IDENTIFICATION OF *parS* IN MEGAPLASMID pHG1 OF *Cupriavidus necator* H16

by

POW KAH CHEE

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

October 2018

ACKNOWLEDGEMENT

I would like to acknowledge the guidance and support of Prof. Razip, as the main supervisor. He has been patient, understanding and supportive. I am deeply indebted to him. I am also grateful for the guidance and support of Dr. Kamarul. I would also like to acknowledge the financial support of MyBrain15 (MyMaster) scholarship. Lastly, I would like to express my gratitude to fellow members of LAB 414, my friends and family in supporting me throughout my degree.

TABLE OF CONTENTS

ACKNOWLEDGEMENT		**ii**
TABLE OF CONTENTS		**iii**
LIST OF TABLES		**vi**
LIST OF FIGURES		**vii**
LIST OF SYMBOLS AND ABBREVIATIONS		**ix**
ABSTRAK		**xi**
ABSTRACT		**xiii**
CHAPTER 1 – INTRODUCTION		**1**
1.1	Objectives	3
CHAPTER 2 – LITERATURE REVIEW		**4**
2.1	Plasmid	4
2.2	The Significances of Plasmid Technology	7
	2.2.1 Significance 1: A storage of genetic information	9
	2.2.2 Significance 2: A tool for gene expression	10
	2.2.3 Significance 3: An agent to a system	11
	2.2.4 Significance 4: A supportive role	11
2.3	Plasmid Replication Origin	12
	2.3.1 Plasmid Replication Mechanisms	13
2.4	Plasmid Segregational Stability	16
2.5	Active Partitioning System	19
2.6	Type Ia active partitioning system	28
	2.6.1 ParA – Motor Protein	28
	2.6.2 ParB – Centromere-binding Protein (CBP)	30
	2.6.3 *parS* – Centromere	32
2.7	Plasmid Incompatibility	34
2.8	*Cupriavidus necator* H16, the PHA bioplastic producer	35
	2.8.1 The megaplasmid pHG1	36
	2.8.2 PHA biosynthesis of *C. necator* H16	37
	2.8.3 The *phaCAB* operon	37
2.9	Genetic engineering study on *Cupriavidus necator* H16	38

CHAPTER 3 – METHODOLOGY 41

3.1 General Methods 41

3.1.1 Bacterial strains, plasmids and media 41

3.1.2 Chemicals and reagents 41

3.1.3 Sterilization 41

3.2 Bioinformatic analysis 44

3.2.1 BLAST alignment 44

3.2.2 Prediction of promoter region 44

3.2.3 Motif search on protein sequence 44

3.2.4 Prediction of palindromic sequence 44

3.3 DNA modification methods 45

3.3.1 Outline of molecular cloning 45

3.3.2 DNA purification 45

3.3.3 Polymerase chain reaction (PCR) 46

3.3.4 Extraction and purification of pHG1 megaplasmid 46

3.3.5 Plasmid extraction by alkaline lysis 49

3.3.6 Agarose gel electrophoresis 50

3.3.7 Preparation of Inoue method's competent cell. 50

3.3.8 Heat-shock chemically transformation of cell 51

3.3.9 Plasmids constructions 52

3.4 Expression of H_6ParB tagged protein 53

3.4.1 Cold-shock protein expression of H_6ParB 53

3.4.2 Immobilised metal affinity chromatography (IMAC) purification of H_6ParB 53

3.4.3 SDS-PAGE 54

3.5 *parS*-ParB binding assay 56

CHAPTER 4 – RESULT & DISCUSSION 57

4.1 Prediction of *parS* position and sequence by bioinformatics analysis 57

4.1.1 Prediction of *parS* locus 57

4.1.2 Prediction of *parS* sequence 62

4.2 Probing of *parS* by immobilisation 66

4.2.1 Cloning and overexpression of hexahistidine-tagged ParB (H_6ParB) 68

4.2.2 Preparation of *parS* candidates and non-*parS* negative controls 71

4.2.3 *In vitro* binding of *parS* by ParB 74

CHAPTER 5 – OVERALL DISCUSSION **80**

5.1 A possible new consensus representative of *parS* 80

5.2 The primary and secondary *parS* 83

5.3 The potential of PAR region as the compartment for the construction of stable vector. 85

CHAPTER 6 – CONCLUSION **88**

CHAPTER 7 – REFERENCE **90**

APPENDIX

LIST OF TABLES

		Page
Table 2.1	The distinctive characteristics of plasmid.	24
Table 3.1	The list of bacterial strains.	42
Table 3.2	The list of plasmids.	42
Table 3.3	The list of primers	43
Table 3.4	The optimised parameters of different reactions of PCR.	48
Table 3.5	The formulation of resolving gel and stacking gel of SDS-PAGE.	55
Table 4.1	Type Ia ParB homologs of active partitioning system identified with cognate parS centromere.	63
Table 4.2	Palindromic parS candidates predicted from region PAR.	65

LIST OF FIGURES

		Page
Figure 2.1	The distinctive characteristics of plasmid.	5
Figure 2.2	The typical elements featured in plasmid vector.	8
Figure 2.3	The three major replication mechanisms of circular plasmid.	14
Figure 2.4	The general properties of active partitioning system.	22
Figure 3.1	The general thermocycling conditions of a PCR.	47
Figure 4.1	The structure and organisation of origin-proximal region and the elements of ParABS system of pHG1.	58
Figure 4.2	Graphical view of motifs identified in ParA and ParB of pHG1.	61
Figure 4.3	Position of *parS* candidates in PAR region.	65
Figure 4.4	The principle and procedure of *parS*-ParB binding assay.	67
Figure 4.5	The plasmid map of pColdparB.	69
Figure 4.6	Protein expression and purification of H6ParB.	70
Figure 4.7	The plasmid map of pTEM.	72

Figure 4.8	The plasmid maps of pTEM-Pst1 and pBBR1parAB.	73
Figure 4.9	Gel profile of the *parS* candidates and negative controls.	75
Figure 4.10	Result of *parS*-ParB binding assays. ParB binding activity detected in PAR and ORI..	77

LIST OF SYMBOLS AND ABBREVIATIONS

°C	degree Celsius
∞	infinity
θ	theta
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	base pair
CBP	centromere-binding protein
ChIP-seq	chromatin immunoprecipitation sequencing
CRISPR	clustered regularly interspaced palindromic repeats
CV	column volume
ddH_2O	deionised distilled water
dH_2O	distilled water
DNA	deoxyribonucleic acid
EMSA	electromobility shift assay
g	relative centrifugal force (G-force)
HTH	helix-turn-helix
IHF	integrated host factor
IMAC	immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobase pair
kDa	kilodalton
LB	lysogeny broth
KDPG-aldolase	2-keto-3-deoxy-6-phosphogluconate aldolase
LDF	linear discriminative function
Mb	megabase pair
MCS	multiple cloning site
MWCO	molecular weight cut-off

Ni-NTA	nickel-nitrilotriacetic acid
NR broth	nutrient rich broth
nsDNA	non-specific DNA
NTP	nucleoside triphosphate
OD	optical density
ORF	open reading frame
ori	replication origin
P(3HB)	poly(3-hydroxybutyrate)
par system	active partitioning system
PCR	polymerase chain reaction
PHA	polyhydroxyalkanoate
RE	restriction enzyme
RHH	ribbon-helix-helix
RNA	ribonucleic acid
rpm	revolution per minute
SCP	single-cell protein
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacryamide gel electrophoresis
SOC	super optimal broth (SOB) with glucose added
T_a	annealing temperature
v/v	volume per volume
w/v	weight per volume
wt%	weight percent

IDENTIFIKASI *parS* DALAM MEGAPLASMID pHG1

***Cupriavidus necator* H16**

ABSTRAK

Sistem aktif pembahagian (sistem ParABS) yang menyelaraskan migrasi plasmid replika ke dua kutub yang berlainan mengandungi tiga elemen, ParA, ParB dan *parS*. Dalam megaplasmid pHG1 *Cupriavidus necator* H16, ParA dan ParB telah dianotasi manakala *parS* yang berinteraksi dengan ParB masih tak diketahui. Kajian ini bertujuan untuk mengenal pasti posisi *parS* dalam megaplasmid pHG1. Analisis bioinformatik menunjukkan ParB mengandungi motif HTH terabadi yang merupakan ciri determinatif sistem aktif pembahagian jenis Ia. Hal ini telah membawa kepada hipotesis *parS* pHG1 terletak di hilir operon *par*, yang mana dua palindrom, Pld_1 dan Pld_2, telah dikenali sebagai calon *parS*. Namun, jujukan palindrom-palindrom itu ataupun rantau yang ditakrif sebagai PAR (yang mengandungi operon *par*) tidak menunjukkan persamaan yang signifikan kepada kelapan-lapan konsensus *parS* ketika dijajarkan. Ini menandakan kemungkinan *parS* pHG1 sebagai konsensus *parS* yang baharu yang mewakili kumpulan sistem ParABS yang belum diketahui. Selain dari rantau PAR, *parS* juga dihipotesiskan terletak di dalam rantau ORI (huluan rantau PAR). Untuk menyokong hipotesis ini, assai ikatan *parS*-ParB telah dikendali untuk mengesan aktiviti ikatan ParB dalam rantau PAR dan rantau ORI. Aktiviti ikatan ParB telah dikesan dalam kedua-dua rantau pada keafinan ikatan yang berbeza. Keputusan ini mencadangkan rantau PAR berikat dengan ParB dengan keafinan yang lebih tinggi berbanding dengan rantau ORI. Oleh sebab perbezaan ini, kajian ini mencadangkan satu *parS* primer terletak di dalam rantau PAR manakala satu *parS* sekunder terletak di dalam rantau

ORI. Pada masa depan, kajian boleh memfokuskan untuk memencilkan sistem ParABS pHG1 yang berfungsi dan mengenalpastikan jujukan *parS*.

IDENTIFICATION OF *parS* IN MEGAPLASMID pHG1 OF *Cupriavidus necator* H16

ABSTRACT

The active partitioning system (ParABS system) that coordinates the migration of plasmid replicons to two separate poles is consist of three elements, ParA, ParB and *parS*. In the megaplasmid pHG1 of *Cupriavidus necator* H16, the ParA and ParB had been annotated while the ParB-binding *parS* remained unknown. This study aims to identify the position of *parS* in the megaplasmid pHG1. Bioinformatic analysis shows that ParB has a conserved HTH motif which is a determinative characteristic of type Ia active partitioning system. This led to the hypothesis that the *parS* of pHG1 is located downstream of the *par* operon where two palindromes, Pld_1 and Pld_2, were identified and determined as the probable *parS* candidates. However, the sequences of the palindromes as well as the sequence of defined PAR region (consisting *par* operon) showed no significant similarity to any of the eight *parS* consensus when aligned. This suggest that pHG1 *parS* could be a new *parS* consensus representing an unknown subset of ParABS system. Aside of the PAR region, another *parS* was also hypothesised to reside in ORI region (the immediate upstream of PAR region). To support this hypothesis, a *parS*-ParB binding assay was conducted to detect the ParB binding activity in region PAR and region ORI. ParB binding activity was detected in both regions but at different binding affinity. PAR region was shown to bind with ParB at a comparatively higher magnitude as compared with the ORI region. Due to this difference, this study proposes that a primary *parS* is located in the PAR region while a secondary *parS* is located in the

ORI region. In future, studies can be focused on isolating a functional pHG1 ParABS system and identify the sequence of *parS*.

CHAPTER 1 - INTRODUCTION

Plasmid is a replicative DNA molecule that is often found in a bacterial cell as a part of the cell's genetic element. Unlike chromosomal DNA, plasmid is smaller and it encodes only non-essential genetic features (Summers, 2009). In other words, plasmid is susceptible to be lost by the host cell.

In genetic engineering technology, plasmids play an important role as vectors to assist in the manipulation of the genome entity of a cell (Friehs, 2004; Phillips and Funnell, 2004; Summers, 2009). Since then, plasmid vectors have become the popular choice among other genetic engineering tools for the implementation of genetic engineering in both areas of research and industrial application (Sambrook *et al.*, 1989; Summers, 2009). To overcome the instability of plasmid, an antibiotic is often used as a selective marker to sustain the plasmid-harbouring population (Sambrook *et al.*, 1989; Sato *et al.*, 2013). This practice is however impractical at the scale of industrial application as the cost of the antibiotic will spike the production cost to an unprofitable level (Sato *et al.*, 2013; Gruber *et al.*, 2014). To overcome the problem, one of the approach is to incorporate a self-encoded plasmid stabilising system into the plasmid vector (Nordstrom and Austin, 1989).

Active partitioning system is a commonly employed strategy of low-copy-number plasmid to secure the plasmid's inheritance over the host's life cycle (Summers, 2009). The active partitioning system maintains the plasmid stability by coordinating the plasmid replicons to separate poles (Summers, 2009). The mechanism involved two proteins and one DNA-binding site which interact with each other. This process can be generalised into three simple steps (Gerdes *et al.*, 2000; Gerdes *et al.*, 2010). First, centromere-binding protein recognises the centromere (DNA-binding site) that is present on the target plasmid and forms the

nucleoprotein–partition complex. In the second step, the two partition complexes are then paired up to recruit for motor proteins. Finally, in the third step, the motor proteins serve to drive the two partition complexes apart in a nucleoside triphosphate (NTP)-dependent manner.

This study aims to identify *parS* of active partitioning system (also known as ParABS system) of megaplasmid pHG1 of *Cupriavidus necator* H16. *C. necator* H16 is chosen because it is the most popular industrial strain used to synthesise biodegradable plastic – polyhydroxyalkanoate (PHA) (Schlegel *et al.*, 1961; Reinecke and Steinbuchel, 2009). Genetic engineering technology had been practiced on *C. necator* H16 to manipulate PHA production in the strain with respect to variety, quantity and quality of the synthesised PHA (Reinecke and Steinbuchel, 2009). Thus, understanding the ParABS partitioning system may yield useful knowledge to upgrade the genetic engineering toolbox for application at the industrial-scale.

Based on the genome studies of Schwartz *et al.* (2003) and Pohlmann *et al.* (2006), *C. necator* H16 genome consists of two chromosomes and one megaplasmid. The megaplasmid, namely pHG1, had been annotated to encode three protein homologs of the active partitioning system, namely ParA, ParB and ParB2 (Schwartz *et al.*, 2003). Another essential component of the ParABS system, the *parS* (DNA-binding site), was not annotated in the study, and it remained unknown.

Megaplasmid pHG1 is believed to encode the type I active partitioning system (Schwartz *et al.*, 2003). This study envisions that the pHG1 ParABS system could be an ideal plasmid-stabilising system for the construction of selection-free and stable plasmid vector, customised for *C. necator* H16. This is because the pHG1

ParABS system is a native system of *C. necator* H16 and therefore, its performance would have been optimised in the native host. In order to achieve this vision, more study on *parS* is required. Thus, the objectives of this study are outlined below.

1.1 Objectives

The objectives of this study are:

1. To predict and identify *parS* of pHG1 by bioinformatic approach and *parS*-ParB binding assay.

CHAPTER 2 - LITERATURE REVIEW

2.1 Plasmid

Plasmid is an extra-chromosomal element that encodes genetic information of a bacterium (Phillips and Funnell, 2004; Summers, 2009). The structure of a plasmid is often depicted as a circular and double-stranded DNA molecule, but there are also plasmids that are linear in form (Kinashi *et al.*, 1987). In nature, plasmid is a common companion that build up the prokaryote's genome. It bears the role as chromosomal DNA; representing the genetic element of the host; encodes and regulates gene expression (Summers, 2009). The reason that plasmid is distinctive from the chromosome is because plasmid does not encode genetic element that is essential for the host (Summers, 2009). The genes found in plasmid are generally limited to be beneficial instead of essential (Summers, 2009). For instance, plasmid could encode genes for antibiotic resistance, virulence factor and detoxification and degradation of organic compound (Thomas and Summers, 2001). Nevertheless, the advantages does not eliminate the risk of plasmid being uninherited (Summers, 2009).

Apart from that, there are other differences between plasmid and chromosome as illustrated in Figure 2.1. Firstly, the size of a plasmid is usually smaller than the chromosome (Summers, 2009). The range of the plasmid size is broad, ranging from a few thousand basepairs to several hundred thousand basepairs (Thomas and Summers, 2001; Summers, 2009). The size is usually inversely proportional to the copy-number of plasmid, to balance the metabolic burden of the plasmid to the cell (del Solar *et al.*, 1998; Summers, 2009). Unlike chromosome, plasmid copy-number

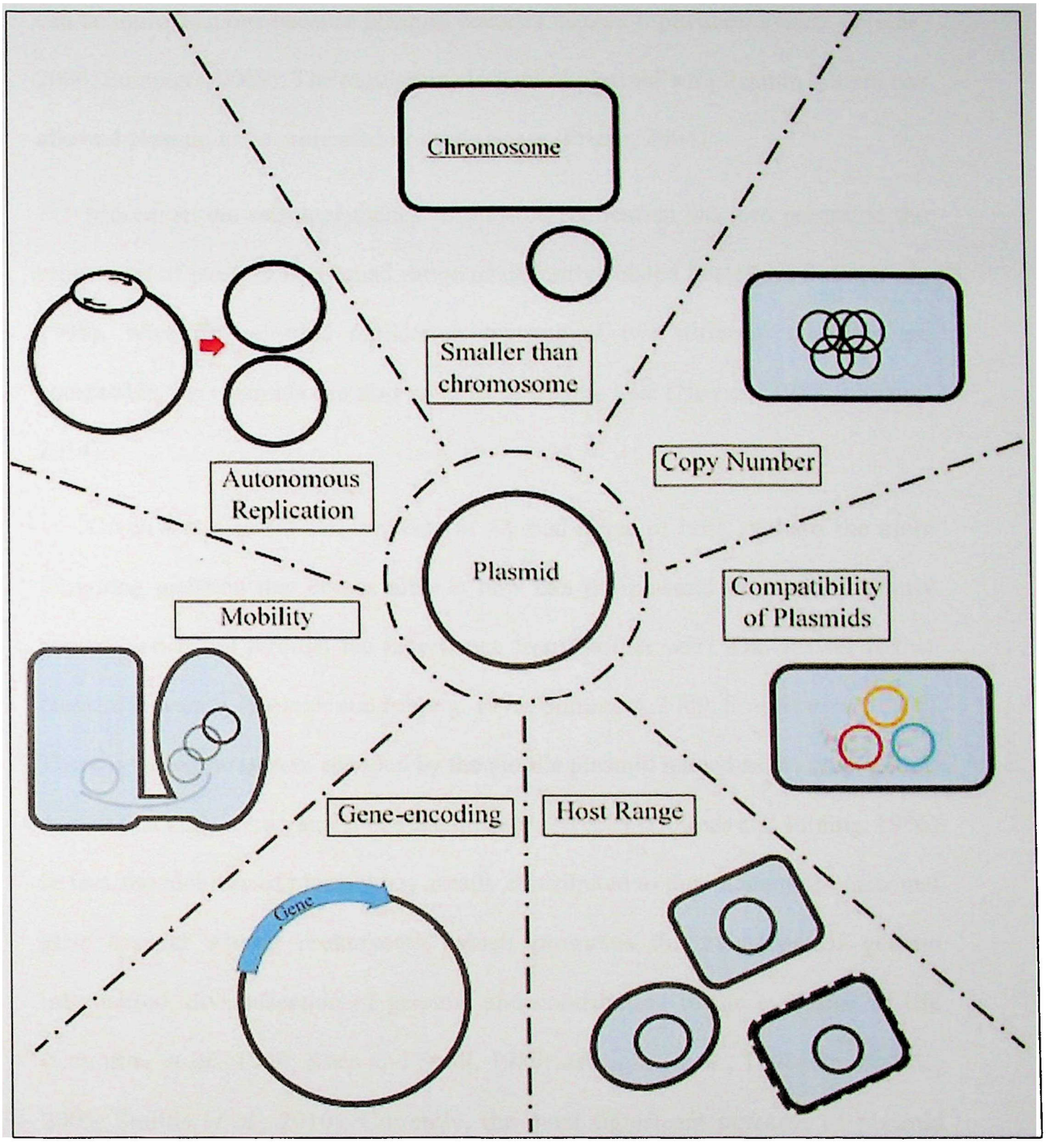


Figure 2.1 The distinctive characteristics of plasmid.

The figure summarised the information obtained from the book published by Summers (2009).

can be more than one because plasmid encodes its own replication system (Friehs, 2004; Summers, 2009). The regulation element of plasmid's replication system has allowed plasmid to be replicated multiple times (Friehs, 2004).

Moreover, the self-dependency of plasmid replication has also permitted the replication of plasmid in a broad range of distantly-related hosts (del Solar *et al.*, 1998). When the plasmid replication systems of two different plasmids are compatible, the plasmids can also co-exist in a same host (Novick, 1987; Thomas, 2014).

Given that plasmid can replicate in a broad range of host, perhaps the more intriguing question that comes after is how can the plasmid be in the distantly related host if not through the inheritance from mother cell? The answer is that plasmid is mobile. (Holmes and Jobling, 1996; Summers, 2009; Smillie *et al.*, 2010). There is a specific system encoded by the mobile plasmid named as the conjugation system that allows the transfer of plasmid between cells (Holmes and Jobling, 1996). In fact, the mobility of plasmid has greatly contributed to the incident of horizontal gene transfer among prokaryotes, which promotes the exchange of genetic information, diversification of genome and contributing to the evolution of life (Couturier *et al.*, 1988; Kües and Stahl, 1989; del Solar *et al.*, 1998; Frost *et al.*, 2005; Smillie *et al.*, 2010). Currently, the most significant outcome of plasmid mobility is perhaps the rapid emergence of multiple drug resistance organisms (Alanis, 2005; Birge, 2013; Chang *et al.*, 2015).

To summarise, plasmid is the extra piece of DNA adopted by the host cell with given independency to enable high frequency of lateral exchange of genetic information as a strategy to increase the host competency in surviving the

environment. Since its discovery, plasmid technology had been developed by harnessing the strengths of plasmid, which had contributed tremendously to molecular biology research and biotechnology (Sambrook *et al.*, 1989; Friehs, 2004).

2.2 The Significances of Plasmid Technology

In the field of molecular biology, the application of plasmid is very diverse and its contribution is immeasurable (Sambrook *et al.*, 1989; Friehs, 2004). Plasmid is a powerful tool because it is an excellent tool for the manipulation of DNA. As DNA is the first of the central dogma of biology, the changes of DNA can thus have the deepest influence to a biological system (Sambrook *et al.*, 1989; Friehs, 2004). Compared to other tools, the distinctive characteristics of plasmid (Figure 2.1) have made it a strategic tool to store and to express DNA (Sambrook *et al.*, 1989).

To commercialise the use of plasmid as a genetic tool box, the plasmid is customized and compacted to carry only the necessary and useful elements (Sambrook *et al.*, 1989). This synthetic plasmid is named as a plasmid vector. The plasmid vector is readily fused with another DNA-of-interest through a process called cloning (Sambrook *et al.*, 1989).

In figure 2.2, a typical circular plasmid vector is depicted together with the elements that are commonly incorporated into the plasmid vector. First, the replication origin is the vital element needed for the autonomous replication and copy-number regulation of plasmid (del Solar *et al.*, 1998; Friehs, 2004). The replication origin also determines the plasmid host range and plasmid incompatibility (Novick, 1987; del Solar *et al.*, 1998; Jain and Srivastava, 2013).

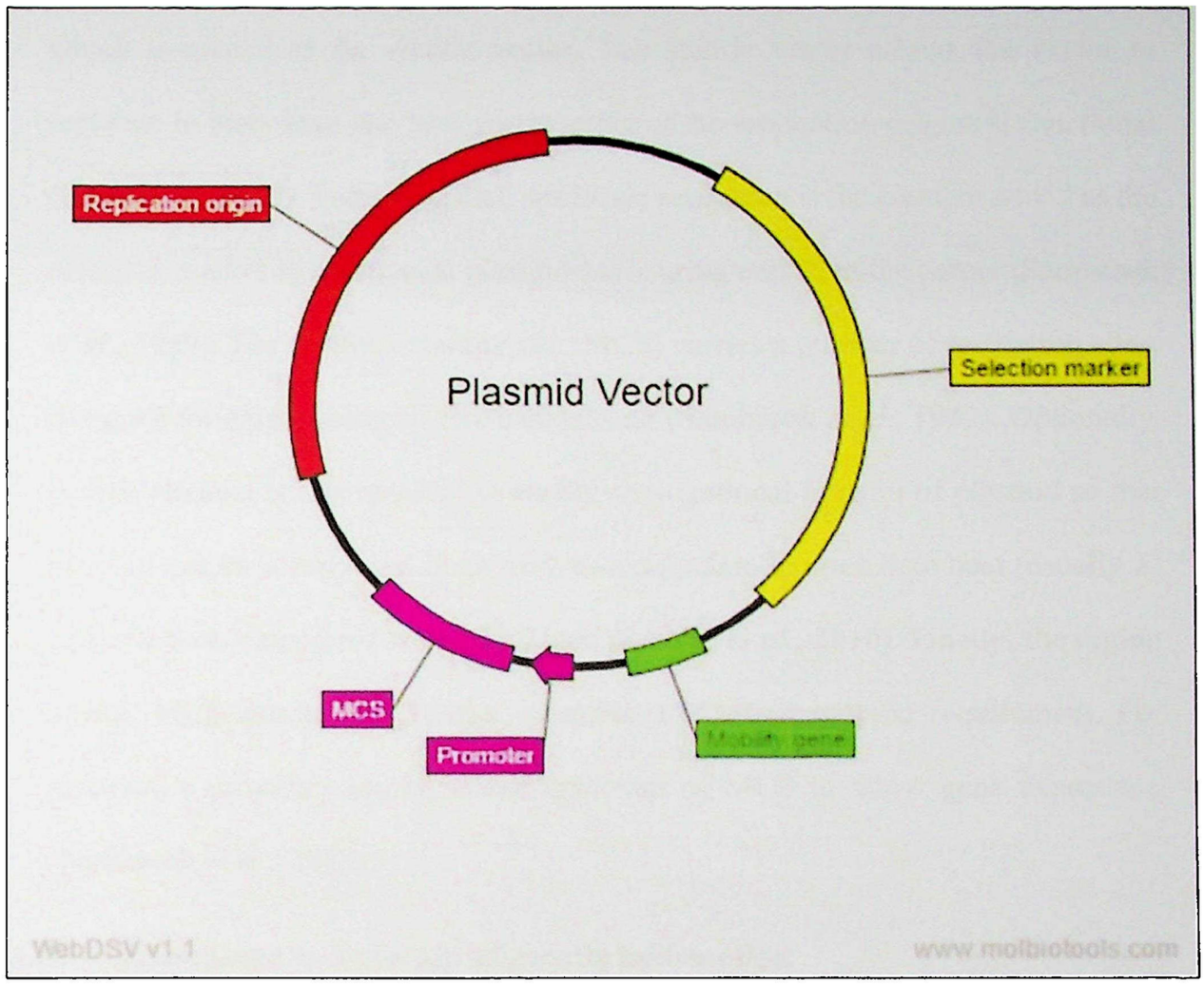


Figure 2.2 The typical elements featured in plasmid vector.

MCS stands for multiple cloning sites which is a section where many restriction sites are located.

Sometimes, a plasmid vector is designed with two different replication origins, which is named as the shuttle vector. The shuttle vector allows the vector to replicate in more than one host where either of the replication origins is functional (Rose *et al.*, 1987). Following that, antibiotic resistance is the element added as the selective marker to distinguish plasmid-harbouring cell from the others (Sambrook *et al.*, 1989). The multiple cloning site (MCS) carries a number of restriction sites, designed for easy cloning of DNA-of-interest (Sambrook *et al.*, 1989). Optionally, mobile element is incorporated to enable conjugational transfer of plasmid so that plasmid can be constructed in an easy-to-manipulate intermediate host (usually *E. coli*)and then transferred to the final host (Smillie *et al.*, 2010). Finally, the region nearby MCS can also be further customised to attain specific requirement. For instance, a promoter can be added upstream of MCS to allow gene expression (Sambrook *et al.*, 1989).

2.2.1 Significance 1: A storage of genetic information

Among the diverse roles of plasmid vector, the most fundamental role of plasmid is perhaps to store genetic information (Sambrook *et al.*, 1989). Cloning the DNA insert into a plasmid helps to maintain the quality and quantity of the DNA insert (Sambrook *et al.*, 1989). For example, plasmids can act as the intermediate library for the sequencing of the genome of a living organism (Lodish *et al.*, 2000). High quantity and purity is a crucial requirement to accomplish whole-genome sequencing. Crude genomic DNA is a complex mixture of sheared DNA fragments with poor homogeneity.

Thus, plasmid vector act as the intermediate agent to adopt the DNA fragments separately and transformed into a cell to reproduce more identical replicons for sequencing purpose. Apart from that, the unknown chromosome can easily be

sequenced when cloned into a plasmid because most plasmids carry optimised primer set for sequencing (Sambrook *et al.*, 1989; Lodish *et al.*, 2000; Birge, 2013). In short, plasmid can serve as the library of the genetic information encoded by the cloned DNA.

2.2.2 Significance 2: A tool for gene expression

Plasmid can also be used as a tool to express the cloned genetic element (Sambrook *et al.*, 1989; Friehs, 2004; Birge, 2013). The open reading frame (ORF) cloned into a vector can be expressed to protein. There are several advantages to using plasmid as an expression vector. One of the advantage is the expression level of protein encoded by plasmid can be higher than chromosome-encoded protein due to the difference of copy number (Sambrook *et al.*, 1989). The second advantage, plasmid enable heterologous expression of foreign-origin protein in a common expression host like *Escherichia coli* (Miroux and Walker, 1996). Many expression vectors have also adapted fusion protein technology which help to yield pure protein from crude sample (Sambrook *et al.*, 1989; Birge, 2013). This is achieved by pre-designing the vectors with extra codons around the MCS. For example, an in-frame cloning of *parB* downstream to the codons of a polyhistidine tag resulted in the expression of polyhistidine-tagged ParB (Mohl and Gober, 1997). This protein can be easily purified by immobilised-metal affinity chromatography (IMAC) (Hochuli *et al.*, 1988). The third advantage is the use of plasmid as the template for protein study because it greatly simplifies the mutagenesis of proteins (Ho *et al.*, 1989; Sambrook *et al.*, 1989; Packer and Liu, 2015). Mutations can easily be introduced into the plasmid-encoded protein simply by inverse PCR. The outcome is easily obtained with high accuracy and low error due to the small-size of plasmid. Thus, plasmids are popular tools to express and study unknown proteins.

2.2.3 Significance 3: An agent to a system

Plasmid can also act as an agent to introduce a new system to a host or to manipulate a system in a host (Birge, 2013). Here, the plasmid is valued not solely by the proteins expressed, but the system built up by the proteins. Perhaps the most significance contribution of this aspect is to the fermentation technology (Imanaka and Aiba, 1981; Friehs, 2004). Using polyhydroxyalkanoates (PHA) biosynthesis as an example, introducing a heterologous *phaC* (PHA synthase) gene into *Cupriavidus necator* H16 resulted in the production of a heteropolymer which the wild type host could not have achieved on its own (Sato *et al.*, 2013). Likewise, transferring the PHA biosynthesis system (*phaCAB* operon) to *Escherichia coli* enable the biosynthesis of PHA in *E. coli* resulting into a phenotype that has not been recorded in the wild type (Tan et al., 2010).

2.2.4 Significance 4: A supportive role

Last but not least, plasmid sometimes might not take part as the main player but act supportively. As an example, genome editing technology make use of plasmid as a toolbox to carry out the necessary elements for genome editing (Blomfield *et al.*, 1991; Skorupski and Taylor, 1996; Philippe *et al.*, 2004). As an example, genome editing is achieved by introducing a mutated homologous sequence into the targeted host via a suicide plasmid, to stimulate homologous recombination, resulting in an allelic exchange (Philippe *et al.*, 2004). It is called a suicide plasmid because the plasmid is designed to be unable to replicate in the recipient host. Therefore, the plasmid serves only as a delivery agent to deliver the mutated homologous sequence, which is then exchanged with the host (Skorupski and Taylor, 1996; Philippe *et al.*, 2004). Another example of the plasmid

application is the newly developed CRISPR technology. CRISPR stands for clustered regularly interspaced palindromic repeats (Jinek *et al.*, 2012; Ledford, 2015). In the study of Jiang *et al.* (2013), they were able to deliver the genome editing service through a plasmid-based system. Two plasmids were used to express the Cas9 protein, tracrRNA and crRNA to edit the host's genome. Both examples (suicide plasmid and CRISPR) use plasmid as the carrier of the necessary tools or elements and require the plasmid only for a temporary moment inside the host. Once the purpose is achieved, the plasmid is removed by plasmid curing.

In conclusion, plasmid is an important tool that have greatly contributed to the development and research of biotechnology. Serving as a genetic element, plasmid can thus influence the host activity down from the most basic level and up to the complex biological system. Therefore, the research and development of plasmid technology should always be continuously developed to gain more valuable knowledge in order to develop new tools to tackle new frontiers in biological knowledge.

2.3 Plasmid Replication Origin

Replication origin (*ori*) is the term used to describe a *cis*-acting element in a replicon that is essential to initiate the autonomous replication (del Solar *et al.*, 1998). In other words, a plasmid constructed with only a replication origin can actively and independently replicate. Typical plasmids harbour only one functional replication origin, similar to the bacterial chromosome (Takeuchi and Nakazawa, 2001). The plasmid replication origin is a minimal region consisting of genes and noncoding DNAs. The genes encode the necessary proteins such as the replication initiator protein while the noncoding DNA serves as the binding sites of replication

initiator protein and other regulatory elements (Kües and Stahl, 1989; Sambrook *et al.*, 1989).

Note that the independency of plasmid replication is not really absolute (Kües and Stahl, 1989; Jain and Srivastava, 2013). The replication does rely on several chromosome-encoded proteins, such as DnaA and integrated host factor (IHF) (Kües and Stahl, 1989; Zhong *et al.*, 2005). These proteins are ubiquitous and thus plasmids are able to replicate in a range of hosts (del Solar *et al.*, 1998; Zhong *et al.*, 2005; Jain and Srivastava, 2013).

2.3.1 Plasmid Replication Mechanisms

The plasmid replication can be initiated by a few mechanisms. The three major types of replication mechanisms (Figure 2.3) of circular plasmid are theta mode of replication, strand displacement and rolling circle replication mechanisms (Kües and Stahl, 1989; del Solar *et al.*, 1998). Generally, the initiation of replication and the synthesis of a new strand, will first require the binding of a DNA polymerase to a region of primer-bound DNA.

Briefly, the DNA polymerase catalyses the formation of phosphodiester bond between the 3'-OH tail of primer and the 5'-phosphate of the adding nucleotide, and thus the extension of a new strand of DNA (del Solar *et al.*, 1998). To achieve that, the theta mode and strand displacement replication mechanisms utilize RNA priming to synthesize complementary RNA primer; while the rolling circle

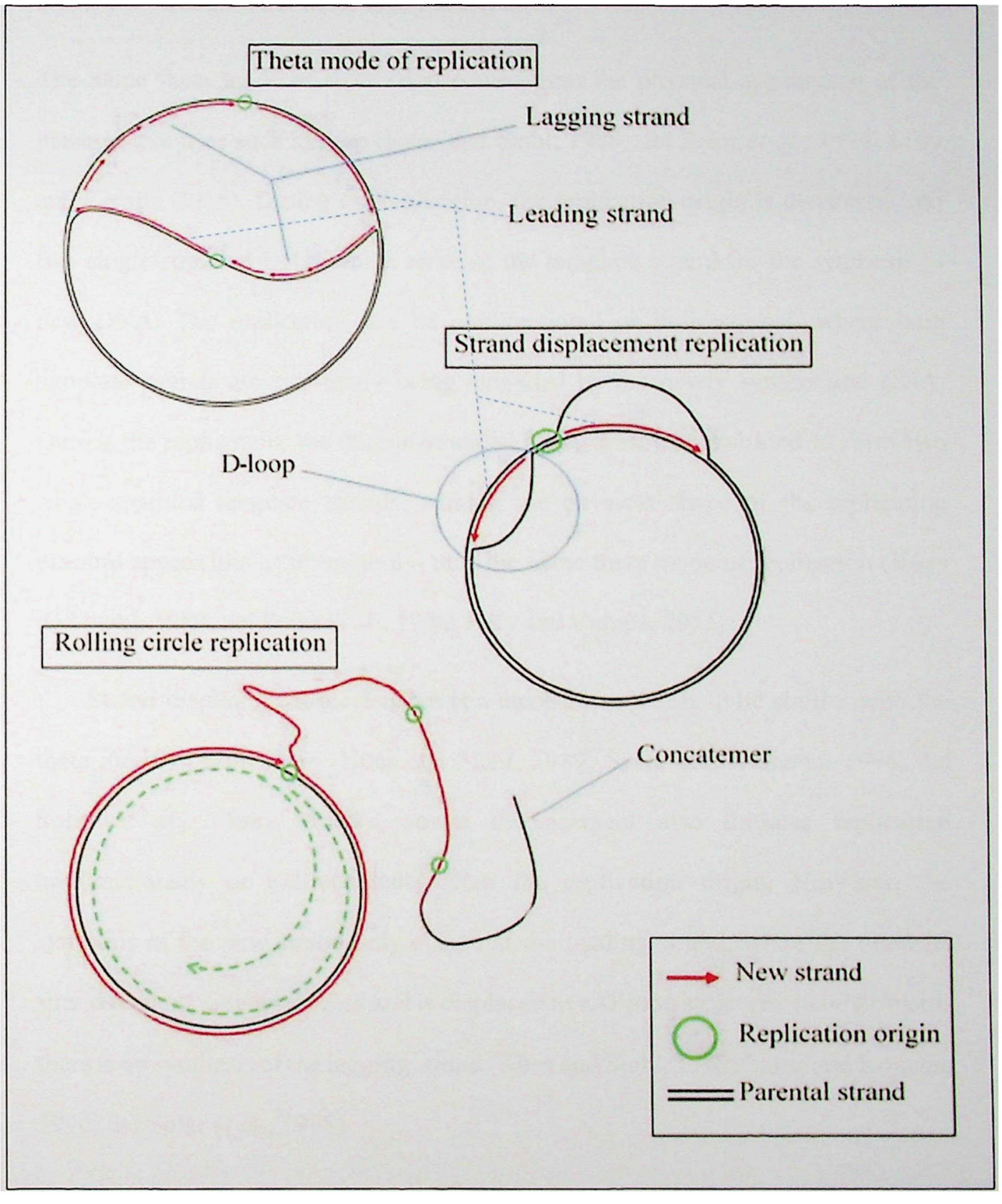


Figure 2.3 The three major replication mechanisms of circular plasmid.

replication utilizes nicking to break the existing parental strand of lagging strand to create a 3'-OH end (del Solar *et al.*, 1998).

The name theta mode of replication comes from the physical appearance of the plasmid that uses such system (Kües and Stahl, 1989; del Solar *et al.*, 1998; Lilly and Camps, 2015). During the replication, the replication origin is denatured into two single-stranded DNA which serve as the template strand for the synthesis of new DNA. The replication can be unidirectional or bidirectional, where both template strands are constantly being extended by the newly synthesized DNA. During the replication, the double-stranded DNA is slowly unfolded to form two single-stranded template strands, making the physical shape of the replicating plasmid appear like a "θ" symbol – thus the name theta mode of replication (Kües and Stahl, 1989; del Solar *et al.*, 1998; Lilly and Camps, 2015).

Strand displacement mechanism is a mechanism that is quite similar with the theta mode of replication (Kües and Stahl, 1989; Sakai and Komano, 1996; del Solar *et al.*, 1998). Briefly, strand displacement also initiates replication unidirectionally or bidirectionally from the replication origin. However, the synthesis of the new strand only occurs at the leading strand, while the opposite strand remains single stranded and is displaced as a D-loop structure. In other words, there is no synthesis of the lagging strand (Kües and Stahl, 1989; Sakai and Komano, 1996; del Solar *et al.*, 1998).

Rolling cycle replication mechanism works in a totally different way in which, the leading strands are synthesized continuously from the same template strand (del Solar *et al.*, 1998; Khan, 2005; Ruiz-Masó *et al.*, 2015). The replication starts by nicking the leading strand by the replication initiator protein which recognizes a

hairpin formation. The polymerase then utilizes the 3'-OH at the nicked point to incorporate nucleotides and synthesize the new leading strand in a unidirectional fashion. The intruding newly synthesized leading strand thus displace the parental strand. The single-stranded parental strand is stabilized by single-stranded DNA binding (SSB) protein. In the replication process, the polymerase goes cycle by cycle on the same template strand and thus produce a concatemer of the leading strand. The concatemer is then nicked to produce multiple strands of leading strand which is then ligated and circularized. Only then, the synthesis of complementary strand is initiated to produce the double-stranded plasmid replicon (del Solar *et al.*, 1998; Actis *et al.*, 1999; Khan, 2005; Ruiz-Masó *et al.*, 2015).

2.4 Plasmid Segregational Stability

The term plasmid segregational stability or plasmid stability refers to the heredity of plasmid in the life cycle of a host. In other words, the capability of a plasmid to be faithfully inherited by the daughter cells over cycles and cycles of life cycle (Nordstrom and Austin, 1989). In a host cell, plasmid loss is rarely due to the structural instability but rather due to the inefficient distribution of daughter plasmid replicons. Thus, the loss of a plasmid in a population is generally due to segregational instability (Friehs, 2004). Nevertheless, the genetic elements harboured by the plasmid would still have an indirect impact that promotes plasmid-free population. Under unbiased condition, plasmid-free cells would usually have a faster growth rate than the plasmid-harbouring cells (Bentley and Kompala, 1989; Corchero and Villaverde, 1998; Lau *et al.*, 2013). The faster-growing plasmid-free cells would consequently dominate the population. The growth rate of plasmid-harbouring cells is presumably slower because the existence of plasmid itself is an added metabolic burden to the host (Bentley and Kompala, 1989; Bentley *et al.*,

1990). The degree of burden increases further by the loading imposed by expressing the content coded by the plasmid (Bentley and Kompala, 1989). The environmental factor, medium composition for instance, should also be considered as the influencing factor on plasmid stability. Since the metabolism activity of cell is influenced by its environment, the expression of plasmid's genetic elements is thus affected by the environment changes (Gupta *et al.*, 1995).

When the plasmid replicons within the cell is randomly distributed, the chance of successful inheritance of the plasmid to both daughter cells increase proportionally with the increase of plasmid copy number (Bentley and Kompala, 1989). Theoretically, this implies that a low copy number plasmid will have a very low stability. For example, there is only a 50% of chance for the plasmid of two to be inherited by both two daughter cells. However, most low-copy-number plasmid contradicts this expectation regarding the plasmid stability. In fact, many of them are highly stable in the host's life cycle. This is due to the fact that the low-copy-number plasmids have employed a systematic approach to maintain the vertical transfer of plasmids (Friehs, 2004).

Basically, there are three different systems whose role is to maintain plasmid segregational stability. These systems are active partitioning, post-segregational killing and multimer resolution systems (Dmowski and Jagura-Burdzy, 2013). In short, the active partitioning system ensures plasmid inheritance by locating plasmid replicons at both pole regions of the cell (Summers, 2009). Since active partitioning system is the subject of this study, a more specific and detailed review is written in the subsequent chapters (Section 2.5).

Post-segregational killing system is a cruel system that kills the daughter cell which has failed to inherit any plasmid (Friehs, 2004). It can also be termed as the plasmid addiction system (Kroll *et al.*, 2010). Briefly, the system encodes two elements, a stable toxin and an unstable antidote. For example, the *hok/sok* system of R1 plasmid consist of the toxin-encoding *hok* gene and antisense-RNA-transcribing *sok* gene (Gerdes, 1988). The antisense RNA serves as the antidote of the *hok* protein and acts by inhibiting the translation of *hok* mRNA through the complementary binding at a region of the *hok* mRNA. Since the antisense RNA is more susceptible to degradation than the mRNA, the toxic *hok* protein can be expressed in the plasmid-free cells whereas the *sok* RNA had long been degraded. Consequently, the plasmid-free daughter cells are killed by the toxin proteins translated from the *hok* mRNA (Gerdes, 1988; Franch *et al.*, 1997).

Metabolism-based plasmid addiction system is another member of the plasmid-addiction system, derived from the concept of post-segregational killing system (Kroll *et al.*, 2010). It is created by human ingenuity consisting of two partners – an auxotrophic host and its auxotrophic-complementing plasmid. As a result, the survival of the recombinant is tied with the presence of the plasmid occurrence. This strategy was used on *C. necator*, which utilized KDPG-aldolase (2-keto-3-deoxy-6-phosphogluconate aldolase) addiction system to preserve plasmid heredity up to 93% of the cell population over the cultivation period (Voss and Steinbüchel, 2006).

The multimer resolution system is a system that acts supportively as compared to previous systems, in the sense that it increases the probability of faithful plasmid inheritance (Nordstrom and Austin, 1989). This system enhances plasmid stability by resolving plasmid multimer to its monomers and promote the spreading of

plasmids. The result of multimerization can influence plasmid stability since the clustering of plasmid replicons will reduce the efficiency of plasmid spreading. As demonstrated by the study of ColE1 plasmid, a loss of multimer resolution system had led to the consequence of 10^{-2}-10^{-5} plasmid loss per generation (Summers and Sherratt, 1984).

Another system that has indirectly contributed in maintaining plasmid segregational stability is the conjugation system of plasmid. As mentioned, conjugation system allows the horizontal transfer of plasmid and therefore, can transfer plasmid to the daughter cell that have not inherited any plasmid (Holmes and Jobling, 1996). For instance, megaplasmid pHG1 of *C. necator* H16 is a self-transmissible plasmid (Friedrich *et al.*, 1981). The same study has shown that pHG1 can be conjugated back to a pHG1-cured mutant and other closely-related species, implying the ability of mobile plasmid to remain circulating in the host population even when some daughter cells failed to inherit the plasmid.

In conclusion, nature has developed several mechanisms to help sustain the plasmid prevalence in the host's life cycle. While most plasmids adopted more than one of these systems to maximise the efficiency in maintaining plasmid stability (Nordstrom and Austin, 1989).

2.5 Active Partitioning System

The active partitioning (*par*) system is a true partitioning system that coordinates the movement of the replicated plasmids to the daughter-cell-to-be poles so that both daughter cells will inherit the plasmid (Summers, 2009).

The discovery of active partitioning system began from an observation of the unexpectedly high segregational stability of some low-copy-number plasmids

(Jacob *et al.*, 1963; Rosner, 1972). The observation suggested that the plasmid replicons are systematically migrated during cell replication. At the time, P1 plasmid was known as a plasmid with the lowest plasmid-copy-number and showed high plasmid stability characteristics (Rosner, 1972). Over a series of experiments involving the isolation of the relevant genetic elements, mutation, deletion and complementation study, the ParABS system of P1 plasmid was discovered in 1982 and later characterised to harbour two *trans*-acting proteins, ParA and ParB, and a *cis*-acting binding-site, *parS* (Austin *et al.*, 1982; Austin and Abeles, 1983; Austin and Abeles, 1983; Abeles *et al.*, 1985).

Focusing on the studies related to the identification of P1 *parS*, the first approach to identify *parS* of P1 ParABS system was via an incompatibility test (Abeles *et al.*, 1985). The first attempt to identify the parS of P1 was by inserting different inserts into a plasmid and transforming it into a P1-harbouring host. The first study identified the P1 parS in a 112 bp insert. The second attempt to characterise *parS* was conducted in 1987 and helped to reduce the size of *parS* to a 34-bp sequence, which was later known as the small-*parS* (Martin *et al.*, 1987). The study highlighted the 13-bp inverted repeats within the small-*parS* as the ParB-interacting element where the fidelity was vital for the functionality of the ParABS system. Later, this subject of *parS* identity was revisited and studied using a DNAase footprinting assay (Davis and Austin, 1988; Davis *et al.*, 1990). These studies rectified the previous assumption which considered the 13-bp inverted repeats as the ParB-interacting element. The footprinting assay showed that it was the shorter heptamer (5'-ATTTCAC) and possibly the hexamer (5'-TCGCCA) found within the 34-bp sequence that were found to bind ParB. The study of Davis and Austin (1988) also suggested the involvement of a specific host protein that

facilitates the recognition of ParB towards the DNA binding sites which the protein is later confirmed as IHF by Funnell (1988). The study of Funnell and Gagnier (1993) confirmed the interaction between ParB and the hexamer and, coined the heptamer repeats as A-Boxes and the hexamer repeats as B-Boxes. In 2001, a study was published showing that the A-Boxes interact with HTH domain of ParB while the B-Boxes interact with dimerization domain of ParB (Surtees and Funnell, 2001). Finally, the study of Schumacher and Funnell (2005) provided the crystal structure of *parS*-ParB nucleoprotein complex as a strong evidence to support the A-Boxes and B-Boxes as the ParB interacting elements.

In nature, *par* system is usually employed by low-copy-number plasmid to counter their disadvantage upon random segregation. Many bacterial chromosomes also encode the homologs of *par* system; where some are active in mediating chromosome segregation only in a particular condition (Summers, 2009). To date, the segregation mechanism of chromosome in prokaryotes remains elusive. Therefore, the plasmid *par* system has been used as the model system to study the migration mechanism of DNA in a bacterial cell. In order to resolve the migration mechanism, many studies of *par* system were thus conducted focusing on the aspects of protein structure and mode of mechanism. These studies had been reviewed critically from time to time (Ebersbach and Gerdes, 2005; Schumacher, 2007; Schumacher, 2008; Gerdes *et al.*, 2010; Baxter and Funnell, 2014).

In general, a typical *par* system involves three elements – two *trans*-acting proteins (motor protein and centromere-binding protein) and a *cis*-acting centromere DNA sequence. As illustrated in Figure 2.4A, the three elements are always close to each other and the proteins are regulated under a single operon. It

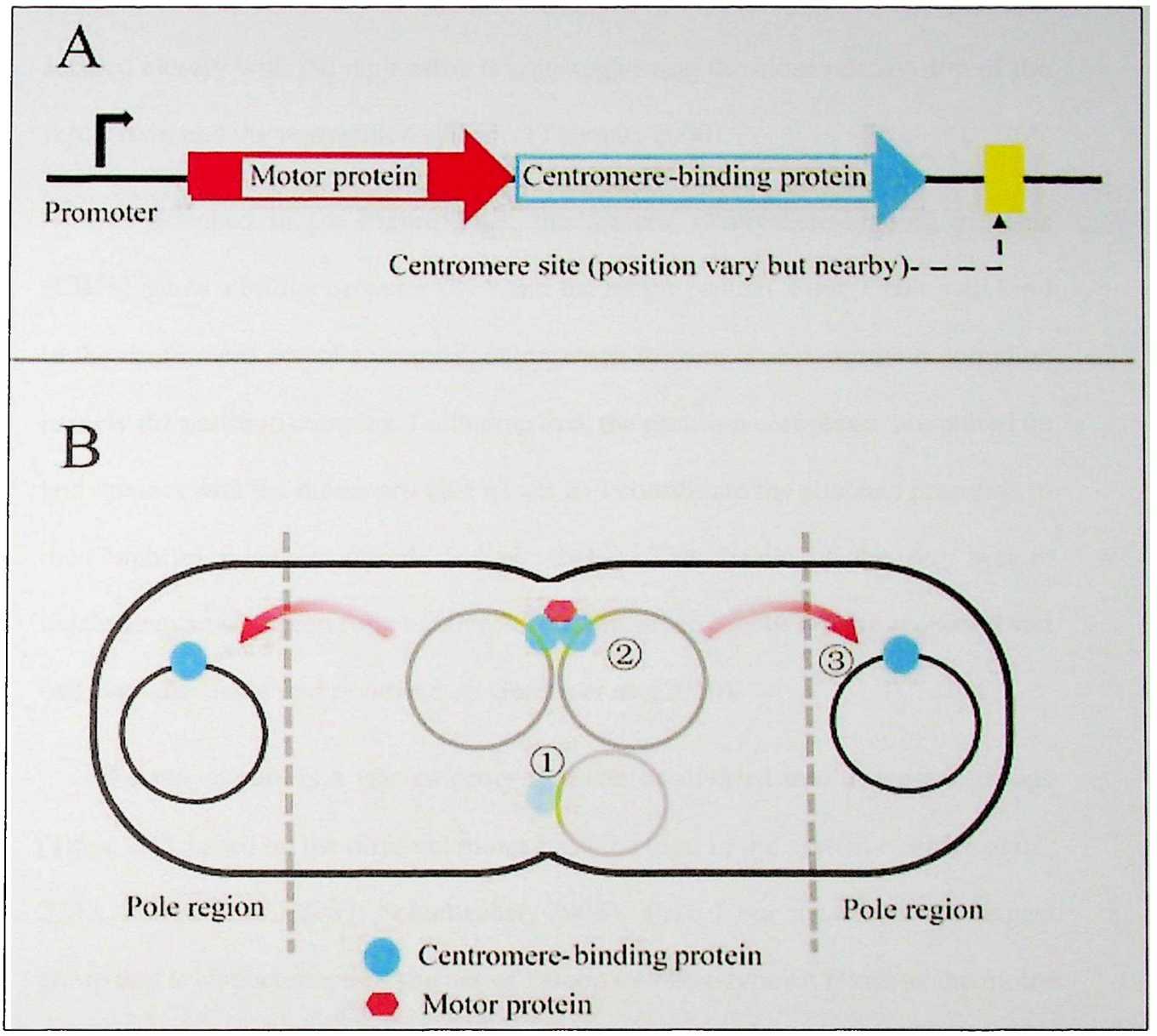


Figure 2.4 The general properties of active partitioning system.

Figures **(A)** illustrates the conserved orientation of the genetic elements encoding active partitioning system; **(B)** illustrates the generalized mechanism of active partitioning system, which includes 3 steps namely, **(1)** assembly of partition complex; **(2)** pairing of partition complex and recruitment of motor protein, and **(3)** migration of partition complex. The information illustrated here is adapted from Gerdes *et al.* (2000) and Gerdes *et al.* (2010).

is important to note that the operon transcribes the motor protein before the centromere-binding protein (Gerdes *et al.*, 2000). Typically, the *par* operon is located closely with the replication origin, suggesting the close relationship of the replication and the segregation systems (Thomas, 2000).

As sketched in the Figure 2.4B, the dimeric centromere-binding proteins (CBPs) act as a bridge between DNA and the motor protein. First, CBPs will bind to the centromere site of a plasmid, oligomerize to form a nucleoprotein complex, namely the partition complex. Following that, the partition complexes are paired up and interact with the motor proteins which will coordinate the attached plasmids to their rightful positions (Gerdes *et al.*, 2010). The details of the *par* system mechanism has not been fully uncovered yet, but several models were proposed and had been discussed and reviewed by Gerdes *et al.* (2010).

The *par* system is a vast category that can be divided into at least 3 groups (Table 2.1), based on the different motor proteins used in the system (Gerdes *et al.*, 2000; Larsen *et al.*, 2007; Schumacher, 2008). Type I *par* system is the largest group that is characterized by the use of P-loop (Walker-type) ATPase as the motor protein (typically named as ParA). The systems identified as type I can be further categorised as type Ia and type Ib.

The major difference between these two groups is that, the ParA of type Ia contains a DNA-binding domain which enables ParA to act as an autoregulator of its own operon. On the contrary, type Ib and another two subsequent systems rely on centromere binding protein as the autoregulator. The centromere site of type Ib and the following two systems are commonly found to be located upstream or within the operon to fulfil the dual-function of centromere-binding protein.

Table 2.1 Classification of active partitioning system.

Properties	Type I		Type II	Type III
	Type Ia	Type Ib		
Motor protein	P-loop (Walker Type) ATPase		Actin-like ATPase	Tubulin-like GTPase
Size of motor protein[a]	Large	Small	Small	Small
Size of CBP[a]	Large	Small	Small	Small
Operon autoregulator	Motor protein	CBP	CBP	CBP
Conserved motif of CBP	Helix-turn-helix (HTH)	Ribon-helix-helix (RHH)	Ribon-helix-helix (RHH)	Winged-HTH
Position of centromere[b]	Downstream	Upstream	Upstream	Upstream

[a] The size is a comparative of the respective proteins' homologs.

[b] The position of centromere describes the ideal centromere locus respective to *parAB* operon.

* The table is a compilation of information obtained from Gerdes *et al.* (2000), Larsen *et al.* (2007) and Schumacher (2008).