EPITOPE MAPPING OF HEMOLYSIN E ANTIGEN OF *Salmonella enterica* serovar Typhi BY PHAGE DISPLAY

by

CHIN CHAI FUNG

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2016

ACKNOWLEDGEMENT

First and foremost, I would like to convey my deepest appreciation and gratitude to my supervisor, Dr. Lim Theam Soon for his aspiring guidance, invaluable constructive criticism, patience and time that have supported me throughout my study. Under his valuable mentoring, I manage to improve my research ideas and skills. His teaching regardless of knowledge bestowed or life lessons shall be brought with me even after I have graduated.

Besides, I would also like to express my gratitude to Prof. Dato' Asma Ismail, Dr Choong Yee Siew and Dr Eugene Ong Boon Beng as my co-supervisors. Their help and guidance led to my success in completing this dissertation. I would also like to thank all the science officers, staff and my beloved colleagues (Bee Nar, Qiuting, Mimie, Anizah, Soo Khim, Angela, Chai Fen and Hamiza) that always lend me their hands whenever I need help.

I am also proud to be a recipient of the MyBrain (MyPhd) scholarship from the Malaysia Ministry of Education. This work would not have been possible without this sponsorship.

Last but not least, to my family and friends whom constantly encourage and support me throughout my study, I would like to say a big thank you for always be there for me. The only sentence that I would say to repay for all your support is 'Yes, I did it!'

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	three dimensional
ABTS	2,2’-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)
bp	base pair
BSA	bovine serum albumin
CDR	complementary determining region
CFU	colony-forming unit
C_H	constant heavy chain
C_L	constant light chain
dAb	domain antibody
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
dH_2O	distilled water
dsDNA	double stranded DNA
eGFP	enhanced green fluorescence protein
ELISA	enzyme linked immunosorbent assay
E. coli	*Escherichia coli*
Fab	fragment antigen-binding
Fv	fragment variable
His-tag	histidine tag
HlyE	Hemolysin E
HRP	horseradish peroxidase
hr	hour
Ig	immunoglobulin

IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kiloDalton
mAb	monoclonal antibody
min	minute
OD	optical density
o/n	overnight
PcAb	polyclonal antibodies
PBS	phosphate buffer saline
PBST	phosphate buffer saline containing Tween
PCR	polymerase chain reaction
Phty	healthy pool sera
PTM	skimmed milk in PBST
Pty	typhoid pool sera
RF	replicative form
RT	room temperature
S. Typhi	*Salmonella enterica* serovar Typhi
sec	second
scFv	single-chain fragment variable
ssDNA	single stranded DNA
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TUPs	target-unrelated peptides
Ub	ubiquitin
V_H	variable heavy chain
V_L	variable light chain

LIST OF SYMBOLS

^{o}C	degree Celcius
%	percent
g	gram
xg	gravity force
L	liter
µg	microgram
µL	microliter
M	molar
mg	milligram
mL	milliliter
mM	millimolar
ng	nanogram
nm	nanometer
rpm	revolution per minute
U	unit of enzyme
v/v	volume / volume
w/v	weight / volume

PEMETAAN EPITOP ANTIGEN HEMOLISIN E DARIPADA

***Salmonella enterica* serovar Typhi DENGAN PAMERAN FAJ**

ABSTRAK

Hemolisin E (HlyE) merupakan toksin pembentukan liang baru dan sebagai penentu virulen penting untuk patogenesis *Salmonella* Typhi dan Paratyphi A. Pemetaaan epitop B-sel untuk *S.* Typhi HlyE daripada serum manusia yang dijangkit deman kepialu (Pty) akut adalah penting kerana toksin ini adalah antigenik dan spesifik untuk mengesan demam kepialu. Satu perpustakaan fajmid 20-mer peptida rawak telah dibina dengan anggaran saiz sebanyak $3x10^9$. Perpustakaan peptida rawak tersebut telah dibina menggunakan kodon NNK yang mengekod kesemua 20 asid amino dengan satu kodon ambar hentian yang digantikan dengan glutamina oleh strain *Escherichia coli* (*E. coli*) penindas ambar. Pada masa yang sama, pengayaan antibodi spesifik HlyE daripada serum Pty juga dilakukan dengan mengunakan penulenan-imuno. Penyaringan perpustakaan pameran faj 20-mer peptida rawak dengan antibodi poliklonal (PcAb) yang diperkayakan juga dijalankan. Jujukan peptida yang diperkaya ditapis dengan SAROTUP dan seterusnya dianalisa dengan perisian web Pepitope dan Episearch untuk mengenal pasti epitop konformasi yang berpotensi. Selain daripada itu, perisian web ElliPro, BCPREDS dan CBTOPE juga digunakan untuk menjangka kedua-dua epitop linear dan konformasi berdasarkan struktur dan jujukan HlyE. Setelah pelbagai analisis dilakukan dengan menggunakan perisian bioinformatik, epitop linear dan konformasi yang dijangka, iaitu GAAAGIVAG dan PYSQESVLSADSQNQK disahkan dengan serum Pty. Epitop ini seterusnya digunakan untuk mengasing antibodi monoklonal yang berinteraksi berlawanan dengan demam kepialu dengan menggunakan kaedah penyaringan baru

iaitu MSIA™ Streptavidin D.A.R.T's® panning dengan perpustakaan faj antibodi. Setelah struktur tiga dimensi (3D) *S.* Typhi HlyE dan antibodi yang berpotensi untuk mengikat epitop tersebut diperolehi, pendokkan molekul telah dilakukan untuk menyediakan pandangan yang lebih mendalam berkenaan dengan interaksi antigen-antibodi dan pembaikan epitop yang dijangka. Pemetaaan epitop bersama dengan analisis komputer tentang interaksi tersebut membolehkan pemahaman yang lebih tepat pada posisi dan kedudukan konformasi epitop dengan antibodi. Antibodi monoklonal yang diasing berlawanan dengan epitop berkemungkinan mempunyai aplikasi hiliran yang berpotensi untuk diagnostik dan terapeutik demam kepialu. Kesimpulannya, keputusan daripada kajian ini boleh membantu dalam kajian selanjut untuk memahami mekanisme tindakan HlyE dalam patogenesis demam kepialu.

EPITOPE MAPPING OF HEMOLYSIN E ANTIGEN OF *Salmonella enterica* serovar Typhi BY PHAGE DISPLAY

ABSTRACT

Hemolysin E (HlyE) is a novel pore-forming toxin and an important virulence determinant in *Salmonella* Typhi and Paratyphi A pathogenesis. This toxin is antigenic and specific to detect typhoid fever, thus, mapping of B-cell epitopes of *S.* Typhi HlyE from pooled acute human typhoid (Pty) sera is of major interest. A random 20-mer peptide phagemid library was generated with a library size of $3x10^9$. The random peptide library was generated with NNK degeneracy that encodes all the 20 amino acids with one amber stop codon which is substituted by glutamine in amber suppressor *Escherichia coli* (*E. coli*) strains. At the same time, enrichment of HlyE-specific antibodies from Pty sera was also done using immunopurification. Biopanning of the phage display 20-mer random peptide library with the enriched polyclonal antibodies (PcAb) was carried out and the enriched peptide sequences obtained were filtered with SAROTUP and subsequently, analyzed by Pepitope and EpiSearch web tools to interpret for potential conformational epitopes. In addition, ElliPro, BCPREDS and CBTOPE web tools were also employed to predict the both linear and conformational epitopes based on structure and sequence of HlyE. After the multiple analysis using bioinformatics tools, the predicted linear and conformational epitopes, i.e GAAAGIVAG and PYSQESVLSADSQNQK were further validated using Pty sera. These epitopes were then used to isolate interacting monoclonal antibodies (mAbs) against typhoid fever using a novel MSIA™ Streptavidin D.A.R.T's® antibody phage library biopanning approach. After obtaining the three dimensional (3D) structures of the *S.* Typhi HlyE and potential

antibodies binding to these epitopes, molecular docking was performed to provide an in-depth view of antigen-antibody interaction and refinement of the predicted epitopes. Epitope mapping together with computational analysis of the interactions provide an insight into the positional and conformational position of the epitopes with the antibodies. The mAbs isolated against the epitopes could have potential downstream applications in diagnostics or therapeutics for typhoid fever. In conclusion, the results from this study could pave the way for further work to understand the mechanism of action of HlyE in typhoid pathogenesis.

CHAPTER 1

INTRODUCTION

1.1 Research background

Epitope mapping is a robust tool to identify epitopes of a target antigen that can elicit immune responses. There are many different approaches used for epitope mapping which includes co-crystallization of antigen-antibody complex, overlapping peptide strategies, nuclear magnetic resonance, site directed mutagenesis at given positions of the antigen, computational docking, combinatorial approach of phage display technology and bioinformatics analysis (Gershoni *et al.*, 2007). However, phage display is a popular method to display random peptides on the surface of a phage particle. These phage display random peptide libraries are usually generated by inserting degenerate oligonucleotides encoding for various peptide near pIII or pVIII coat proteins in the Ff bacteriophage (Fagerlund *et al.*, 2008). Due to its convenience of use and the ability to identify both linear and conformational epitopes simultaneously using purified antibodies or polyclonal sera, phage display becomes the preferred method in epitope mapping (Cortese *et al.*, 1994). The approach has been effectively used to map epitopes for various diseases such as cancer (Xu *et al.*, 2003), autoimmune diseases and also infectious diseases (Wang and Yu, 2004; Mullen *et al.*, 2006). By using phage display random dodecapeptide library, According to Cheong et al. (2016) has successfully identified 14 potential epitopes within *Plasmodium knowlesi* parasite causing life-threatening human malaria

(Cheong *et al.*, 2016). In addition, random peptide libraries were also employed in epitope mapping against ricin toxin after anti-ricin PcAb have been purified. This epitope mapping approach yielded numerous peptides that bind specifically towards ricin and four immunodominant epitopes were discovered (Cohen *et al.*, 2014). The mapping process with phage display would yield a tremendous amount of sequence related data that need to be processed and analyzed against a protein model. To analyze a diverse collection of sequences, bioinformatics tools such as Pepitope and EpiSearch web servers are used to map the peptides on the 3D antigenic protein structures to delineate specific epitope regions (Mayrose *et al.*, 2007a; Negi and Braun, 2009). Phage display technology has been successfully applied to identify the epitope sequences of RNAIII-activating protein (RAP) of *Staphylococcus aureus* by biopanning with polyclonal anti-TRAP antibodies using a random peptide library (Yang *et al.*, 2005). Epitope prediction using bioinformatics tools alone is easier and cost-effective but it suffers from accuracy issues. Thus, a combinatorial approach using phage display technology and bioinformatics tools is a preferred method for epitope mapping.

Identification of epitopes is the first step to understand and identify antigenic regions of a target antigen. In this work, the epitope mapping of the *S.* Typhi HlyE antigen was done using Pty sera. This provides a more accurate picture of the antigenic regions recognized naturally by the immune response. However, identification of the epitopes alone from a PcAb pool will not allow the identification of the immune-triggered antibodies in the sera. Thus, upon obtaining the epitope regions, antibodies would be generated against those target epitopes for potential downstream applications. The mAbs generation will be carried out by phage display biopanning using a human naïve scFv library against antigenic peptides identified as

the epitopes of HlyE. This approach has been successfully applied for other diseases. For instance, anti-Cry1C scFvs were successfully isolated against Cry1C toxin using commercial human naïve scFv antibody library (Tomlinson I+J) that can potentially be used as detectors in food and agricultural samples (Wang *et al.*, 2012). Although the isolated mAbs were not against infectious agent, this still proves that mAbs were able to be isolated using phage display technology. Thus, mAbs isolated against HlyE epitopes have the potential to be applied in antibody based therapies by inhibiting the infection process through toxin neutralization. The mAbs generated could also be used to dock with the epitope regions of HlyE to provide an insight into the interactions between the epitope and paratope as well as refinement of the target epitope regions. Thus, obtaining a precise epitope is essential for disease diagnosis, vaccine development, antibody and immunological therapies.

1.2 Literature review

1.2.1 Typhoid fever

Salmonella enterica serovar Typhi is a Gram negative and highly virulent bacterium that only infects humans (Baker *et al.*, 2011). This bacterium causes typhoid fever and multi-systemic enteric infections are commonly found in developing countries (Ivanoff *et al.*, 1994; Parry *et al.*, 2002). *S.* Typhi is taxonomically referred as *Salmonella enterica*, subspecies *enterica* with serovar Typhi (World Health Organization, 2008). *S.* Typhi together with *S.* Paratyphi A and B are contagious bacteria that can cause systemic infections by reaching the reticuloendothelial system after 10-14 days of incubation (Ivanoff *et al.*, 1994). In recent years, this bacterium is

reported to have epidemic potential instead of just being endemic (World Health Organization, 2008). This highlights the severity of typhoid infections globally.

Typhoid fever is a well-known contagious disease caused by *S.* Typhi which remains as a public health problem mainly in underdeveloped and developing countries (Korbsrisate *et al.*, 1999; Shahane *et al.*, 2007; World Health Organization, 2008). Typhoid fever is a detrimental infection which could lead to intestinal perforation and haemorrhage complications in 0.5-1 % of the cases (Ivanoff *et al.*, 1994). According to the review by Buckle *et al*. (2012), the estimated typhoid fever episodes in 2010 were 13.5 million in total (Buckle *et al.*, 2012). In addition, *S.* Typhi commonly causes bacteremia among patients of the age groups between 2-15 years. A recent study conducted in India, Pakistan and Indonesia revealed a high infection rate of 573 cases per 100 000 children tested (Ochiai *et al.*, 2008). Typhoid fever is distinct from paratyphoid fever clinically with the latter demonstrating milder symptoms with an infection ratio of 1:10 (Ivanoff *et al.*, 1994). Generally, the mode of transmission of typhoid fever is through fecal-oral route by ingestion of contaminated water (World Health Organization, 2008; Baker *et al.*, 2011) or food (World Health Organization, 2008) with humans as the sole natural host and reservoir (Ivanoff *et al.*, 1994). After ingestion, the infectious agents will reach the lamina propria of the small intestine and survive within the macrophages which engulf and digest the typhoid-causing bacteria. Some of the typhoid-causing bacteria will remain within macrophages in the small intestine and lymphoid tissue, while others will be drained into the mesenteric lymph nodes for further action by macrophages (Ivanoff *et al.*, 1994). Moreover, after the ingestion of *S.* Typhi, an asymptomatic period commences that usually lasts for 6-14 days, followed by fever and malaise. Fever with influenza-like symptoms such as chills, headache, anorexia,

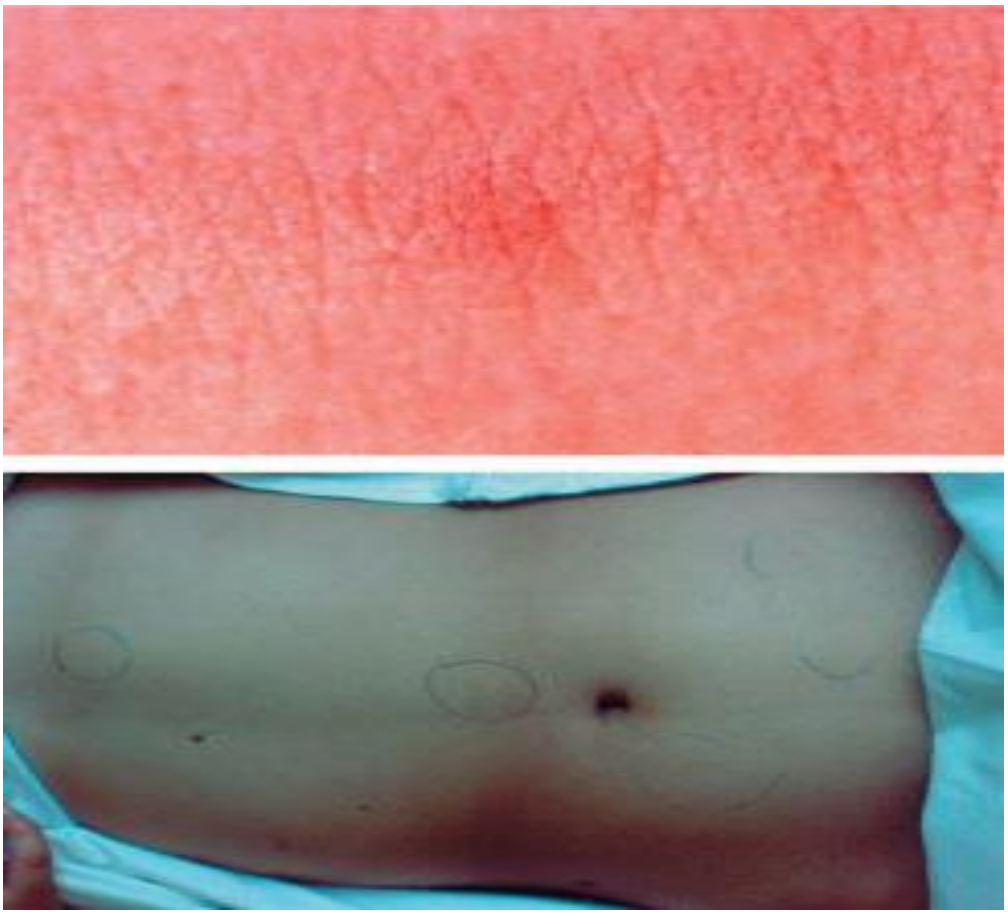
nausea, malaise are commonly observed with hepatomegaly, splenomegaly, rose spots on abdomen (Figure 1.1) and chest may be detected (Connor and Schwartz, 2005).

Epidemiological data elucidates the size of typhoid inocula to be dependent on waterborne and foodborne transmissions with the latter providing the biggest impact (Ivanoff *et al.*, 1994). A common problem associated with typhoid fever is misdiagnosis for common food poisoning and diarhea. This misdiagnosis could worsen the situation and perpetuate the spread of the disease (Parry *et al.*, 2002). With the emergence of drug resistant bacteria strains, it has been challenging for health service providers to manage the disease, especially in underdeveloped and developing countries (Bhutta, 2006). Live vaccine Ty2la and Vi polysaccharide vaccine were given to patients in developing country to cure *S.* Typhi infections with various effectiveness (Ivanoff *et al.*, 1994). A clear picture of global typhoid fever epidemiology may aid in disease control by deciding the most competent target vaccine (Lin *et al.*, 2001) and employing appropriate preventive measures.

1.2.2 Hemolysin E

HlyE is a novel pore-forming toxin exhibited by *E. coli*, *S.* Typhi and *Shigella. flexneri* (*S. flexneri*). Cytolysin A (ClyA) and silent hemolysin A (SheA) (Hunt *et al.*, 2008) are the other scientific names for HlyE. According to Ong *et al.* (2013), the rHlyE protein of *S.* Typhi has a molecular size about 30 kDa (Ong *et al.*, 2013). The closest model available to the *S.* Typhi HlyE is the crystal structure of *E. coli* HlyE with the sequence similarity more than 90% (von Rhein *et al.*, 2006).

(a)



(b)

Figure 1.1: Common symptoms observed for typhoid fever. (a) A rose spot and (b) a small cluster of rose spots found on the abdomen. Figure adapted from Huang and DuPont (2005).

Figure 1.2 illustrates the 3D structure of *E. coli* HlyE in a water-soluble form. *E. coli* HlyE demonstrates a long rod shaped molecule which is a new architecture for toxin family structure. This HlyE molecule is predicted to form a pore-forming structure through oligomerization by forming a channel that projected out from the membrane (Wallace *et al.*, 2000)**.** The hydrophobic β-tongue region of the toxin which normally functions for membrane binding will eventually cause lysis of target cells by forming pores on the membrane surface (Atkins *et al.*, 2000; Wallace *et al.*, 2000; Hunt *et al.*, 2010). Moreover, HlyE is described to be rich in α-helices with numerous amphiphilic peptide segments. Some of these peptides are believed to participate in membrane binding as well as the pore forming. Surprisingly, there is no symptom of lysis in the bacterial cells though it has been notable that HlyE triggers lysis in erythrocytes and other mammalian cells (Oscarsson *et al.*, 1999) in addition to apoptosis in macrophages (Lai *et al.*, 2000).

HlyE gene of *E. coli* K-12 strain reveals a cytolytic phenotype with studies conducted suggesting that cytolytic activities of HlyE can be facilitated with the presence of cholesterol (Oscarsson *et al.*, 1999). In addition, functional *hlyE* gene is now detected in *S.* Typhi and *S.* Paratyphi A besides being common in *E. coli* strains. Nevertheless, *hlyE* gene is not found in *S. enterica* serovar Paratyphi B and serovar Paratyphi C, non-typhoidal strains like *S. enterica* subsp. *enterica* serovars (Typhimmurium, Enteritidis, Choleraesuis, Dublin and Gallinarum), *S. enterica* subsp. *arizonae* and S. *bongori* strains (von Rhein *et al.*, 2009). On the other hand, HlyE of *S.* Typhi and *S.* Paratyphi A can be triggered by the *Salmonella* transcription factor SlyA (von Rhein *et al.*, 2009) which is a regulator of virulence gene in *Salmonella* (Ellison and Miller, 2006). HlyE of *S.* Typhi and *S.* Paratyphi A are also able to produce effective pore-forming toxin that could lead to human infections and

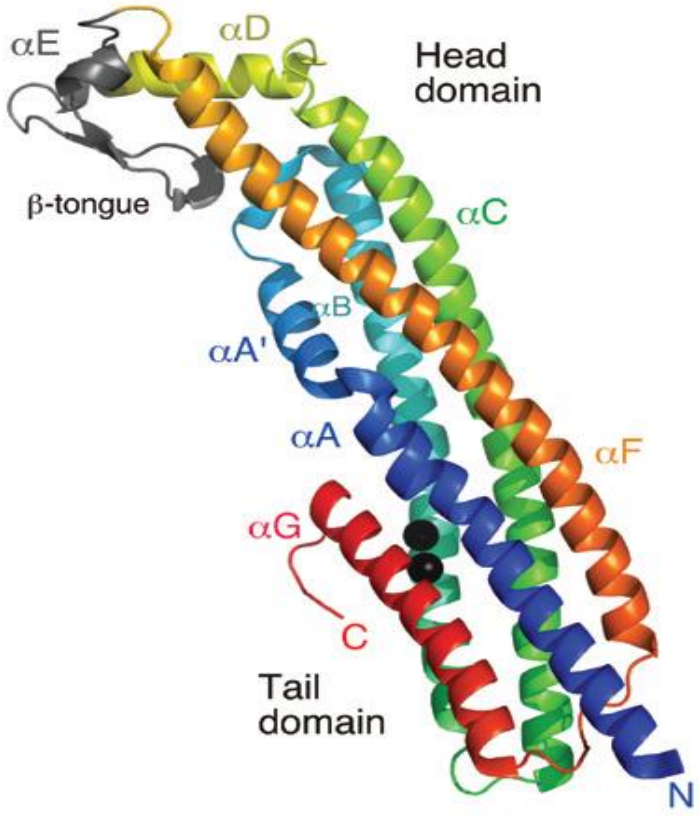


Figure 1.2: 3D structure of *E. coli* HlyE in water-soluble form. Figure adapted from Hunt *et al.* (2010).

certain *Salmonella* parthogenesis (von Rhein *et al.*, 2009). Recent serodiagnostic studies conducted showed HlyE under denatured condition (Liang *et al.*, 2013) and native condition (Ong *et al.*, 2013) remains antigenic and specific to detect typhoid fever making HlyE a potential candidate for diagnostics applications.

1.2.3 Antigenic epitopes

Antigens are mostly consisted of protein, carbohydrate, nucleic acid, lipid and phospholipid. Regardless of different types of antigens, they are recognized as foreign substances by the immune system. Since a whole antigen is too large to bind to the complementary-determining regions (CDRs) of a cognate antibody, hence, there are only specific fragments known as epitopes that interact with the antibody (Mahler and Fritzler, 2010). Epitopes are commonly elucidated as antigenic determinants that elicit immune responses in the hosts. Epitopes are commonly involved in protein-protein interactions as the contact points of any molecular interactions in a biological system. Nonetheless, epitopes are frequently represented as antigenic epitopes that are involved in specific binding against antibody or T-cell receptor. This B-cell epitope is regularly used interchangeably with epitope. Moreover, the regions on antibodies that are specific to interact with epitopes are referred as paratopes (Wang and Yu, 2004). Protein epitopes usually interact with antibodies due to complementary nature of epitopes and paratopes. This epitope-paratope interaction is further enhanced by various binding energy such as hydrogen bond, van der Waals interaction and ionic bond (Irving *et al.*, 2001).

There are various epitopes found on an antigen whereby structural properties and complexities play a pivotal role in the categorization of epitopes. Epitopes can

mainly be grouped into two forms, i.e. continuous and discontinuous epitopes as shown in Figure 1.3. Linear or continuous epitopes are any linear peptide fragments of the antigens that are capable to bind with the paratopes of the antibodies. This linear epitopes are not dependent on the structure of the protein and persist even after the protein is denatured (Sela and Pecht, 1996; Ladner, 2007). On the contrary, discontinuous or conformational epitopes are the major epitopes found in antigenic proteins which are made up of surface residues that are brought together via folding and strongly dependent on the native structure of the proteins (Sela and Pecht, 1996). Cryptotopes, neotopes and mimotopes are the less common epitopes found as compared to linear and conformational epitopes. Crytotopes are other forms of epitopes that are buried in polymerized proteins or viral particles and are only antigenic once the virus dissociate. This is when the hidden subunits are now exposed as individual subunit. Neotopes on the other hand are specific towards quaternary structures of viral particles and will lose its antigenicity once the virus dissociate. Mimotopes are common in phage display as peptide sequences that mimic linear or conformational epitopes binding to the target antibody but are irrelevant to the antigen (Van Regenmortel, 2009). These mimotopes induce identical or similar immune response elicited by the native epitopes. In fact, all epitopes have a vague boundary until they bind to certain parts of the antibodies. This literally happens for continuous and discontinuous epitopes since discontinuous epitopes are usually a combination of different short fragments of a few linear residues that raise the antibody responses. Furthermore, antigenic cross-reactivity is also a common phenomenon whereby an antibody is capable of binding with numerous related epitopes (Van Regenmortel, 2009).

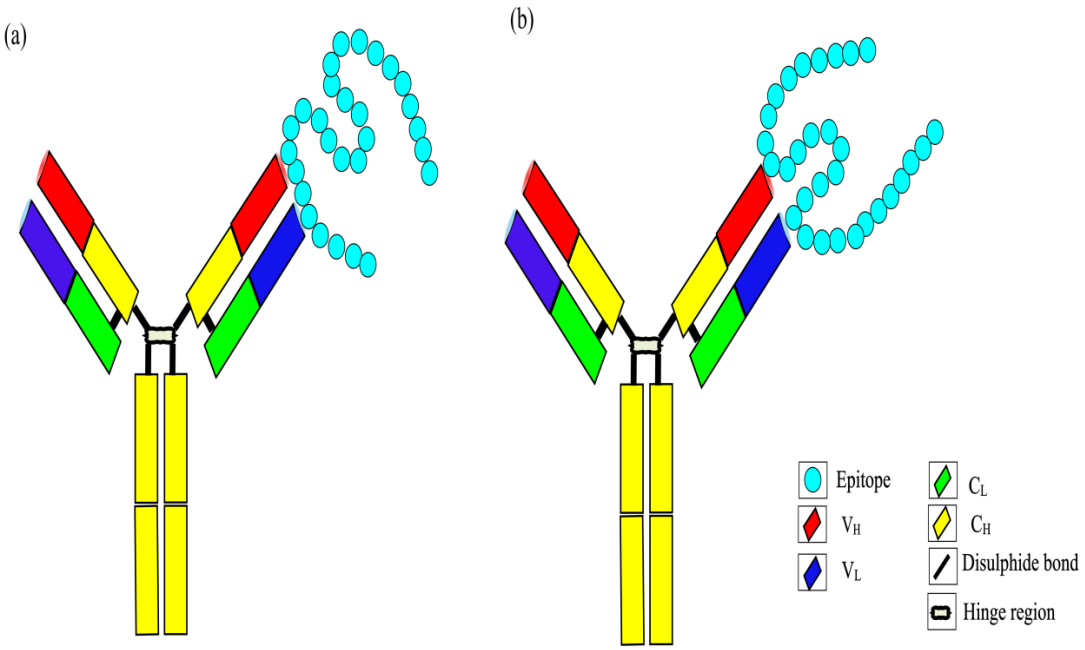


Figure 1.3: Description of epitopes. Binding of (a) linear and (b) conformational epitope to an antibody is illustrated.

1.2.4 Phage display technology

Phage display technology has been successfully applied to display and select of desirable peptides (Scott and Smith, 1990; Smith *et al.*, 2007), antibody (Knappik *et al.*, 2000; Sblattero and Bradbury, 2000) and enzyme inhibitors (Wang *et al.*, 1995). This technique has been widely used as the phenotype of the peptides displayed on the surface of a bacteriophage is physically linked to the genotype encoded for the peptides that embedded within the phage particles (Smith, 1985). Phage display technology can produce target-specific antibodies in a faster rate with comparable yield that involve no animals (Hoogenboom *et al.*, 1998). Basically, based on the types of vector, phage display systems can be categorized into phage, hybrid and phagemid (Kehoe and Kay, 2005). Hybrid system would produce two types of pIII in which one is with the fusion protein and other one is without (Kehoe and Kay, 2005), hence, the efficiency of displaying the desired protein is greatly reduced. Phagemid system is preferred over phage system for library cloning as phagemid systems has a better cloning efficiency and produces a higher diversity library (O'Connell *et al.*, 2002) that does not affect phage viability. This phagemid system has plasmid and phage-derived replication origin enabling it to replicate as double-stranded (ds) DNA like a plasmid as well as production and packaging as single-stranded (ss) DNA into the virion like a phage (Bratkovič, 2010). Besides being accessible to various restriction enzyme recognition, the genome of phagemid is also smaller allowing it to accommodate a larger insert. Phagemid system also provides additional advantages of easy expression of fusion protein and producing stable recombinant phage particles (Qi *et al.*, 2012). However, an additional step of co-infection with helper phage is needed to provide all other necessary proteins for phage assembly.

The success of phage display technology is highly dependent on the initial size and quality of the library (Sidhu, 2001), therefore strategies in the contruction of a phage display library have been extensively reviewed and improved for decades. Although different library designs and selection methods have been introduced, the basic principle of phage display remains the same (Böttger and Böttger, 2009). Commercial pre-made phage library (Ph.D 7 and Ph.D 12) which employs M13 phage system has been widely made available from New England Biolabs Inc. (NEB) and is convenient for the biopanning selection process (Fukunaga and Taki, 2012). Nonetheless, for epitope mapping purposes, the coverage of the peptides length is not optimum as usual conformational epitopes will range in between 10 to 22 residues (Van Regenmortel, 2009). Thus, a random 20-mer NNK phagemid peptide library has been designed for a better coverage of epitope screening. NNK as the choice for amino acid degeneracy of the library as it encodes for 32 triplet codons which covers all 20 amino acids with only one amber stop codon (Smith and Petrenko, 1997; Castel *et al.*, 2011). By using amber supressor *E. coli* strains such as TG1, the amber stop codon could be read and susbtituted with glutamine (Weigert *et al.*, 1965) which does not affect the expression of desired peptides. The design and application of a 20-mer NNK peptide phagemid library could help to maximize the size as well diversity of the peptide library for efficient epitopes screening. Furthermore, the level of the peptide display is varied and greatly influenced by the length and sequence (Malik *et al.*, 1996; Sidhu, 2001). The advantages of phage display technology over other selection methods are cost-effective and time-saving in addition to easy handling.

1.2.4 (a) Biology of filamentous phage

Filamentous phages (M13, f1, fd) or commonly refered as Ff bacteriophages are modified to be used in phage display technology. The Ff bacteriophages are with rod shaped structures of 1 μm length and diameter of 6 nm (Kehoe and Kay, 2005). These phages carry the desired genetic materials would infect F pilus bearing bacteria to replicate multiple copies of the desired targets within a short period of time (Krumpe and Mori, 2006; Fukunaga and Taki, 2012). In fact, these three phages were independently isolated from USA and European sewage system. Having sequence homology of 98.5%, they have been used interchangeably in the studies of biology. The core genomes of these filamentous phages are categorized into three groups, i.e. replication, assembly and structural genes which require for complete replication in the bacterial hosts (Mai-Prochnow *et al.*, 2015).

Ff bacteriophage genomes can be manipulated in phage display by packaging foreign genes into their capsids that allow expression of the fusion proteins outside of capsids. Infection begins when the phage particles that carry the gene of interest attach to the F-pilus of *E. coli.* After the pilus retracts, this brings the phages closer to the surface. The host Tol protein complex causes phage coat proteins to depolymerize and deposit into the cytoplasmic membrane. Subsequently, the ssDNA of the phage particles enters the bacterial cytoplasm. The ssDNA will be converted into double-stranded replicative form (RF) by *E. coli* proteins with the goal of creating multiple copies through rolling circle amplification. As the concentration of RF increases, pV viral protein binds to the + strand of the DNA that blocks the conversion to RF which later allows phage packaging to occur. Phage packaging occurs in the cytoplasmic membrane where all the viral proteins are assembled with the genetic material, and finally, nascent phage particles are released from the

bacterial cell without causing cell lysis (Russel *et al.*, 2004; Kehoe and Kay, 2005). This unique characteristic of Ff bacteriophage has reduced contamination due to the cell debris, thus, simplifies the purification process between rounds of biopanning (Castel *et al.*, 2011). Moreover, filamentous phages are taken precedence over tailed phages like λ and T7 in phage display with the advantages of having small genome size, ease of genetic manipulation, high durability towards a broad range of pH and temperature which enable a variety of binding and elution conditions during biopanning (Mai-Prochnow *et al.*, 2015).

Among the filamentous bacteriophages, M13 is the first bacteriophage developed as cloning vectors or phagemids in molecular biology. Upon infection with helper phage, phagemids replicate using the origin replication of the phage to produce numerous amount of ssDNA which are later packaged into phage particles in the bacterial hosts. The M13 bacteriophage has a ssDNA circular genome of 6407 bases that is shorter than fd bacteriophage (van Wezenbeek *et al.*, 1980; Kehoe and Kay, 2005). In addition, Figure 1.4 illustrates the structure of a filamentous bacteriophage M13. All 5 coat proteins (pIII, pVI, pVII, pVIII and pIX) provide structural stability to the phage particle. Nonetheless, only pIII coat protein is required for host cell recognition and infection (Sidhu, 2001). The pIII and pVIII are the commonly used coat proteins to display desired peptides on the surface of bacteriophage. The ends of the phage particle are capped by a combination of two minor coat proteins. The proximal end of the phage is capped by pIII and pVI, whereas pVII and pIX are responsible for capping the distal end (Barbas *et al.*, 2001; Kehoe and Kay, 2005). Employing pIII coat proteins as anchor proteins would have limitation towards number fusion proteins displayed or monovalent display. However, this minor coat protein allows large protein insertions and favours tight

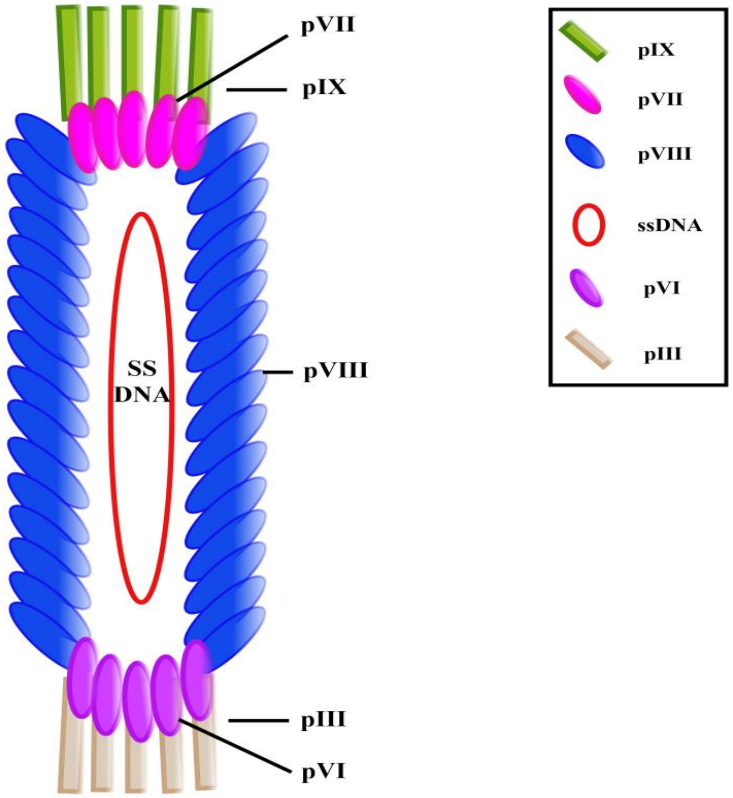


Figure 1.4: Structure of filamentous M13 bacteriophage.

binders. In contrast, the pVIII major coat proteins with 2700 copies allow polyvalent display, enabling avidity-based binding to immobilized targets. Nonetheless, these major coat proteins are limited with fusion proteins of smaller sizes and favour weakly bound monomers (Sidhu *et al.*, 2000; Russel *et al.*, 2004). The length of the virion is ultimately dependent on the size of the genome whereby a longer genome contributes to a longer phage, nonetheless, the exact conformation of the genome inside the phage particles is still unknown (Kehoe and Kay, 2005).

1.2.4 (b) Biopanning with a phage display library

Biopanning or affinity selection is commonly used in phage display technology to select desired clones by *in-vitro* incubation binding of phage library to targets, wash away those unspecific binders and elute bound phages as illustrated in Figure 1.5 (Russel *et al.*, 2004). The reversible binding between the protein library (antibodies and peptides) and the targets enables bound phages to be eluted and further amplified. These bound phages are amplified by infecting *E. coli* and replicating along with it. The amplification of the enriched phages is with the aid of helper phage as illustrated in Figure 1.5 due to phagemid system is employed. The affinity selection process is repeated in order to obtain enriched sequences for a particular immobilized target. Three to five rounds of biopanning are usually performed to isolate high affinity binders against the target (Brissette and Goldstein, 2007). DNA sequencing is then used to identify enriched fusion coat protein bound to the targets (Deshayes *et al.*, 2002). The genetic sequence for specific displayed peptides or antibodies can be identified as the target sequences are embedded within the bound phages (Willats, 2002). Different solid phases such as high protein binding microtiter plate (Krebs *et al.*, 2001), immunotubes (de Kruif *et al.*, 2009) and

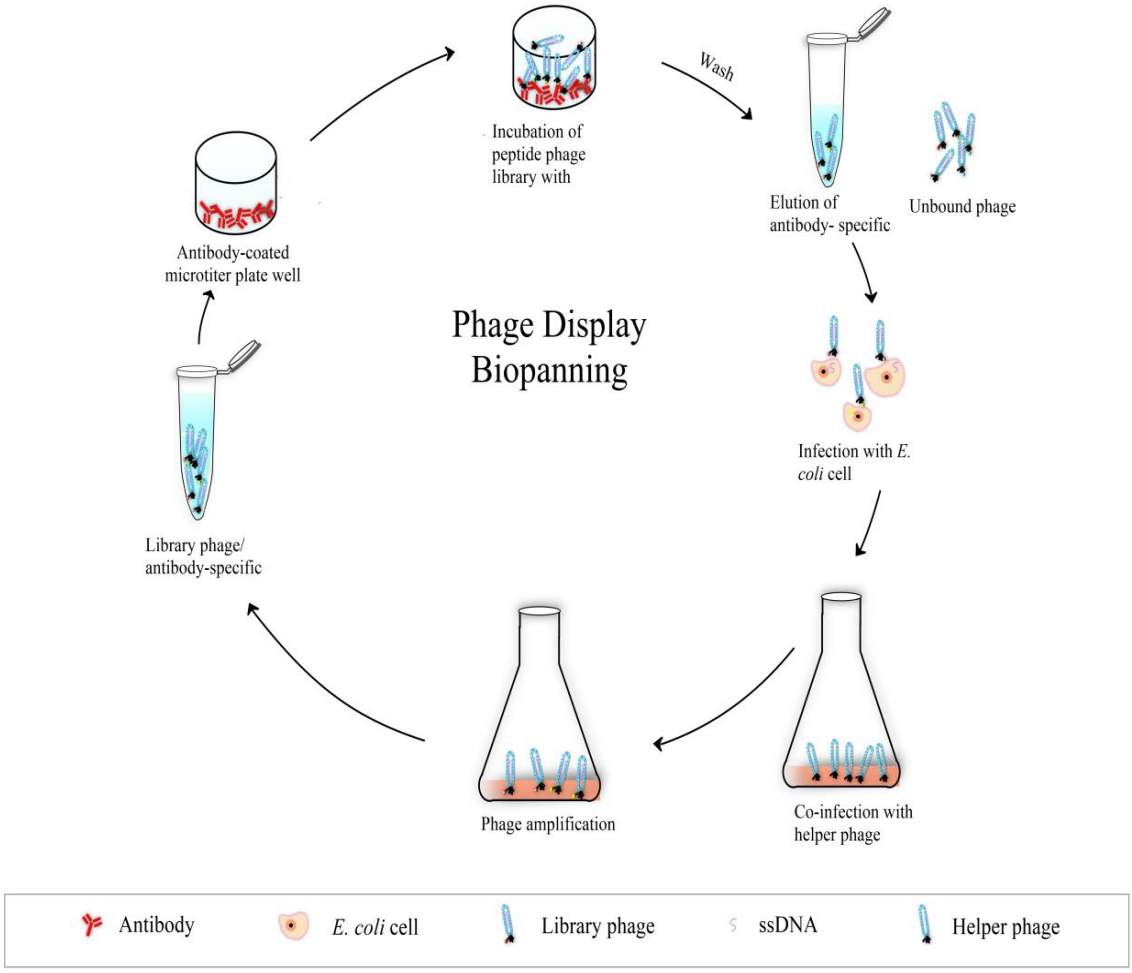


Figure 1.5: Phage display biopanning using conventional microtiter plate method.

magnetic beads (Walter *et al.*, 2001) are commonly used to capture the selected targets.

Epitope mapping of infectious agents using a phage display library has been extensively used whereby a random peptide library is employed to identify mimotopes of the target (Wang and Yu, 2004; Mullen *et al.*, 2006). Identification of mimotopes of *Neisseria meningitides* serotypes B capsular polysaccharides using a random peptide library enables the development of novel epitope-based vaccine against the infectious bacterium (Park *et al.*, 2004). Besides, a commercial Ph.D 12-mer random peptide library was also used to screen for target epitopes against rabies virus using purified an anti-rabies virus IgG as the template in the biopanning process. After the sequence analysis, RYDD-W-T motif was discovered to be the potential epitope that could help in development of vaccines or therapeutics for rabies (Yang *et al.*, 2013). Thus, phage display is applicable to elucidate protein-protein interactions between receptors and ligands and the success phage display biopanning has had in determining antibody-antigen binding sites or epitope mapping. Hence, application of phage display technology in epitope mapping will lead to advancements in diagnostics and epitope-based vaccine development (Cortese *et al.*, 1994) for contagious diseases (Wang and Yu, 2004).

Apart from that, isolation of mAbs against targets employing phage display is also one of the popular approaches nowadays. Naïve, immunized and synthetic antibody libraries are the three common phage display antibody libraries available nowadays (Miersch and Sidhu, 2012). They differ in terms of the source where the antibody gene derived but all of the antibody libraries are practically effective in selection for different targets. For instance, monoclonal antibody fragments against *Bacillus thuringiensis* Cry1C d-endotoxins had been isolated from Tomlinson (I+J)

human semi-synthetic scFv library through biopanning process (Wang *et al.*, 2012). On the top of that, the Tomlinson I scFv library was also employed to produce antibodies against the cyanobacterial hepatotoxin microcystin (McElhiney *et al.*, 2000). Thus, phage display technology is an effective means to generate mAbs against various toxins. In short, phage display technology is useful for epitope mapping and generation of recombinant antibodies.

1.2.5 Epitope mapping

An effective epitope mapping approach should be simple, fast and precise in identifying epitopes (Rojas *et al.*, 2014). Epitope structures can be elucidated by studying antigen-antibody interactions. X-ray crystallography would provide a precise identification and mapping of an epitope by studying the 3D structure of antibody-antigen complex. Nonetheless, this approach is rather time-consuming, expensive and requires highly purified protein complexes (Li *et al.*, 2003; Negi and Braun, 2009). Sometimes, the crystalized antigen-antibody complex could be difficult or even impossible to obtain (He *et al.*, 2013). Another method to map epitopes would be synthesizing the overlapping potential epitope peptides and binding them to target antibodies to screen for specific target epitope sequences. This method is only applicable for linear epitope mapping as partitions of conformational epitope are no longer in frame with the protein structure and be detected by target antibodies (Robotham *et al.*, 2002).

Epitope predictions can be done based solely on bioinformatics tools with various accuracies. BCPREDS is a bioinformatics web tool that predicts linear B-cell epitopes based on the primary sequence of an antigen. This online web tool has two