CONSTRUCTION OF A SYNTHETIC SINGLE DOMAIN ANTIBODY PHAGE DISPLAY LIBRARY FOR MOLECULAR DIAGNOSTIC APPLICATIONS

NUR HIDAYAH BINTI HAIRUL BAHARA

UNIVERSITI SAINS MALAYSIA

2015

CONSTRUCTION OF A SYNTHETIC SINGLE DOMAIN ANTIBODY PHAGE DISPLAY LIBRARY FOR MOLECULAR DIAGNOSTIC APPLICATIONS

by

NUR HIDAYAH BINTI HAIRUL BAHARA

Thesis submitted in fulfillment of the requirements for the Degree of Master of Science

August 2015

ACKNOWLEDGEMENT

In the name of Allah SWT, the Most Gracious and the Most Merciful, I offer my humble gratitude to You for giving me the strength to complete this thesis.

First and foremost, my deepest gratitude and sincere appreciation goes to my dedicated supervisor, Dr. Lim Theam Soon whom has been the best mentor anyone could possibly have. I thank him for his constructive ideas, criticism, guidance and patience throughout the course of my study. I also would like to thank my co-supervisor Dr. Choong Yee Siew. They have successfully guided me through some difficult times and were always willing to sharpen my understanding of my research.

Most importantly, I am also greatly indebted to most beloved parents, husband and my siblings for their endless love and support of me to achieve my goals in life. My love for them transcends all boundaries. I could never have gotten where I am today without their encouragement. They have never ceased to comfort me in my darkest hours in completing my study and also many thanks go to my siblings for their love and moral support. I feel blessed to have them in my life

My heartfelt appreciation and love goes to my beloved lab mates; Lim Bee Nar, NoorSharmimi Omar, Chin Siang Tean, Loh Qiuting, Nur Faezee Ismail, Chin Chai Fung, Anizah Rahumatullah and Chan Soo Kim .They have always stood by me for four years through many phases of hardship and turbulence. I could not have completed this study without their generous help.

Lastly, I would like to acknowledge the financial support received from The Malaysian Ministry of Higher Education, My Brain and INFORMM, Universiti Sains Malaysia.

TABLE OF CONTENTS

LIST OF TABLES

.

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ab	Antibody
ABTS	2,2’-azino-bis (3-ethylbenzthiazoline 6-sulfonic acid) diamonium
Amp	Ampicillin
ampR	Ampicillin resistance gene (bla)
AVI-Tag	Avidin-Tag
Bp	Base pair
BSA	Bovine serum albumin
Cam	Chloramphenicol
CDR	Complementarity determining region
Cfu	Colony-forming unit
CH1	Constant Heavy Region 1
CH2	Constant Heavy Region 2
CL	Constant Light
C-terminus	Carboxy-terminus
Da	Dalton
dAb	Domain antibodi
ddH2O	Double distilled water
DNA	Deoxyribonucleic acid
dNTP’s	Deoxyribonucleosid-5’-triphosphate
dsDNA	Double stranded DNA
D segment	Diversity segment
E. coli	Escherichia coli

EDTA	Ethylendiaminotetraacetic acid
eGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay
Fab	Fragment antigen binding
Fc	Fragment crystalline
FP	Fluorescent protein
Fv	Variable fragment
Glu	Glucose
GFP	Green fluorescent protein
His-Tag	Histidine Affinity Tag
hr	Hour
HRP	Horseradish peroxidase
Ig	Immunoglobulin
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-ß-D-thiogalactoside
J segment	Joining segment
Kan	Kanamycin
Kb	Kilo base pairs
kDa	Kilo Dalton
KD	Dissociation constant
M	Mole / litre
mAb	Monoclonal Antibody
MCS	Multi-cloning site
Min	Minutes
mL	Millilitre

mm	Millimeter
MP	Milk powder
MPP	Magnetic particle processor
MTP	Microtitre plate
MW	Molecular weight (in Dalton)
nmol	Nanomoles per litre
Ni-NTA	Nickel-nitrilotriacetic acid
OD	Optical Density
OD_{600}	OD at 600 nm wavelength
o/n	Over night
PAGE	Polyacrylamide gel-electrophoresis
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with Tween 20
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pfu	Plaque-forming units
PP	Polypropylene
PTM	2 % milk powder, 1 % Tween 20 in PBS
PVDF	Polyvinylidene difluoride
RFP	Red fluorescent protein
RNA	Ribonucleic acid
rt	Room temperature
rpm	Revolutions per minute
scFv	Single chain variable fragment
scFab	Single chain fragment antigen binding

SDS	Sodium dodecylsulfate
sec	Seconds
SSB	Single-strand DNA binding protein
ssDNA	Single stranded DNA
syn dAb	Synthetic domain antibody
Taq	DNA polymerase from Thermus aquaticus
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)-aminomethane
Tween 20	Polyoxyethylenesorbitan monolaurate
U	Enyzme units
Uv	Ultra violet
V	Volt
V-genes	Variable genes
VH	Variable domain of the immunoglobulin heavy chain
VL	Variable domain of the immunoglobulin light chain
V segment	Variable gene segment
YFP	Yellow fluorescent protein
Mtb 16 kDa Hsp	*Mycobacterium tuberculosis* 16 kDa heat shock protein

Units

(v/v)	volume/volume
(w/v)	weight/volume
µg	Microgram
µL	Microlitre
µm	Micrometer
°C	Degree Celsius

LIST OF PUBLICATION

Hairul Bahara, N. H., G. J. Tye, et al. (2013). "Phage display antibodies for diagnostic applications." Biologicals.

PENJANAAN PERPUSTAKAAN FAJ PAPARAN DOMAIN ANTIBODI SINTETIK TUNGGAL UNTUK APLIKASI DIAGNOSTIK MOLEKUL

ABSTRAK

Antibodi domain telah dieksploitasi secara meluas sebagai perancah untuk penjanaan perpustakaan antibody sintetik kerana saiz tanpa bergantung dengan mekanisma lipatan mudah. Dalam kajian ini, penjanaan perpustakaan yang pelbagai menggunakan rangka manusia tunggal (VH3-23(DP47)) dan kepelbagaian sintetik diperkenalkan melalui kaedah mutasi rawak yang berlaku secara semula jadi dalam kawasan kaset saling melengkapi (CDR), CDR1,CDR2 dan CDR3 pada rantaian berat telah menghasilkan 10^9 saiz perpustakaan. Kepelbagaian sikuen bagi semua CDR dapat ditentukan hasil daripada 28 klon yang dipilih secara rawak. Daripada 28 klon, 18 klon telah dipulihara dengan kepelbagaian panjang CDR3 yang berbeza dan juga kepelbagaian dalam sisa amino asid. Kualiti perpustakkan yang dihasilkan dapat dinilai melalui proses seleksi terhadap dua jenis antigen protein; penyakit dan protein pendarfluor. Pelbagai klon sasaran unik khusus telah diperolehi bagi kebanyakan antigen. Walau bagaimanapun, terdapat 2 antibodi monoclonal yang telah berjaya diraih hasil daripada seleksi dengan *Mycobacterium tuberculosis* 16 kDa Hsp antigen (Mtb 16 kDa Hsp) yang berpotensi untuk digunakan untuk tujuan teraputik. Kesimpulannya, himpunan kepelbagaian perpustakaan naïf boleh digunakan pada masa hadapan untuk menyaringi antibodi pengikat dengan antigen berpotensi yang lain.

CONSTRUCTION OF A SYNTHETIC SINGLE DOMAIN ANTIBODY PHAGE DISPLAY LIBRARY FOR MOLECULAR DIAGNOSTIC APPLICATIONS

ABSTRACT

Domain antibodies have been widely exploited as a scaffold for the generation of synthetic antibody libraries because of their relatively small size and simple folding mechanism. In this study, the generation of a highly diverse library using a single human framework (VH3-23(DP47)) and synthetic diversity introduced by randomly mutating naturally occurring within complementarity-determining regions (CDRs) CDR1,CDR2 and CDR3 of heavy chain yielded a library size of 10^9. The sequence diversity of all CDRs was determined from 28 randomly selected clones. Out of the 28 clones, 18 clones were conserved with different length of CDR3 and highly diverse amino acids residues. The quality of the library was also validated by panning against two different types of protein antigens; diseases and fluorescent proteins. Multiple unique target specific clones were obtained for most antigens. However, two monoclonal antibodies were successfully raised against *Mycobacterium tuberculosis* 16 kDa Hsp (Mtb 16 kDa Hsp) antigens which could potentially be used for therapeutics. In conclusion, the diverse repertoire of the naïve library can be used in the future to screen for binders against other potential antigens.

CHAPTER 1

1.0 Introduction

The rise of recombinant antibody technology was made possible by a combination of innovations like polymerase chain reaction technology (Orlandi, Güssow et al. 1989; Hoogenboom 2005), phage display technology (Siegel 2002) and evolution of online data collection of human immunoglobulin genomic sequences (Hust and Dübel 2004; Benhar 2007). For the past decade, the exploration of an array of recombinant antibody libraries for various applications was carried out. Improvements in molecular biology have paved the way towards improving parameters in order to produce libraries with higher diversity, larger sizes and better quality. Numerous studies have been done with the aspiration to mimic the uniquely human adaptive immune system that constantly generates diverse binding capacities of antibodies in a miniature sized test tube.

In 1975, the very first monoclonal antibody was introduced via hybridoma technology that requires the immunization of animals (Muyldermans 2001). Generation of these monoclonal antibodies involved the incorporation of myeloma cells with antibody producing spleen cell (Kohler and Milstein 1975). Thus, the hybrid will feature traits from both cells by maintaining immortality and antibody production. Inevitably, the use of hybridoma technology to produce monoclonal antibodies suffered several setbacks (Hoogenboom 2005). Some of the main disadvantages of murine derived antibodies are the use of animal host, longer periods of time required for production, unable to generate functional human antibodies and incapable of generating antibodies against toxic antigens (Geyer, McCafferty et al. 2012). It is these bottlenecks that have made hybridoma technology an unattractive

prospect for antibody production. The degree of freedom on offer for researchers by recombinant antibody technology has led to it gaining popularity in diagnostic applications (Marks, Hoogenboom et al. 1991; Holt, Enever et al. 2000; Siegel 2002; Ohara, Knappik et al. 2006).

In vitro display methods such as phage display, yeast display, ribosome display technology were introduced as a major alternative for the generation of recombinant human monoclonal antibodies (Barbas, Kang et al. 1991; Silacci, Brack et al. 2005). It is an *in vitro* process that is independent of any regulation by the immune system. The most widely used method is the phage display technology. This method employs the use of filamentous bacteriophage M13 as the display machinery (Barbas and Barbas 1994). The ability of a bacteriophage to present a recombinant target on its surface was first evident with the pioneering work by George Smith with peptides.

1.1 Phage Display Antibody Library

Phage display has earned its spotlight as the gold standard *in vitro* display system that caters for the increasing demand for the generation of peptides and recombinant proteins especially antibodies. The underlying concept of this display technology is the physical linkage between genotype and phenotype. The robustness of this technology lies in its ability to control and manipulate selection conditions. It is therefore, independent of any regulation by the immune system. In addition, antibodies can be harvested without having to go through animal immunization. Over the last decade, *in vitro* display methods have been very successful in the generation of diagnostic and therapeutic antibodies.

In general, there is an array of bacteriophages that has been exploited for surface display such as T7, T4 and Lambda. However for phage display systems, the most commonly used bacteriophage is the Ff class of filamentous phage. The Ff phage comprises of M13, f1 and fd that belongs to the inoviridae family that infects gram negative bacteria bearing the F-episome. It is a long rod like shape particle that is made up of coat proteins encapsulating the single stranded genome. The viral coat is mainly made up of 5 types of coat protein (pIII, pVIII, pIX, pVII, pVI). The unique feature of the filamentous phage virus is the non-lytic lifecycle that has paved the way for an *in vitro* tool to study the protein-protein interaction as well as peptides. Phage propagation under the non –lytic cycle has allowed the phage display system to function as a tool for surface display. In the early 1980s, George P. Smith demonstrated the display of peptides via the fusion to the gIII gene of filamentous phage surface. From this discovery, we are able to obtain information on the phage physical linkage between genotype and its phenotype. The successful presentation of peptides was achieved, the first phage derived antibody library for monoclonal antibody production was reported (Winter, Griffiths et al. 1994)

Given the technological advancements over recent years, many researches have attempted to display numerous proteins through fusion with different coat proteins. However, with several limitations for display on each coat protein, only pIII is vastly used to display large proteins. The major advantage of gIII fusion is that it can tolerate relatively large insert without compromising the integrity of the F-pilus infection. It's worth mentioning that pVIII coat protein has also been used for display of proteins and peptides. On the contrary, this fusion suffers from few drawbacks. Because of the phage particles are vastly made up of pVIII coat protein, fusion of large proteins or peptide for display may compromise the stability and structure of

the phage particles (Iannolo, Minenkova et al. 1995). Moreover, the fusion to gVIII will correspond to avidity effects due to high copies of the protein being presented on the surface, hence hindering affinity binding. In addition, the favored detection system for M13 phage is based on antibodies to pVIII coat protein, therefore any alteration to the gene VIII may interfere with the phage detection.

There are two ways for foreign proteins or peptides to be inserted as a fusion to the phage coat proteins. It can be carried out using the phage vector or phagemid vector system. In this study, the phagemid system with gene III fusion is employed for the synthetic antibody library construction. Phagemids in general, are plasmids with an existing *E.coli* plasmid origin of replication, multiple cloning sites and an antibiotic-resistance gene inclusive of an additional Ff phage-derived origin of replication and gene III or gene VIII. This addition allows for the phagemid to be packaged as single stranded DNA (ssDNA) in viral particles. Phagemids can function as normal plasmids or packaged as recombinant single stranded DNA in the M13 capsid with the aid of a helper phage (Azzazy and Highsmith 2002). The added advantage of using this phagemid system over phage vector is that soluble proteins can be readily expressed in *E.coli* host without having to undergo any form of alteration.

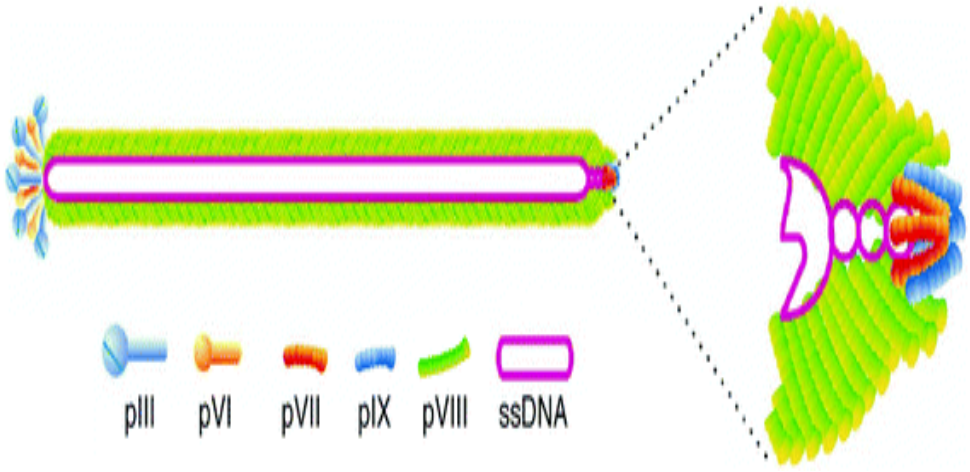

Figure 1.1 : Illustration of filamentous phage particle adapted from Eubanks 2007 (Eubanks, Dickerson et al. 2007). The phage particles are in linear rod like shape in which consisting of single stranded DNA and five coat proteins. pVIII coat protein also known as major coat protein makes up the vast structure of the phage protein capsid. Fusion of foreign proteins and peptides are usually to the gene VIII and gene III thus will be displayed fusion to pVIII and pIII coat protein.

1.2 Antibody Format

The classical format of antibodies are represented graphically as a Y-shape structure (Figure 1.2) with two identical ends (Wood 2006). At N-terminus, a heavy chain is linked via interchain disulphide bonds to a light chain to generate the Fragment Antigen Binding (Fab) (Rader and Barbas 1997). The binding pockets of the antibody is derived from the variable light and variable heavy domains within the Fab structure (Huston, Margolies et al. 1996). The advancement of recombinant antibodies through phage display has led to a wide array of different forms of antibody formats to be introduced (Hudson 1998).

To date, formats such as the human domain antibodies, camelid domain antibodies (Harmsen and De Haard 2007), single domain shark antibodies (Dooley, Flajnik et al. 2003), single chain fragment variable (scFv), tandem scFv, diabody, tetrabody, minibody and single chain fragment antigen binding have been extensively employed as formats for monoclonal antibody generation (Andris-Widhopf, Rader et al. 2000; Little, Kipriyanov et al. 2000; Holt, Herring et al. 2003; Hussack, Keklikian et al. 2012). Moreover, the limitation introduced by the folding machinery of *Escherichia coli* (Holliger and Hudson 2005) has resulted in the preferred use of smaller fragments such as domain antibodies to be heavily utilized for phage display (Holt, Herring et al. 2003; Dudgeon, Famm et al. 2009). Thus, the introduction of the current formats is essential for researchers to curb such limitations.

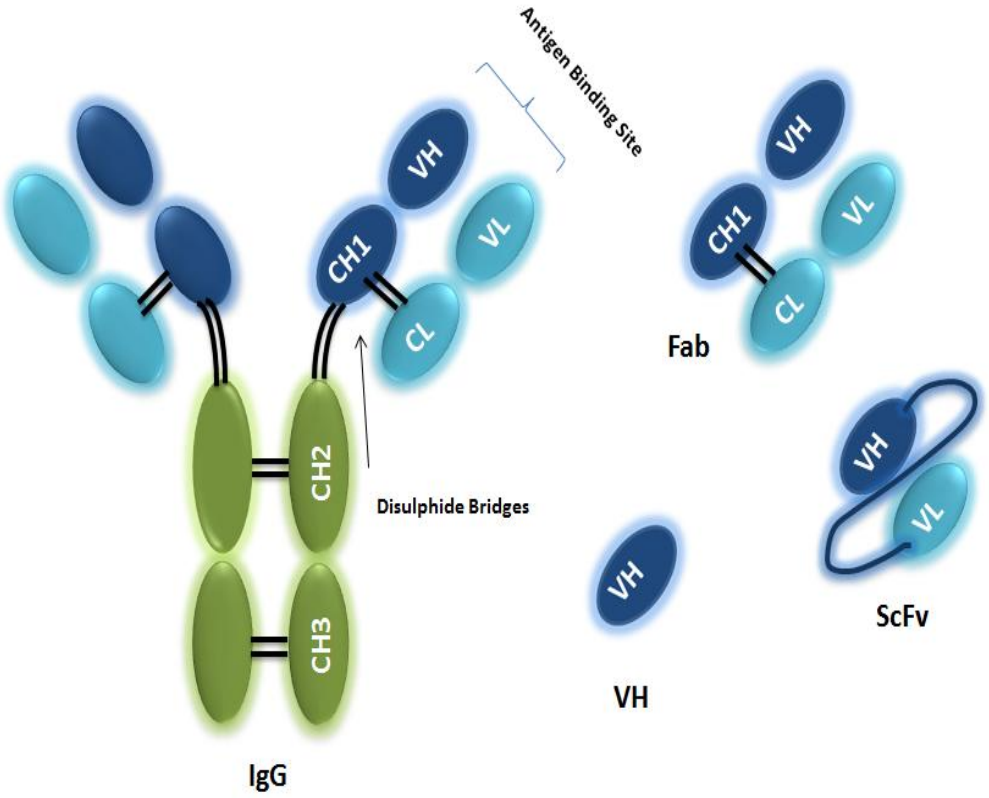

Figure 1.2 **:** Illustrations depicting the basic Y shape of an immunoglobulin and smaller antibody derivatives commonly used for phage display, Fab (Fragment antigen binding) and scFv (single chain fragment variable) adapted from Kierny 2012 (Kierny, Cunningham et al. 2012).

1.3 Domain Antibody: Unique Biophysical Properties

The first smallest known antigen-binding fragment coined as “nanobodies”, ‘‘domain antibody’’, or ‘‘dAb’’ was identified when a murine VH repertoire was selected against the model antigen hen-egg lysozyme with high affinity and specificity (Andris-Widhopf, Rader et al. 2000). Unlike scFv which is twice the size and Fab, four times the size of dAbs (Holt, Herring et al. 2003), their relatively smaller size is well suited for phage display (Chen, Zhu et al. 2009). Basically, dAb is the variable regions of either the heavy (VH) or light chains (VL) of immunoglobulins (Holt, Herring et al. 2003).

Recent commercial interest revolves around manufacturing humanized antibody with high specificity and affinity for potential diagnostic and therapeutic applications (Brekke and Sandlie 2003). The production of domain antibodies brings about the advantages over the use of conventional antibodies. To add to its commercial value, the antibody produced must meet the requirement of biophysical properties (Harmsen and De Haard 2007) such as high yield and soluble expression (Muyldermans 2001), heat stability (Goldman, Anderson et al. 2006) such as resistance to proteolysis, resistance to harsh condition (Dona, Urrutia et al. 2009) such as chemical degradation (Wang, Singh et al. 2007), aggregation and denaturation.

Initial studies of domain antibody showed that the expression and solubility of the first murine VH domain antibodies were low. The selection of the VH domain was done in mouse with the presence of a cognate VL, therefore, it was thought that the absence of the VH-VL hydrophobic interface contributed to the instability of the structure. After the setbacks, a modification was introduced in cloning of camelid

VHH domains. It was found that the solubility improved due to a hydrophilic mutation of a tetrad at the VL interface. Soon after, similar modifications of residues at positions 44, 45 and 47 was done in human VH domains with those frequently found in camel VHH domains. This approach is best known as ‘camelisation” (Davies and Riechmann 1994; Conrath, Vincke et al. 2005). However, despite having to overcome aggregation, the modified VH domains remained expressed at low yields and relatively unstable due to the deformation of the ß-sheet.

In resolving this issue, many researchers studied the effect of the VHH tetrad on solubility (Barthelemy, Raab et al. 2008) which resulted in a VH dAb library produced based on a murine germline gene with a substitution at the VL interface. Phage display panning was done with monomeric IgG-specific dAb and found to be soluble at a concentration of 2 mM (Holt, Herring et al. 2003). On the same note, good expression of fragments selected from the llama dAb library was attributed to the framework substitutions that differs from the VHH tetrad. Mutation and manipulation of the CDRH3 loop length also plays an important role to achieve good expression and solubility of the VH antibody (Riechmann and Muyldermans 1999).

Besides having good expression yield and solubility, another attractive property of several camelid VHH domains and llama VH domains is the heat stability (Dudgeon, Famm et al. 2009). In general, antibodies and their fragments derived from human VH dAbs tend to aggregate irreversibly upon heat denaturation. However, it was reported that when camel and llama VHH domains (Dolk, van Vliet et al. 2005) were subjected to heat ranging between 80–90^{0}C, they were able to maintain its antigen binding specificity despite prolonged incubation at high temperatures. Advancements made to cater for the thermo stability includes site-directed mutagenesis based approaches for directed evolution of antibodies.

Undoubtedly, successful isolation of recombinant antibodies from libraries depend heavily in the quality of the libraries produced. Factors such as the number of correctly folded functional antibodies have brought a paradigm shift towards developing human domain antibody. An example of recent studies showed functional antibody of HEL4 domain antibody library mimicking the natural human immune response designed with only CDR3 diversity (Mandrup, Friis et al. 2013). This library also includes mutations of the amino acid composition with regards to the positions critical for the folding and aggregation of domain antibodies.

With regards to dAbs high affinity and specificity, their small size and short half-life are best suited for targeting antigens in tissue and blood vessel where penetration is often obstructed and for clearance purpose. For example in tumour cells, dAb can be used to assist delivery of specific toxins to the tumour cells in a short time without damaging healthy cells. However, in some applications, such as in cancer treatment (Revets, De Baetselier et al. 2005) in which the target antigen resides in the blood stream, prolonged serum half-life is crucial to maximize time for antibody antigen reaction to minimizes the dosage amount.

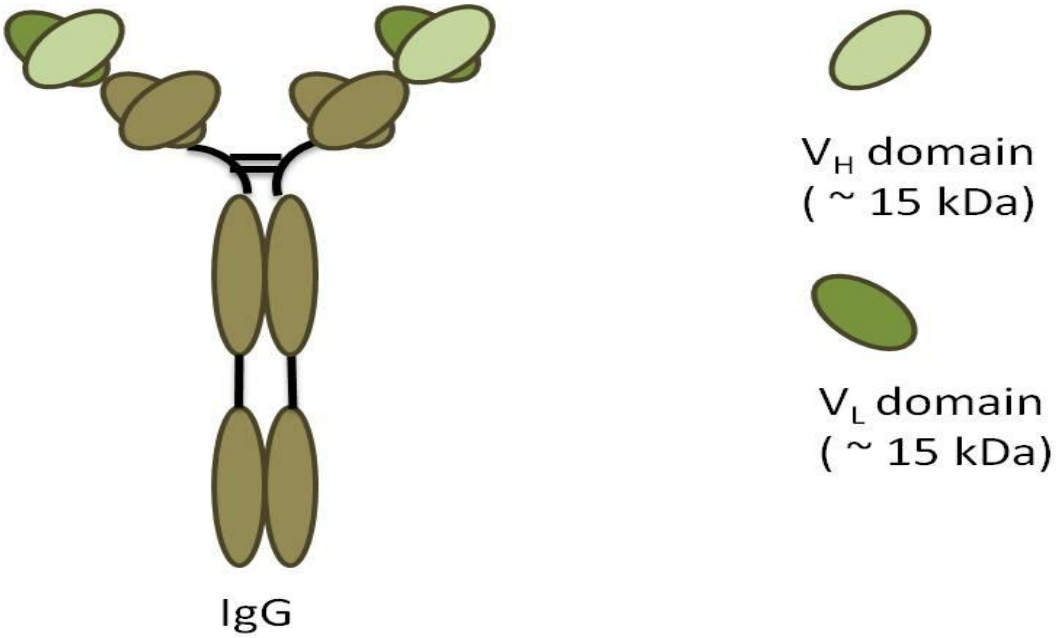

Figure 1.3 : Schematic representation of antibody formats ranging from conventional whole antibody to variable heavy and variable light chain domain antibody as the smallest unit (Chakravarty, Goel et al. 2014).

1.4 Synthetic Antibody Technology

The rise of synthetic antibody stems from the limitation of natural repertoire diversity. For the past twenty years, studies have shown a particular interest in producing antibody with high affinity antigen-binding sites by introducing diversity mutation in CDR loops. The construction of semi- or fully synthetic antibody library genes were assembled using chemically synthesized DNA. One of the classical methods of in-vitro antibody production derived from the natural antibody genes is via PCR. Naïve antibody repertoires via synthetic platforms are not biased for binders of any particular antigen and bypass the redundancies of naturally occurring antibody. Hence, the advantage of synthetic antibody over the natural repertoire is that it provides a wider diversity for any type of target.

Notably, technical advances have allowed the development of highly functional synthetic antibody libraries that rival or even exceed the recognition potential of natural immune systems. The first semi-synthetic library was reported in 1992 by Nissim and colleague with a human *VH* genes repertoire from 49 human germline *VH* gene segments with combinatorial synthetic CDR3 of five or eight residues (Hoogenboom and Winter 1992; Griffiths, Williams et al. 1994). Later on, it was again followed up with a 'single pot' human scFv library built from a diverse repertoire of *in vitro* human *VH* gene segments assembly with random nucleotide sequence for CDR3 lengths between 4 – 12 residues (Benhar 2007).

The second generation of synthetic platform is based on a more limited collection of variable domain genes however, emphasizing more on robustness. The overall design took into consideration the yields of functional antibody fragments based on cellular folding in the *E. coli* expression machinery (Welch, Govindarajan et al. 2009). Since then, Pini's group explored a semi synthetic antibody library that

was prepared using a single VH (DP47) and Vκ (DPK22)(Pini and Bracci 2000). The VH component of the library was created using partially degenerate primers in a PCR-based method to introduce random mutations at positions 95 – 98 in CDR3. It was found that creating antibody libraries starting from well-expressed frameworks was able to retain the diversity and stability (Hoogenboom, de Bruïne et al. 1998).

The 'HuCal' libraries were constructed with a more diverse sequence space although it is confined by the limited set of variable domain scaffolds (Benhar 2007). All the genes assembled were synthetically synthesized with a total of seven VH and VL (four Vκ and three Vλ) germline families that accounts for more than 95% of the human antibody diversity (Knappik, Ge et al. 2000). In addition, the genes were also optimized for expression in *E. coli*. The design of the library was based on cloning the V genes of scFv in all 49 combinations into a phagemid vector. Diversity was introduced in the CDR3 cassettes via generation of mixed trinucleotides sequences by substitution CDR3 regions of the master genes. Interestingly, the outcome of the library selection has resulted in obtaining high affinity binders with Kd between 10^9 M and 10^10 M. The variation of the CDR3 cassettes resulted in a highly diverse library producing antibodies against a vast number of antigens with high affinity.

1.5 Synthetic Domain Antibody Generation

Interestingly, the unique nature of generating highly diverse antibodies against a plethora of antigens by the immune system has intrigued researchers to mimic such processes *in vitro* with synthetic gene platforms. The genetic sequence of the variable domain is chemically synthesized with the introduction of randomization at fixed positions corresponding to the CDR of the variable domain with a fixed framework (Rothe, Urlinger et al. 2008; Yang, Kang et al. 2009; Prassler, Thiel et al.

2011). These degenerate oligonucleotides function as substitutes for the naturally occurring *in vivo* diversity. Synthetic platforms also take into account the variation in length of the CDR regions to fully maximize the diversity.

These oligonucleotides are designed using highly randomized codons which are used to code for unspecified amino acids. The generation of amino acids sequence depends on the codon usage. There are two commonly used codon scheme of encoding unspecified amino acids sequences; 1) NNK (A/T/C/G as an equimolar representation of N and K codes for G/T) 2) NNS (N represents four bases and S codes for G/C) (Barbas, Burton et al. 2001). These schemes produce 32 codons with one stop codon. N in general, produces 64 possible codons, hence coding for 20 amino acids. The most commonly used is NNK as it is able to produce high frequency of stop codon when used to encode for very large peptide consisting of more than 50.

Construction of recombinant domain antibody requires the chemically synthesized genes to be assembled in a manner that resembles the complete gene sequence. The first method of assembly was introduced by Stemmer and colleagues where full-length genes were generated (Stemmer, Crameri et al. 1995). This approach is known as the conventional one-pot gene assembly (Prodromou and Pearl 1992; Stemmer, Crameri et al. 1995; Wu, Wolf et al. 2006). It is an annealing and assembling process by incorporating the mixture of all synthetic oligonucleotide in a single step PCR. However, due to the variation in the length of degeneracy, it is rather difficult for gene assembly via the conventional method of polymerase chain reaction. One-pot gene assembly is likely less efficient for degenerate oligonucleotide with higher complexity as PCR is known to work well with a fixed sequence and not randomized sequences (Young and Dong 2004). Confined by the

limitation, two-step approaches have been proposed to assist in the assembly of genes with higher complexity for example by two-step PCR (Cherry, Nieuwenhuijsen et al. 2008), ligation chain reaction of fragmented segments (Au, Yang et al. 1998), gap filling and ligation (Ostermeier 2003).

Two-step PCR methods involve the assembly of multiple overlapping oligonucleotides by PCR to generate the template DNA followed by the amplification of the DNA template with two outermost oligonucleotides as primers. The ligation chain reaction confers ligation of smaller fragments to form a unit and subsequently amplified by PCR. The ligation chain reaction method however is slightly similarl of the ligation chain reaction wherein, the genes are assembled by polymerase gap filling in by ligating the ends together. Despite numerous proposed approach proposed, it is worth noting however that these method are not routinely used.

1.6 Antibody Selection by In Vitro Panning

Generation of antibodies by the immune system is involves the B cell repertoire where the V genes segments have undergone rearrangement. As a result, a single antibody is displayed on the surface of the each cell. The selection process occurs by the interaction between antibodies with the antigen. Selected antibodies will either segregate to short-lived plasma cells or to long-lived memory cells in lymph nodes, spleen, and bone marrow (Winter, Griffiths et al. 1994). For memory cells, the V genes of the selected antibodies will undergo hyper mutation. At this point, binding affinity may be improved with successive selection with antigen. With regards to mimicking the whole process of B cell antibody generation process, “panning” or “biopanning” is used. Biopanning refers to the iterative *in vitro* process

of antibody selection from antibody libraries based on target affinity (Kretzschmar and von Rüden 2002).

There are several conditions that need to be taken into consideration during the selection process. The first factor is the imperative proficiency of isolating the gene pool to construct an antibody library with high diversity, capability of expressing functional antibody fragment in soluble form and lastly, the efficiency of simultaneous expression and display of genetic information being packaged. The population of target specific antibodies are enriched relative to the number of panning rounds (Mullen, Nair et al. 2006).

Target antigens are commonly coated on various solid phase. The most common solid phase used are nitrocellulose, magnetic beads, agarose columns, monolithic columns, polystyrene tubes and 96 well microtitre plates (Kontermann and Dübel 2010). The solid phase bound phages are subjected to stringent washing to eliminate nonspecific binders. The subsequent step is then followed by recovery of the bound phages by elution. This process can either be by competitive elution (Krishnaswamy, Kabir et al. 2009) or harsh acidic (Barbas, Kang et al. 1991) or alkaline condition (Marks, Hoogenboom et al. 1991). Phage recovery or rescue plays a pivotal role in the whole panning process as this will ensure retrieval of high affinity binders. The phages are normally grown in bacterial culture for amplification thus the recovered phages can be subjected to further rounds of selection. Moreover, for each round of panning, phages can be enriched 20-1000 fold (McCafferty, Griffiths et al. 1990).

1.6.1 Panning Via Conventional Method (Immunotubes and Microtitre Plate)

Prior to selection, the target antigens are coated on the surface of the solid phase for presentation. Figure 4 shows the overall illustration of the conventional panning process. This will be followed by an incubation step with the antibody bearing phage particles to allow binding of antibodies to the antigen. Parameters such as physical, chemical or biological are essentially introduced (Lee, Iorno et al. 2007). Stringent washing steps are necessary to ensure the removal of unbound phage particles from bound phage particles. Discrepancy in the washing approach will result in the variation of enriched clones. Lastly, an elution step can be conducted in many ways either by enzymatic digestion, pH shift or competitive antigen elution. The eluted phage particles are then enriched by infection of *E. coli*. At this time, the phage particles can either be repackaged to be used in the subsequent panning round or for final analysis.

After 4 to 6 rounds of panning, identification of bound phage can be evaluated by antibody presenting phage or in the soluble form of antibodies on an immunoassay format (Walter, Konthur et al. 2001). The positive clones will then be sequenced to obtain the genotypic information pertaining to the positive clone. As the genetic information of the clone is now available, modifications to the antibody can be done and produced in different host depending on the platform the antibodies will be used in. The availability of the genetic information of the antibodies would also facilitate additional modifications in terms of stability and affinity maturation (Pini and Bracci 2000).

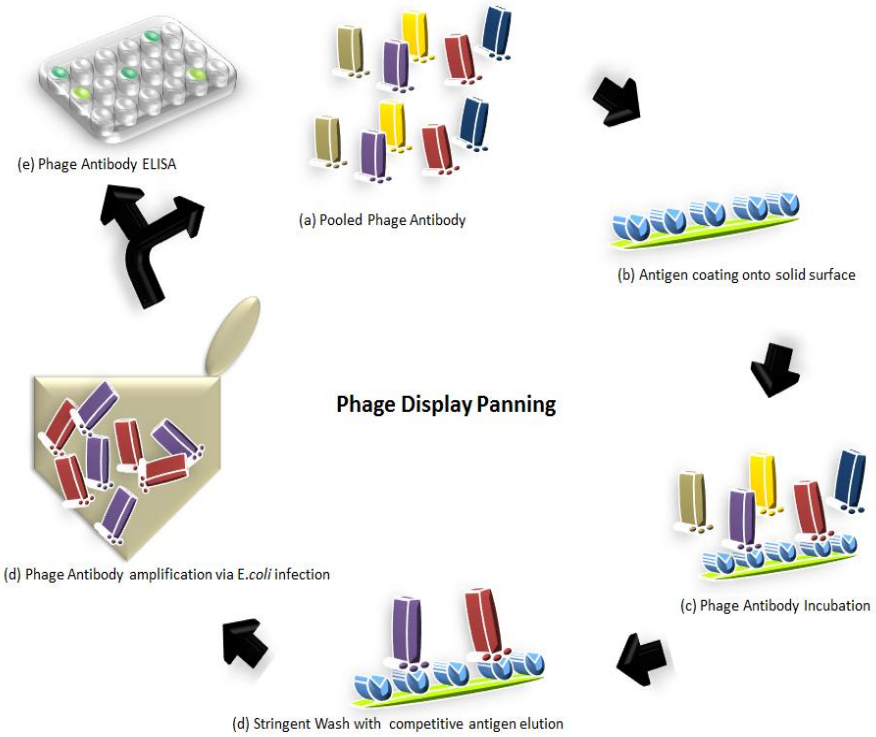

Figure 1. 4 : Biopanning protocol adapted from Kügler 2013 (Kügler, Zantow et al. 2013). (a) Pooled phage library will be subjected to panning (b) Antigens are immobilized onto solid surfaces and blocked to ensure nonspecific binding of phages onto the plastic surface. (c) The pooled phages will then be incubated with the immobilized antigen. (d) Unbound phages were then washed off by stringent washing subsequently followed by elution. (e) Antigen bound phage will be rescue by E.*coli* infection and followed by phage enrichment. (f) After every round of enrichment, the phage can either be subjected to phage ELISA or carried forward until the successive rounds completed.

1.6.2 Semi-Automated Panning

Screening large sample sizes is tedious with conventional panning procedures that require repetitive rounds of panning, phage infection and propagation. However in semi-automated systems to streamline the laborious process of phage display, such as semi-automated magnetic bead-based antibody selection (Konthur, Wilde et al. 2010), allows high-throughput screening of antibodies to be carried out with maximum convenience and minimal handling.

Conventional method of antigen immobilization using 96 well microtiter plates involve two methods, either by adsorption of antigens to the plate surface (Bora, Chugh et al. 2002) or coating the plates with streptavidin to capture the antigen s(Välimaa, Pettersson et al. 2003). In contrast to using the microtiter plate, another alternative is by allowing biotinylated antigens to be coated onto the streptavidin magnetic beads (Cox and Ellington 2001). These magnetic beads have larger surface area which contributes to the efficiency in the panning process as compared to using microtiter plates.

In practice, the panning method utilizes a pin-based magnetic particle processor (Kingfisher, Thermo) as shown in Figure 5(a). This machine enables the handling of 96 magnetic pins in which it is positioned similar to the common 96 well plate (Walter, Konthur et al. 2001; Rhyner, Konthur et al. 2003). The basic concept of using the processor is to streamline processes such as washing step, incubation times, and to conduct selections on same targets under different buffer conditions simultaneously. The software-driven procedure dictates the transferring process of magnetic particles between wells by capture and release motions shown in Figure 5(a). The rod-shaped magnets are covered with plastic caps during the transferring process to prevent contamination. However, the automation process only involves

the panning procedure wherein the subsequent step of phage rescue and enrichment is done manually. The advantage of semi-automated panning allows standardization of panning parameters and reduces background of non-specific binder when transferring from one well to another (Konthur and Walter 2002). The application of this method allows better reproducibility and a faster turnover rate in comparison to conventional plate based protocols. Therefore, the implementation of this method allows for high-throughput antibody discovery.

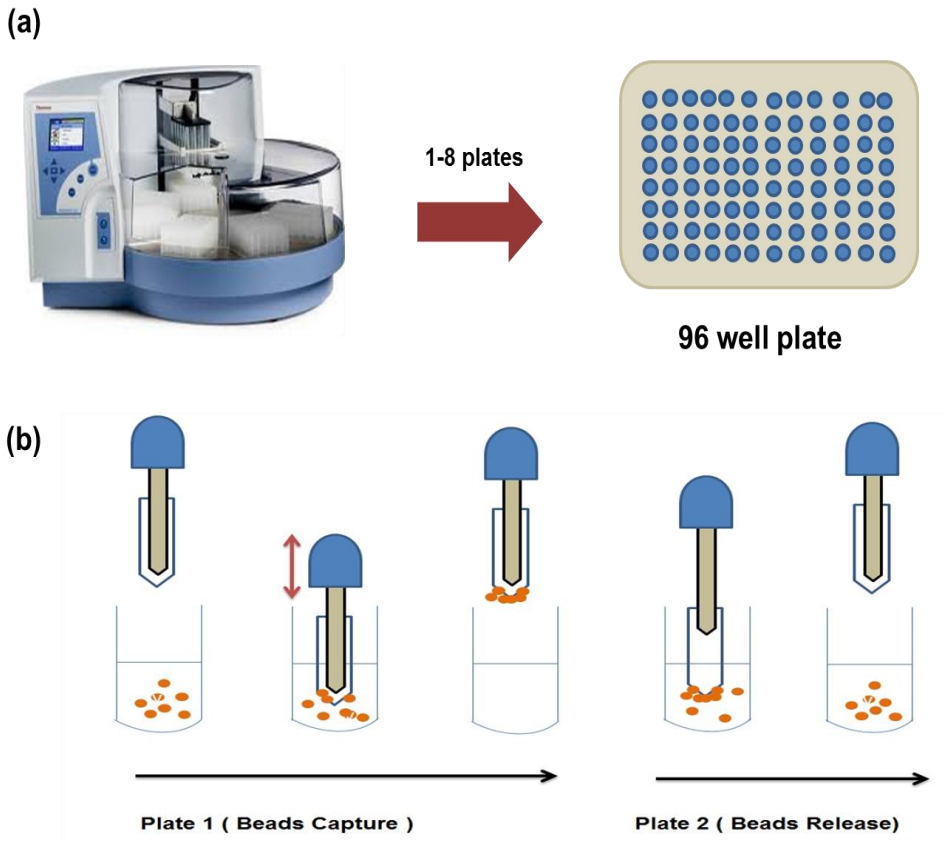

Figure 1.5 : Semi-automated panning. (a) Diagram of King Fisher Flex machine used to control automated beads panning. (b) Overall panning process from incubation to washing and plate switching is done automatically. Figure adapted from (Konthur, Wilde et al. 2010)

1.7 Statement of Problem

A vast number of commercially therapeutic antibodies approved by the U.S. Food and Drug Administration are full-size antibodies of IgG1 format at about 150 kDa size (Holt, Herring et al. 2003). Due to their relatively large-size, these molecules have limitations in terms of poor penetration into tissues (e.g., solid tumors) and eventually results in weak binding to functionally important regions (Dimitrov and Marks 2009). The use of using smaller formats can bring about therapeutic relevance. As an example, the human immunodeficiency virus envelope glycoprotein that can only be access by smaller molecules (Labrijn, Poignard et al. 2003). Therefore, by decreasing the size of the molecule it can aid tissue penetration (Yokota, Milenic et al. 1993).

Over the last decade, a large amount of work has been focused on the development of alternatives for smaller novel scaffolds. (Holt, Herring et al. 2003; Holliger and Hudson 2005; Dimitrov and Marks 2009). Amongst the most explored scaffold includes the relatively small domain antibody, which comprises of only the domain antibodies and synthetic domain antibodies for various fields of research.
Most domain antibodies are derived from camelids, sharks and murine. This is because fully human domain scaffolds of the variable gene repertoire are more likely to aggregate. Since then, human heavy chain variable fragments (VH) have been compared with those found naturally in camelids.

The determining factor for successful isolation of these antibodies relies heavily on the quality of the library generated. Among the critical factors are based on the diversity of the libraries as well as the functionality (Prassler, Thiel et al. 2011). While most of the studies conducted on the synthetic human domain antibody tackles the issue of library construction (Silacci, Brack et al. 2005), biophysical

properties such as proper protein folding (Forrer, Jung et al. 1999) and aggregation (Dudgeon, Famm et al. 2009) or diversity (Mondon, Dubreuil et al. 2008; Yang, Kang et al. 2009). To circumvent this problem, mutational studies have been conducted to understand the factors attributed to these problems.

The main focus for the antigen binding specificity lies within the CDR region of the variable domain. In the early stages, in depth studies of domain antibody sequence analysis have found that aggregation is likely to occur at the regions in or adjacent to the CDR regions. Thus, the generation of synthetic domain antibodies will allow for design of highly stable frameworks. Introducing diversity artificially will eliminate any bias introduced by the host immune system. A full control of the amino acid composition in the CDRs is possible by using the trinucleotide synthetic design. Knappik et al pioneered a rather complex library by introducing diversity in the CDR3 cassette in both variable heavy and light chain thus incorporating it in all 49 combination into phagemid vector (Knappik, Ge et al. 2000).

As more antibody sequence information was generated, several different approaches have been proposed to improve diversity. Christ and his group developed a synthetic human domain antibody library where the diversity was introduced in all three of the CDR region (Lee, Iorno et al. 2007) in the heavy variable region to be used in screening a wide array of antigen. In this study, the human domain antibody constructed will be based on a known antibody framework that is reported to have good solubility and stability (Lee, Liang et al. 2004; Mandrup, Friis et al. 2013). Similarly, the method introduced in this study is aimed to focus on the assembly of highly diverse genes of all three CDR regions with a defined single framework using single-pot synthesis. The CDR lengths were determined via analysis of the average length of CDRs naturally available.

The introduced method would help to establish a synthetic human domain antibody library with unique and diverse CDR regions for functional antigen binding by the extension of CDR-H2 and CDR-H3 distribution length. As more disease specific biomarkers are being discovered, one of the major bottlenecks for the development of diagnostic tests or even for basic research is the availability of specific antibodies against these targets.

Therefore, this study has been conducted specifically for the production of monoclonal antibodies against biomarkers with the use of a synthetic human domain library. The naïve synthetic library will be used for selection of binders against a wide range of disease specific recombinant antigen that can potentially be used in diagnostic or even therapeutics platform.

1.8 Research Objectives

1. To design, assemble and clone a collection of synthetic human antibody variable heavy genes
2. To generate a highly diverse in-house synthetic domain antibody phage display library.
3. To identify monoclonal domain antibodies for potential binders against disease specific recombinant antigen.