

1. Articles on periodicals (related to thesis)

(1) Aye, S.L., Fujiwara, K., Doi, N.
A dual system using compartmentalized partnered replication for selection of arsenic-responsive transcriptional regulator
The Journal of Biochemistry [Epub ahead of print]
June 8, 2018
DOI:10.1093/jb/mvy055

(2) Aye, S.L., Fujiwara, K., Ueki, A., Doi, N.
Engineering of DNA polymerase I from *Thermus thermophilus* using compartmentalized self-replication
Biochemical and Biophysical Research Communications 499(2): 170-176.
March 21, 2018
DOI:10.1016/j.bbrc.2018.03.098

2. Presentations at domestic meetings

(1) Aye, S.L., Doi, N., Fujiwara, K.
Towards directed evolution of transcriptional regulators using compartmentalized-partnered replication with a novel optimized *Tth* DNA polymerase
Consortium of Biological Sciences 2017 (ConBio 2017)
The 40th annual meeting of The Molecular Biology Society of Japan
December 6, 2017

(2) Aye, S.L., Ueki, A., Fujiwara, K., Doi, N.
Toward directed evolution of bacterial biosensor for arsenite detection by compartmentalized partnered replication
The 54th annual meeting of Biophysical Society of Japan
November 26, 2016

Acknowledgements

This work was carried out during the years 2015-2018 at Laboratory of Biomolecular Engineering, Department of Biosciences and Informatics, Graduate School of Science and Technology, Keio University.

I owe my deepest gratitude to my supervisor Professor Nobuhide Doi, PhD., who took a chance on me. Without his continuous optimism concerning this work, enthusiasm, and encouragement, support to this study, would hardly have been completed. I also express my warmest gratitude to my other supervisor Dr. Kei Fujiwara who kept inspiring me with his brilliance and guided me through the way from beginning till the end.

I am deeply grateful to the reviewers of the thesis, Professor Kotaro Oka, PhD., Department of Biosciences and Informatics, Professor Kenji Miyamoto, PhD., Department of Biosciences and Informatics and Professor Yoshiaki Furukawa, PhD., Department of Chemistry, Keio University.

I also owe a great debt of gratitude to my friend, Ms. Asuka Ueki who gave me an exceptional kindness and unlimited support to accomplish my work from day to day. I am also indebted to Ms. Arisa Yotsui and Ms. Saki Kinoshita for their precious countless assists to me. My appreciation also extends to my laboratory colleagues, Ms. Saiga, Mr. Nakayama, Mr. Umino, Mr. Adachi, Mr. Sudo, Mr. Sawamura, and Mr. Kohyama, all of whom blessed me with a life of joy in and out of the lab.

Last, but not least, I owe my gratitude to Japanese government MEXT scholarship and KLL program that aided me financially for this work to make it happen. I will never forget how lucky I was to have this opportunity of staying under the shade of Keio University for these precious years.

Seaim Lwin Aye