DEVELOPMENT OF A NOVEL T/A CLONING AND EXPRESSION VECTOR FOR EFFICIENT EXPRESSION OF PROTEIN-BASED BIOPHARMACEUTICALS IN *Escherichia coli*.

MUHAMAD ALIF CHE NORDIN

UNIVERSITI SAINS MALAYSIA

2017

DEVELOPMENT OF A NOVEL T/A CLONING AND EXPRESSION VECTOR FOR EFFICIENT EXPRESSION OF PROTEIN-BASED BIOPHARMACEUTICALS IN *Escherichia coli*.

by

MUHAMAD ALIF CHE NORDIN

Thesis submitted in the fulfillment of the requirements

for the degree of

Master of Science

May 2017

ACKNOWLEDGEMENT

Just slap anything on when you see a blank canvas staring you in the face like some imbecile.

You don't know how paralysing that is that stare of a blank canvas is, which says to the painter, 'You can't do a thing'.

The canvas has an idiotic stare and mesmerises some painters so much that they turn into idiots themselves.

Many painters are afraid in front of the blank canvas, but the blank canvas is afraid of the real, passionate painter who dares and who has broken the spell of "you can't" once and for all.

Thank you Dr Ali, Dr Yusri, and Prof Narazah; Ron, Yik Wei, Aisyah, Tasyriq, Diana, and Warren; Dr Shahrul Bariyah and everyone at Advanced Medical and Dental Institute (AMDI), USM; Majlis Amanah Rakyat (MARA); my fellow colleagues at AUCMS, Fakulti Sains Kesihatan UiTM (Pulau Pinang) Kampus Bertam, Malaysia Genome Institute (MGI), Institute of Bioscience, Fakulti Bioteknologi dan Sains Biomolekul, and Fakulti Perubatan dan Sains Kesihatan, Universiti Putra Malaysia; Asyraf, Papa and Mama; Thana and his wonderful coffee; and lastly to Cik Ummi Zalikha. Without your continuous support and blessings which fuel the passion for the pursuit of science in me, I would not be able to break the spell of "you can't" once and for all.

TABLE OF CONTENT

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

°C	degree Celcius
µ	micro
6His	hexahistidine
APS	ammonium persulfate
CAT	chloramphenicol acetyltransferase
CBP	calmodulin binding peptide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
HIV	human immunodeficiency virus
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilo-base pairs
kDa	kilo-Dalton
LB	Luria-Bertani
MBP	maltose-binding protein
MCS	multiple cloning site
MgCl2	magnesium chloride
NaCl	sodium chloride
NaOH	sodium hydroxide
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
RIL	arginine, isoleucine and leucine tRNA codons

SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEAP	secreted alkaline phosphatase
SUMO	small ubiquitin-like modifier
T/A	deoxythymidine/deoxyadenosine triphosphate
TAE	Tris-acetic acid-EDTA buffer
TBST	Tris-buffered saline with Tween 20
TE	Tris-EDTA buffer
tRNA	transfer RNA
UV-vis	ultraviolet-visible light
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
α	alpha
β	beta
Δ	delta

PEMBANGUNAN VEKTOR EKSPRESI DAN PENGKLONAN T/A YANG NOVEL UNTUK EKSPRESI PRODUK BIOFARMASEUTIKAL BERASASKAN PROTIEN YANG EFISIEN DI DALAM *Escherichia coli*

ABSTRAK

Escherichia coli merupakan sejenis bakteria yang sering digunakan untuk tujuan pengendalian, manipulasi and ekspresi gen sasaran. Ekspresi gen-gen tertentu secara efisien, terutama yg berasal dari mamalia seringkali terhalang kerana kekurangan biomolekul-biomolekul tertentu seperti tRNA, telah menyebabkan penghasilan protein berkualiti dari *E. coli* dalam kuantiti yang sedikit. Bagi menyelesaikan masalah ini, suatu vektor plasmid pembantu yang mengandungi gen tRNA yang jarang telah digunakan dalam *E. coli*. Proses penyelenggaraan dua vector plasmid yang berbeza: untuk penghasilan protein sasaran dan juga tRNA jarang juga boleh menyebabkan pertambahan pada beban metabolik pada *E. coli* dan mengakibatkan pertumbuhan *E. coli* terencat dan pengurangan dalam penghasilan protein sasaran. Bagi menangani masalah ini, kami pernah mengubahsuai suatu vektor plasmid bagi mengekspresi protein sasaran dan juga menghasilkan tRNA jarang secara serentak. Walaubagaimanapun, kehadiran tapak-tapak enzim endonuklese yang dikendaki di dalam gen sasaran telah menjejaskan proses pengeklonan gen-gen ini secara efektif. Untuk tujuan ini, kami telah mengubasuai suatu vektor plasmid yang mampu menghasilkan gen-gen tRNA jarang dengan modul pengeklonan T/A bagi memudahkan pengeklonan gen-gen sasaran tanpa menggunakan sebarang enzim endonuklase. Kerja-kerja pembangunan vektor plasmid yang novel ini bertujuan untuk melengkapkan vektor plasmid ini dengan: a) modul pengeklonan T/A bagi mempercepatkan proses pengeklonan gen bebas

enzim endonuklase, b) tiga gen tRNA jarang bagi meningkatkan penghasilan protein sasaran dengan lebih efisien, c) kawalatur penghasilan protein yang ketat, dan d) suatu gabungan hexahistidine tag pada C terminal protein sasaran bagi memudahkan pengasingan dan rekombinan protein yang tulen.

DEVELOPMENT OF A NOVEL T/A CLONING AND EXPRESSION VECTOR FOR EFFICIENT EXPRESSION OF PROTEIN-BASED BIOPHARMACEUTICALS IN *Escherichia coli*.

ABSTRACT

Escherichia coli is one of the most popular bacterial hosts for handling, manipulating, and expressing target genes of interest. Certain genes are difficult to express, often due to scarcity of specific tRNA. This has led to little and/or poor quality protein production in *E. coli.* To alleviate this problem, *E. coli* hosts are transformed with rare tRNA-expressing helper plasmid vectors. Maintenance of two different plasmid vectors-expressing the gene of interest and rare tRNA genes exert metabolic burden on the host that may result in diminished growth and reduced protein production. To address this issue, we have previously engineered an expression vector that simultaneously expresses rare tRNA genes and the gene of interest. Restriction enzymes are frequently used to clone desired genes into plasmid vectors. However, the presence of restriction enzymes–recognition sites within the desired gene would hamper the cloning process. To this end, we have engineered a rare tRNA vector with a T/A cloning module to facilitate restriction enzyme-independent cloning of genes. This work has resulted in the development of a novel expression vector that featured: a) T/A cloning module for rapid and restriction enzyme–independent cloning; b) three rare tRNA genes for efficient expression of therapeutic proteins; c) a tightly regulated and inducible promoter for controlled expression of the desired gene; and d) a C-terminal hexahistidine tag for rapid purification of recombinant protein.

CHAPTER 1: INTRODUCTION

1.1 DNA recombination technology is an integral tool for the protein research

For the past decades, protein research has survived many challenges. One of the biggest challenges faced by the protein research scientists is to amass kilogrammes of animal and plant tissues for extraction of various proteins of interest. Previously, this practice was a norm in order to meet the requirements for adequate, valid and reproducible data analysis(Fakruddin, Mohammad Mazumdar, Bin Mannan, Chowdhury, & Hossain, 2013; Rosano & Ceccarelli, 2014). These valuable proteins, oftentimes occur at a very low amount, even in their native sources, and the work to recover these proteins generally translate into labour-, time- and cost-inefficiency (Ferrer-Miralles, Domingo-Espín, Corchero, Vázquez, & Villaverde, 2009). Therefore, scientists are turning their focus onto alternative protein production system for a guaranteed unlimited supply, robustness, superior economic value and represent the certain desired biochemical or biological activity of the native protein (Duilio, Tutino, & Marino, 2004; Henry & Xiaoming, 2012; Kielkopf, Bauer, & Urbatsch, 2012; Nordlund et al., 2008; Palomares, Estrada-Mondaca, & Ramírez, 2004).

Synthetic heterologous production of the recombinant protein of interest was made possible with the debut of recombinant DNA technology (Fakruddin et al., 2013). The advent of genome sequencing technology has enabled scientist to identify the specific gene that encodes for their valuable proteins. Coupled with the sequencing technology, enormous DNA and protein sequencing data has been gathered. The information gathered is proponent in answering some of the most fundamental questions in biology, so a parallel high-throughput DNA recombinant

method development is increasingly in high demand (Meiering & Aravind, 2016). High-throughput DNA recombinant methods allowed the debut of numerous heterologous expression systems. These systems were developed to satisfy the need of various protein manipulations for identification and purification, enhanced and effective protein production, protein profiling analysis for biochemical, structural, mutation and functional studies (Schumann & Ferreira, 2004; Ventura & Villaverde, 2006). In turn, the availability of such feats propelled the realisation of these valuable protein into biotherapeutics development (Schumann & Ferreira, 2004).

The production of the recombinant protein is a multistep and straightforward process (Matsumura, 2015). The production of the desired recombinant protein begin by assembling a recombinant DNA. A piece of gene of interest is cloned into an expression vector plasmid, and the recombinant DNA product, in the form of plasmid, is transformed into expression host of choice, induced and the protein is ready for recovery, purification and characterization (Rosano & Ceccarelli, 2014). So, the DNA recombinant is considered as the central element for the heterologous protein production.

1.2 Common challenges in DNA recombination

The advancement of sequencing technology for the past decade in decoding various complex eukaryotic and prokaryotic genome has yielded an enormous amount of information (Reisinger, Kern, Fesko, & Schwab, 2007; C. Wang et al., 2013). However, assortment of genes that encodes for valuable proteins that have been identified by bioinformatics applications is yet to be studied (Reisinger et al., 2007). This is due to the suboptimal DNA recombination method that did not favour for multiple heterologous gene assembly in automation format.

The introduction of polymerase chain reaction (PCR), where the in vitro amplification of DNA, mediated by DNA polymerase and primed with synthetic oligonucleotide primers, in presence of deoxynucleotide triphosphates mix, amplifies the desired gene fragment from the DNA template. The PCR allowed easy and rapid isolation of desired genes of interest from the genome of their natural sources (Jeung et al., 2002). Thus, the debut of advanced sequencing technologies, coupled with PCR technology has helped propel the need for a rapid, easy and high throughput DNA recombination technology to manipulate and express various target genes of interest on a large scale.

The most challenging part of assembling a recombinant DNA is when the method is subjected to various individual optimisations in order to develop an efficient and effective DNA recombination protocol. Oftentimes, the optimisations are exclusive for a particular vector-target DNA construct and require method re-optimization for another set of DNA construct. Optimisations are subjected to experimental failures and troubleshooting which translates into countless repetitions (Matsumura, 2015). Not only that, the cost of pursuing research resources, especially

the technical labour, far exceeded the marginal variable costs of reagents used in molecular cloning (Matsumura, 2015; C. Wang et al., 2013).

The traditional method of DNA recombination relies on restriction enzyme digestion and ligation of DNA in order to assemble a recombinant DNA (Zimmermann et al., 1998). This restriction enzyme-mediated recombination is often limited by the presence of the restriction sites within both the gene of interest and the destination plasmid vector (C. Wang et al., 2013). Although this pitfall can be overcame by incorporation of unique and at times, exotic restriction enzyme recognition sites in both the target DNA and plasmid vector via PCR, this may prove expensive (Reisinger et al., 2007). However, this approach is only applicable if the desired restriction enzyme sites are absent within the amplified sequence (Jeung et al., 2002; Reisinger et al., 2007). Moreover, the selection of suitable restriction enzyme sites for recombination has become more limited with increased sizes of the amplified sequences (Gengyo-Ando, Yoshina, Inoue, & Mitani, 2006).

Even if the aforementioned pitfalls are embraced, the use of multiple enzymes for restriction digestion and the ligation of the both vector and target DNA can lead to increased time consumption and at experimental expenses (S. Chen, Songkumarn, Liu, & Wang, 2009). In addition, this method requires multiple purifications and/or enzymatic modifications steps in order to ensure a successful DNA recombination (Zhou & Gomez-Sanchez, 2000).

All these problems can potentially cause a bottleneck within the stream of protein research work and deny an effective and efficient workflow to manipulate and express various target genes of interest (C. Wang et al., 2013). Therefore, an easy, high throughput, rapid alternative DNA recombination technology is in need. This is because the high-throughput DNA recombination technologies, especially

those that favours automation are the answer to effective and efficient workflow for the study of various target genes of interest.

1.3 T/A cloning improves the DNA recombination methodology

The development of new cloning techniques are always driven by the attempts to improvise the existing restriction enzyme- and ligase-mediated recombination (Matsumura, 2015). This traditional method still remains relevant until today, as it is well-described, easily monitored and troubleshoot methodology. However, the fact that the traditional method is notoriously known to be resistant towards high-throughput settings (Berlec & Štrukelj, 2012; Goda, Tenno, Takasu, Hiroaki, & Shirakawa, 2004) it is relatively unsuitable for large-scale manipulation and expression of various target genes of interest.

Performing DNA recombination using PCR amplimers always remained a challenge in molecular biology labs. This is because, during the PCR amplification process, more than half of the amplimers have 3' A-tails, the by-product of template-independent manner of terminal deoxytransferase activity of Taq DNA polymerase enzyme (Jeung et al., 2002). The presence of an extra nucleotide at the termini of the PCR amplimers can hamper an efficient DNA recombination process. In blunt-ended DNA ligation, both vector and insert are required to be blunt-ended, thus presence of an extra nucleotide can post the risk of inefficient DNA recombination process (Dimov, 2012).

This phenomenon may render the blunt-ended DNA ligation technique inefficient (Dimov, 2012). A potential solution to this problem is to perform termini-polishing using a Klenow fragment or exonuclease enzyme prior to ligation.

However, this method may incur additional costs, labour intensive and time-consuming (Dimov, 2012).

This phenomenon has given birth to the development of 3'T-tailed linearized plasmid vector (T-vector) that is able to accept the 3'A-tailed PCR amplimers (Gengyo-Ando et al., 2006; Matsumura, 2015). This method, also known as the T/A cloning, was first described by Holton and Graham in 1991, taps into the terminal deoxynucleotidyl transferase activity of Taq DNA polymerase to create a 3'A-overhang onto the PCR amplimers and can easily be ligated with complementary T-vectors without the need of any modifications (Green, M.R. and Sambrook, 2012; Holton & Graham, 1991; Matsumura, 2015). Figure 1.1 illustrates the DNA recombination using T/A cloning method.

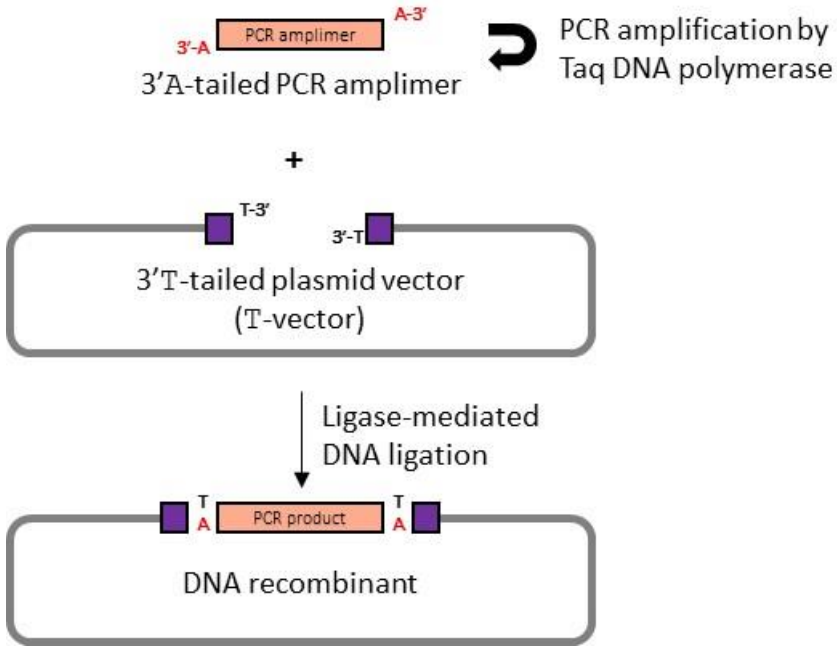

Figure 1.1 DNA recombination using T/A cloning method.
The T/A cloning method requires a linear 3'T-tailed plasmid vector (T-vector) to be aligned with a compatible 3'A-tailed gene fragment. The terminal deoxynucleotidyl transferase activity of a Taq DNA polymerase enzyme during PCR amplification caused a non-templated addition of a single deoxyadenosine triphosphate (dATP) onto the 3' termini of the PCR amplimers. The 3'A-tailed PCR amplimers can easily be cloned into a T-vector without the need of any restriction enzyme digestion and/or termini modifications. The DNA nick at the T/A junctions can be ligated by using a DNA ligase and propagated in an *E. coli* host.

The T/A cloning method is by far, the simplest and most efficient DNA recombination method that allows direct cloning of PCR amplimers without the use of restriction enzymes, designer recombinase enzymes or topoisomerases for site-specific recombination (C. Wang et al., 2013). Any plasmid vector can be made into a T-vector regardless of the existing restriction enzyme sites, plus direct cloning of PCR amplimers made the T/A cloning a suitable candidate to be developed into a high-throughput setting.

Another advantage of the T/A cloning method is this method can be adapted for various downstream DNA recombination applications, not exclusively for modification and expression of target genes. For example, the T/A cloning method has been appreciated to study various genes and gene regulatory components (Fujimori, 2009; Gengyo-Ando et al., 2006; Yu et al., 2014), novel strategy for site-directed mutagenesis (Adachi & Fukuhara, 2012), transient eukaryotic expression and reporter vector (C. Wang et al., 2013), adapted to create a hybrid Gateway-T/A cloning method (Miles & Verkade, 2014), in vivo RNAi gene functional assay (Kamath & Ahringer, 2003), and many more.

In comparison to other alternative recombination methods, T/A cloning is superior to Gateway and Creator cloning systems because the latter require multistep recombination processes. Although both Gateway and Creator cloning systems are independent of restriction enzymes and ligases, the multistep recombination is time- and labour-intensive, their designer recombinases are expensive and require long and deliberate primer design which often complicates the gene cloning and has proven to be cost-inefficient in a high-throughput application (C. Wang et al., 2013).

The ligation independent cloning method (LIC), however, suffers from long primers (C. Wang et al., 2013), enzymatic processing of DNA termini (Zhou &

Gomez-Sanchez, 2000) and requires an intermediate plasmid before being cloned into a destination vector (Reisinger et al., 2007).

The T/A cloning technology, alike any other alternative cloning methods, also comes with its own limitations and requires a deeper understanding of its nucleic acid biochemistry and proper optimisations in order to be effective and efficient (Matsumura, 2015). Moreover, this technique relies entirely upon the terminal deoxynucleotidyl transferase activity of Taq DNA polymerase and/or terminal transferase enzyme and subsequent ligation (Matsumura, 2015). So suboptimal enzymatic condition of these enzymes would greatly hamper efficient and effective DNA recombination via T/A cloning.

1.4 Challenges in T/A cloning methodology and potential solutions

Alike any other DNA recombination method, a selection system is essential to identify individual clones which harbour the gene recombinant from the self-ligating plasmid. Although theoretically self-ligation of T-vector is impossible due to incompatible termini, such phenomena can happen (due to the loss of either or both overhanging 'T') and the presence of self-ligating vector plasmid would hinder an efficient T/A DNA recombination (Dimov, 2012).

Successful DNA recombination can be screened by α-complementation method. This method of screening for positive clones required association of two inactive fragments of *E. coli* β-galactosidase to form a functional enzyme. The desired gene usually cloned into the plasmid multiple cloning sites caused the production of the amino acid terminal fragment which is incapable of forming a functional enzyme. So only bacterial colonies which harbour a functional β-galactosidase will appear blue in the presence of chromogenic X-Gal (Green & Sambrook, 2012a). Ideally, only white colonies should harbour the recombinant DNA.

The α-complementation method is only applicable for cloning desired gene in the form of plasmid DNA. This is because of the gene regulatory elements of an expression plasmid could be hampered by the presence of the native gene regulatory elements in α-complementation. So, this method is only limited to cloning of the desired gene. Furthermore, the α-complementation method of screening DNA recombinant is not completely infallible. The outcome of the α-complementation can often be affected by the size of the insert, mutational changes within the plasmid (Green & Sambrook, 2012a), and culture environment (C. Wang et al., 2013).

The introduction of lethal gene *ccdB* cassette into the cloning site of T-vector as stuffer would alleviate this problem (C. Wang et al., 2013). The desired gene of

interest is supposed to replace this gene during DNA recombination assembly. This gene is widely used as a potent negative selection marker for Gateway cloning, whereby the clones with plasmid DNA vector which harbours the *ccdB* gene will not survive leaving only clones possessing the desired DNA recombinant to thrive. However, the use of *ccdB* is not economical because only special strains can be used to propagate the T-vector harbouring the *ccdB* gene as a stuffer.

The success of T/A cloning depends upon the generation of the T-vector. There are various methods described to prepare the T-vector including T-tailing by terminal deoxytransferase activity of Taq DNA polymerase and/or terminal transferase enzymes or restriction endonuclease treatment (Reisinger et al., 2007). The earlier method requires linearization of the vector, termini-polishing and addition of 3'T-tails by Taq DNA polymerase or terminal transferase enzyme with the presence of excess dTTPs (Dimov, 2012; Goda et al., 2004; Jeung et al., 2002).

However, this method is considered to be less efficient than the latter due to the low efficiency of the terminal transferase enzyme to include a single dTTP onto the 3' ends of the linearized plasmid (Jeung et al., 2002). Moreover, it is difficult to monitor and regulate the polymerase and exonuclease activities. The multistep process of incorporating the 3'T-overhangs is often labour- and time-inefficient especially for a standalone lab (Dimov, 2012).

Alternatively, the T-vector can be synthesised by restriction digestion processing which undeniably yields a single 3'T-tail (Dimov, 2012). This method typically requires an inverse PCR amplification to linearize and flank the cloning site with desired restriction enzyme. A careful primer design is important to ensure the generation of a 3'T-tail after the restriction enzyme processing and a correct frame is maintained.

The *Xcm*I enzyme is one of the regular restriction enzymes of choice for the generation of 3'T-tails. The restriction processing generates a single 3' overhang within its recognition sequence 5'-CCANNNNN/NNNNTGG-3' (where N represents either A, T, C, or G). The idea is to incorporate two *Xcm*I restriction enzyme sites flanking the plasmid cloning site or flanking a stuffer fragment., The T-vector is generated by simply restricting the plasmid vector with *Xcm*I restriction enzyme (Dimov, 2012).

However, *Xcm*I often suffers from partial cleavage resulting in a relatively high background of empty plasmid when cloning and requires introducing a *Xcm*I-containing cassette which limits the flexibility in terms of site of the PCR product integration and requires plasmid modification. (Berlec & Štrukelj, 2012)

Alternatively, *Ahd*I restriction enzyme can also be used to generate the 3'T-overhangs. *Ahd*I is often a preferred restriction enzyme for this purpose due to its superiority in restriction enzyme efficiency compared to *Xcm*I (Dimov, 2012; Reisinger et al., 2007). Moreover, its shorter restriction enzyme recognition sequence made it less complicated to be included in the vector via PCR amplification (Dimov, 2012).

The only downside of adopting *Ahd*I restriction enzyme to generate the T-vector is this enzyme is not suitable for cloning vectors of variants and descendants of the pUC family due to the presence of secondary *Ahd*I site within the Amp^R gene. This problem can be easily resolved by introducing a silent mutation onto the secondary *Ahd*I site via site-directed mutagenesis (Adachi & Fukuhara, 2012; Dimov, 2012; Jeung et al., 2002; Stevenson & Brown, 2015). Other available enzymes that can be used to create 3' overhangs include *Asp*EI (Goda et al., 2004; Jeung et al., 2002),

*Hph*I (Jeung et al., 2002; C. Wang et al., 2013), Eam1105I (Goda et al., 2004), *Asp*EI and *Mbo*II (Jeung et al., 2002), *Bci*VI, *Bfi*I, *Mnl*I and *Taa*I (C. Wang et al., 2013).

The most critical downside of the T/A cloning is that it is unsuitable for cloning and construction of expression library due to its bi-directional gene insert cloning (Goda et al., 2004). The basis of T/A cloning method is in the formation of a single, bilateral 3'overhangs and this caused the population of the hybridised intermediates of vector-insert complexes to be equal in both directions. Although Goda et al (2004) has successfully developed an asymmetrical *Ahd*I whereby correctly oriented insert will generate *Nco*I/*Nde*I restriction enzyme sites within the *Ahd*I site, T/A cloning is still insufficient to be adapted for construction of libraries due to inability to enrich the correctly oriented ligated target gene.

Due to its bidirectional cloning nature of the T/A cloning method, this cloning strategy is often adopted for recombination of the target gene fragment into a cloning plasmid rather than into an expression plasmid vector. We observed that most literature published in 2017 explores the applicability of the T/A cloning method within the vicinity of cloning and propagation of the target DNA, compared to cloning the target DNA into T/A-adapted expression vectors (D. J. Lee, Lee, Jang, Ferrandon, & Lee, 2017; K.-L. Lee et al., 2017; Mattes & Escalante-Semerena, 2016; Ota et al., 2017; Shi et al., 2017; Zhang et al., 2016). There are a handful of T/A and T/A-variant cloning and expression kits available in the market, with impressive claims that each product is faster, cheaper and more efficient, but there is no guarantee that the technique or enzyme-catalysed reaction is 100% efficient or specific especially when placed under suboptimal conditions. Therefore, it is imperative for any method to be optimised empirically and a proper monitoring at

each stage of DNA recombination to ensure an uninterrupted workflow in protein research studies.

1.5 Problem statement

The simplicity of the T/A cloning method, and adaptability of the 3'T-overhangs cloning site onto any plasmid vector are the two strong points that made this cloning strategy among the most adopted cloning method.

Despite the versatility and simplicity of the T/A cloning method, the aforementioned pitfalls have hindered the potential of the T/A cloning method to be adapted into a high-throughput system. Here, we reported a number of improvements to the existing T/A cloning methodology and by adapting our T/A cloning technology onto our pSA-RIL expression plasmid.

The pSA-H*Nef*-C6His-RIL plasmid was previously developed by our colleague to improve the expression of HIV-1 Nef protein which formerly was expressed at very low concentrations in soluble form in *E. coli* expression system (Mualif et al., 2015). We have identified that the presence of codons rarely used by *E. coli* was the reason behind low yields of soluble HIV-1 Nef protein. To address this issue, we included a selection of rare tRNA genes within the HIV-1 Nef expression plasmid. The co-expression of an array of tRNA genes (*arg*U, *ile*Y and *leu*W) significantly improved the yields of soluble HIV-1 Nef protein.

Realising the potential of this expression plasmid for the expression of genes rich in codons rarely used by *E. coli*, we replaced the HIV-1 *Nef* gene with a multiple cloning site (MCS) polylinker to create pSA-MCS-RIL plasmid in view to make the process of desired heterologous gene recombinant easier. However, the MCS contains a limited number of restriction enzyme sites. To overcome this problem, a deliberate primer design must be made to include a new, non-existing restriction enzyme site onto the target genes for in-frame cloning into the pSA-MCS-RIL plasmid vector. The incorporation of the new restriction enzyme sites to the target

gene and/or plasmid vector and the subsequent restriction enzyme digestion and purification further complicates the cloning process, and may potentially hamper an efficient and high-throughput gene recombination.

Thereby, we attempted to improve the cloning methodology of the pSA-RIL plasmid by incorporating our T/A cloning module to replace the existing MCS. Our T/A cloning module was specifically designed to allow (1) directional cloning of PCR amplimers by the incorporation of asymmetrical *Ahd*I sites and (2) low false-positive plasmid background by incorporating the lethal *sacB* gene.

This thesis describes:

1. The engineering of the pSATA-RIL cloning/expression plasmid vector and its plasmid vector derivatives.
2. Evaluation of the pSATA-RIL cloning/expression plasmid vector by expressing *sacB* and *CAT* genes.

This is a proof-of-concept study that demonstrates an improved cloning and expression method of pSATA-RIL plasmid from the pSA-MCS-RIL plasmid. The findings described in this thesis are the provisional assessment for an improved T/A cloning methodology.

1.6 Objective statement

Based on the works of literature cited above, we hypothesised that the incorporation of our T/A cloning module onto pSA-RIL vector would improve the heterologous gene cloning methodology for protein expression in *E. coli* system.

The objectives of this study are:

1. To design a T/A cloning module to the cloning site of the pSA-RIL plasmid and incorporate a lethal *sacB* gene as a stuffer gene into creating pSATA-*sacB*-RIL cloning/expression plasmid vector
2. To evaluate the pSATA-RIL cloning/expression plasmid vector by expressing *sacB* and *CAT* genes.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals/reagents

Refer to Appendix A for the list of Chemicals/reagents used in this study.

2.1.2 Consumables

Refer to Appendix B for the list of Consumables used in this study.

2.1.3 Culture media

Refer to Appendix C for culture media used in this study.

2.1.4 General buffers, stock solutions, antibiotics

Refer to Appendix D for general buffers, stock solutions and antibiotics.

2.1.5 Bacterial strains

Refer to Appendix E for the list of bacterial strains.

2.1.6 Plasmids

Refer to Appendix F for the list of plasmids used in this study.

2.1.7 Primers

Refer to Appendix G for the list of primers used in this study.

2.1.8 General instruments

Refer to Appendix H for the list of general instruments used in this study.

2.1.9 DNA restriction digestion enzymes

Refer to Appendix I for the list of DNA restriction digestion enxymes used in this study.

2.1.10 DNA ligation

Refer to Appendix J for the protocol for DNA ligation.

2.2 Experimental Strategy

The present study was divided into three parts: (1) the construction of a T/A cloning and expression plasmid vector, (2) the optimisation of the site-directed mutagenesis method and the sequences of the spacer spanned between the Shine-Dalgarno (SD) and the ATG start codon, and finally (3) the cloning and expression of levansucrase and chloramphenicol acetyltransferase enzyme. Figure 2.1 illustrates the outline process workflow for the present study.

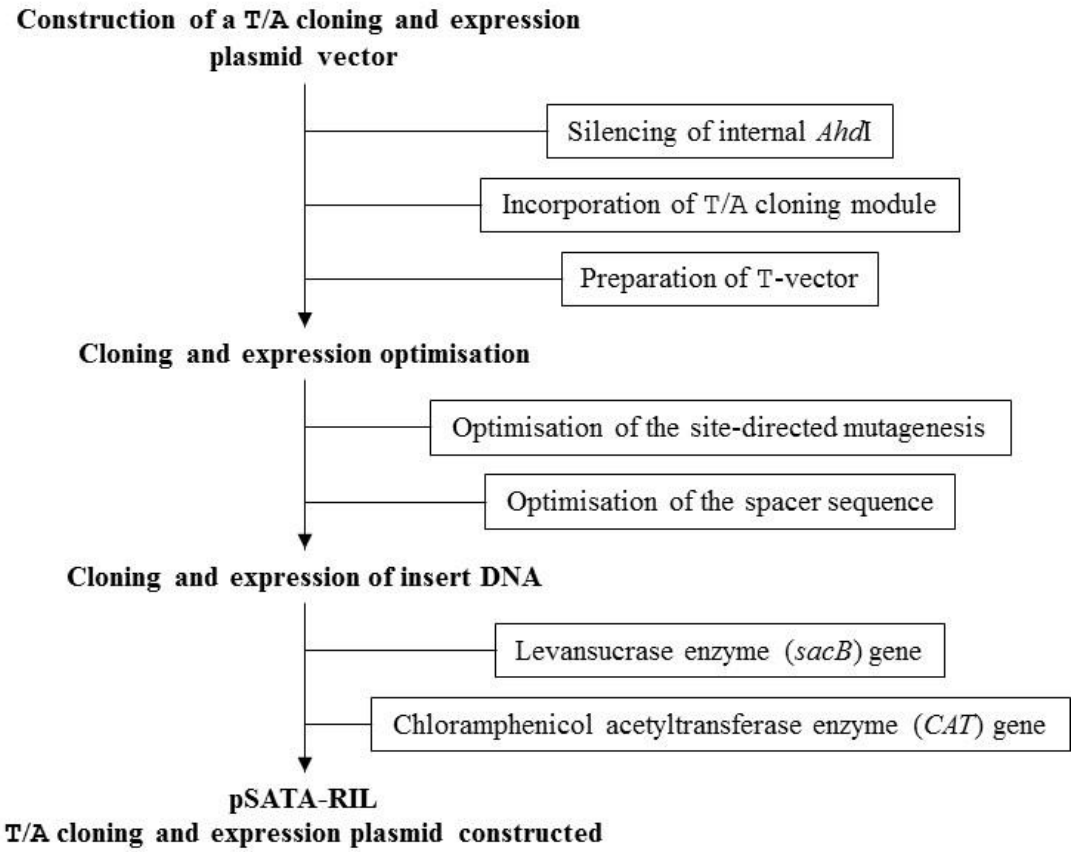

Figure 2.1 Outline process workflow for the present study.

2.3 Methods

2.3.1 Bacterial strains and culture conditions

In this present study, we utilised *E. coli* strain NEB 5-alpha for plasmid propagation, strain NEB T7 Express for expression of cloned insert DNA, and strain Agilent Technologies XL10-Gold ultracompetent cells for mutagenic plasmid propagation. All cells were maintained on LB agar without any antibiotic and routinely propagated in LB broth without any antibiotic supplementations. The selection of clones with various vectors was done in LB medium supplemented with ampicillin (100 μg/mL).

All bacterial strains were maintained on LB agar with appropriate supplements and temporarily stored at 4℃. Subculturing was done once every 14 days.

2.3.2 General primer design

The primers used in this study was designed based on the sequence of the target gene. *In silico* design of the primers was done by using Vector NTI software and the primer DNA sequences were sent to Biobasic Inc. Canada for the oligonucleotide synthesis. The primers were synthesised using Perkin-Elmer DNA synthesiser. The primers were purified by polyacrylamide gel electrophoresis (PAGE) at 50 nm scale.

2.3.3 Polymerase Chain Reaction (PCR) amplification

The PCR amplification of the target genes requires a basic PCR reaction mix and thermocycler setting. For initial PCR amplification, we adhered to the recommended protocols by the suppliers and further modification of the reaction

components and/or thermocycler settings were introduced as necessary. A typical 1× PCR master mix for a total reaction of 50 µL in a PCR tube was prepared as below:

A 25 µL of 5× KAPA HotStart Buffer (to final 1×) was added to a PCR tube. 25 mM $MgCl_2$ was added to final 1.5 mM, 10 mM dNTP mix was added at final 0.2 mM of each dNTP. 10 mM of 'sense' and 'antisense' primers was added to final 0.5 µM each. 5 U/µL KAPA Taq DNA polymerase was added into the tube to final 1 U. The DNA template was added as required to final 200 ng and pyrogen-free water was added to final volume of 50 µL.

The components were prepared on ice and vortexed briefly. The KAPA Taq DNA polymerase was added in the last step. The PCR tube was placed on a thermocycler pre-set to 95˚C. A typical PCR thermocycler setting was set as follows:

The initial denaturation was set to 95°C for 5 minutes. Then a 30-temperature cycle was set: denaturation at 95°C for 30 seconds, annealing 50-56 (touchdown) or Tm-5°C for 30 seconds and extension at 72°C for 1 min per kb of amplimers. The final extension was performed at 72°C for 1 min per kb amplimers.

We categorised the PCR reactions performed in this study based on the expected size of the amplimer, and the requirement for proofreading properties of DNA polymerase enzyme.

2.3.3(a) Insert PCR

This PCR reaction is used to amplify our desired target DNA (insert) for assembly of DNA recombinant. The typical size for this PCR reaction is between 500 bp to 2 kb. We used KAPA HiFi HotStart DNA polymerase which was supplied in KAPA HiFi HotStart ReadyMix (2×) (Kapa Biosystems #KK2601) to catalyse the reaction. A typical insert PCR reaction is as follows:

A 25 µL 2× KAPA HiFi HotStart Ready Mix was added into a PCR tube to a final 1×. An equimolar 10µM 'sense' and 'antisense' primers was added to final 0.3 µM each. 200 ng template DNA was added and finally pyrogen-free water was added to final volume of 50 µL.

All the reaction mix was prepared on ice and vortexed briefly. The PCR tube was placed on a thermocycler pre-set to 95˚C. The PCR thermocycler setting was set as indicated in 2.3.3 Polymerase Chain Reaction (PCR) amplification.

2.3.3(b) Whole plasmid (inverse) PCR

This PCR reaction amplifies a segment of DNA between the two outward-pointing primers. In this study, we performed the whole plasmid (inverse) PCR to flank our desired plasmid with the sequence of interest while maintaining the exactly same nucleotide sequence with the template plasmid. The typical size of this template is around 6 kb. Thus, a thermostable DNA polymerase with 3' to 5' exonuclease (proofreading) activity with the lowest error rate was required to catalyse this reaction. We used Q5® DNA polymerase enzyme (NEB #M0491). A typical inverse PCR reaction is as below:

In a 50 µL reaction, 25 µL 5× Q5 Reaction buffer was added to final 1×, 10mM dNTP mix was added to final 0.2 mM each, 10 µM 'sense' and 'antisense' primers was added to final 0.5 µM each. 200 ng DNA template was added and 2 U/µL Q5 High-Fidelity DNA polymerase was also added to final 1 U. The pyrogen-free water was added to final volume 50 µL.

All the reaction mix was prepared on ice and vortexed briefly before adding the Q5 High-Fidelity DNA polymerase. The PCR tube was placed on a thermocycler pre-set to 98˚C. The PCR thermocycler setting was set as follows:

The initial denaturation was set to 98°C for 30 seconds. Then a 25-temperature cycle was set: denaturation at 95°C for 30 seconds, annealing 50-56 (touchdown) or Tm-5°C for 30 seconds and extension at 72°C for 1 min per kb of amplimers. The final extension was performed at 72°C for 1 min per kb amplimers.

2.3.3(c) Colony PCR

A colony PCR is used to screen for the desired DNA recombinant from colonies of transformed *E. coli* strain.

2.3.3(c)(i) Preparation of bacterial lysate

Randomly selected transformant colonies with diameter around 2-3 mm was selected for the colony PCR screening. The single colony was aseptically picked with a micropipette tip and inoculated into 10 µL of pyrogen-free water in a sterile PCR tube. For identification purposes, the tip was streaked onto a patching grid plate. The PCR tube was placed onto a heating block at 95˚C for about 2 minutes and centrifuged at 4,000×g for one minute. The supernatant was taken as a template for the colony PCR.

2.3.3(c)(ii) Colony PCR

This colony PCR is similar to insert PCR in terms of amplimer size but does not require any proofreading activity. We used KAPA Taq DNA polymerase which was supplied in 2× KAPA Taq ReadyMix with dye (Kapa Biosystems #KK1024). A typical colony PCR reaction was prepared as follows:

A 25 µL 2× KAPA Taq Ready Mix was added into a PCR tube to a final 1×. An equimolar 10µM ‘sense’ and ‘antisense’ primers was added to final 0.2 µM each.

200 ng template DNA was added and finally pyrogen-free water was added to final volume of 50 µL.

All the reaction mix was prepared on ice and vortexed briefly. The PCR tube was placed on a thermocycler pre-set to 95˚C. The PCR thermocycler setting was set as indicated in 2.3.3 Polymerase Chain Reaction (PCR) amplification.

2.3.4 Visualisation and quantification of DNA

2.3.4(a) Agarose gel electrophoresis

We used the agarose gel electrophoresis to resolve the DNA according to its size and to inspect, analyse and purify DNA fragments.

2.3.4(a)(i) Preparation of agarose gel and sample loading

A 0.7% (w/v) agarose gel was prepared by weighing 0.7 g agarose (BIO 1000-500g) and added into 100 mL of 1× TAE buffer in a conical flask. The mixture was heated at medium/high power in a microwave oven for about 2 minutes until the agarose dissolved. The mixture was swirled gently to make sure there is no agarose clump visible before leaving to cool to about 45˚C. The tip of a sterile micropipette tip was used to dip into ethidium bromide (10 mg/mL) before added to the mixture. The flask was swirled gently to mix and poured into a pre-fixed plastic gel cast fitted with an appropriate gel comb. The bubbles were carefully removed by using a sterile micropipette tip and the gel was allowed to solidify at room temperature for 30 minutes.