GENERATION OF A NAÏVE HUMAN scFv ANTIBODY LIBRARY FOR THE PRODUCTION OF HUMAN MONOCLONAL ANTIBODIES BY USING PHAGE DISPLAY TECHNOLOGY

by

LIM BEE NAR

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

SEPTEMBER 2015

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to the following individuals for their assistances and encouragements throughout my study.

First and foremost, I would like to thank my supervisor, Dr. Lim Theam Soon for his enormous support, guidance and advice. Thank you for sharing knowledge as well as the time spent for the completion of this project. Without your patience and criticisms, this project would not have been the same.

I would like to send my special thanks to my co-supervisor, Prof. Dato' Dr. Asma Ismail and Dr. Choong Yee Siew. Their assistances and suggestions for the project throughout my study are highly appreciated.

I would like to extend my thanks to all my labmates and staff members in the Institute For Research In Molecular Medicine for their help and support as well. I would also like to acknowledge the Ministry of Education MyMasters for awarding me scholarship and supporting me by financial means.

Last but not least, I would like to show my deepest gratitude and I am especially indebted to my parents for their understanding and moral support in the whole duration of this study. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

amp	ampicillin
bp	base pair
CH	constant heavy
CL	constant light
CDR	complementary determining region
cDNA	complementary DNA
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
dH_2O	distilled water
DiStRO	diversity standard of random oligonucleotide
dsDNA	double stranded DNA
ELISA	enzyme linked immunosorbent assay
E. coli	*Escherichia coli*
EtBr	ethidium bromide
glu	glucose
His-tag	histidine tag
hlyE	hemolysin E
HRP	horseradish peroxidase

hr	hour
HIV	human immunodeficiency virus
Ig	immunoglobulin
MAp17	matrix protein p17
mRNA	messenger RNA
min	minute
kan	kanamycin
KF	Klenow fragment
OD	optical density
o/n	overnight
PCR	polymerase chain reaction
RNA	ribonucleic acid
RIN	RNA integrity number
S. Typhi	*Salmonella enterica* serovar Typhi
s	second
scFv	single-chain fragment variable
ssDNA	single stranded DNA
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis

V-gene	variable gene
VH	variable heavy
VL	variable light
Ub	ubiquitin
UV	ultraviolet

LIST OF SYMBOLS

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

LIST OF PUBLICATIONS

LIM, B. N., TYE, G. J., CHOONG, Y. S., ONG, E. B. B., ISMAIL, A. & LIM, T. S. 2014. Principles and application of antibody libraries for infectious diseases. *Biotechnology letters,* 36**,** 2381-2392.

LIM, B. N., CHOONG, Y. S., ISMAIL, A., GLÖKLER, J., KONTHUR, Z. & LIM, T. S. 2012. Directed evolution of nucleotide-based libraries using lambda exonuclease. *Biotechniques,* 53**,** 357.

Biotechnol Lett (2014) 36:2381–2392
DOI 10.1007/s10529-014-1635-x

REVIEW

Principles and application of antibody libraries for infectious diseases

Bee Nar Lim · Gee Jun Tye · Yee Siew Choong · Eugene Boon Beng Ong · Asma Ismail · Theam Soon Lim

Received: 2 June 2014 / Accepted: 11 August 2014 / Published online: 12 September 2014

Abstract Antibodies have been used efficiently for the treatment and diagnosis of many diseases. Recombinant antibody technology allows the generation of fully human antibodies. Phage display is the gold standard for the production of human antibodies in vitro. To generate monoclonal antibodies by phage display, the generation of antibody libraries is crucial. Antibody libraries are classified according to the source where the antibody gene sequences were obtained. The most useful library for infectious diseases is the immunized library. Immunized libraries would allow better and selective enrichment of antibodies against disease antigens. The antibodies generated from these libraries can be translated for both diagnostic and therapeutic applications. This review focuses on the generation of immunized antibody libraries and the potential applications of the antibodies derived from these libraries.

Keywords Antibody libraries · Antibody phage display · Immunized antibody libraries · Infectious diseases

B. N. Lim · G. J. Tye · Y. S. Choong · E. B. B. Ong · A. Ismail · T. S. Lim (✉)
Institute for Research in Molecular Medicine, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
e-mail: theamsoon@usm.my

B. N. Lim
e-mail: lbeenar@gmail.com

G. J. Tye
e-mail: geejun@usm.my

Y. S. Choong
e-mail: yeesiew@usm.my

E. B. B. Ong
e-mail: eugene@usm.my

A. Ismail
e-mail: asma@usm.my

T. S. Lim
ADAPT Research Cluster, Centre for Research Initiatives - Clinical & Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

Introduction

Antibodies are a unique product of the immune system and are commonly used for diagnostic or therapeutic applications. Antibodies are highly desirable due to their specificity and affinity to a specific target antigen. However, the production of murine monoclonal antibodies (mAb) and/or antibody fragments is a challenge. The conventional approach to the production of murine mAb is the hybridoma technology by Kohler and Milstein (1975). The process requires the introduction of a specific target antigen to an animal host supplemented with adjuvants to elicit an immune response towards the target antigen. Murine B lymphocytes are fused with myeloma cells to generate hybrid cells called hybridomas. These hybridomas will have the ability to propagate indefinitely in

Reports

Directed evolution of nucleotide-based libraries using lambda exonuclease

Bee Nar Lim[1], Yee Siew Choong[1], Asma Ismail[1], Jörn Glökler[2], Zoltán Konthur[3] and Theam Soon Lim[1]
[1]Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Penang, Malaysia, [2]Alacris Theranostics GmbH, Berlin, Germany, and [3]Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Berlin, Germany

BioTechniques 53:357-364 (December 2012) doi 10.2144/000113964
Keywords: Lambda Exonuclease; Antibody; Mutagenesis; Diversity; Nucleotide-based Libraries

Directed evolution of nucleotide libraries using recombination or mutagenesis is an important technique for customizing catalytic or biophysical traits of proteins. Conventional directed evolution methods, however, suffer from cumbersome digestion and ligation steps. Here, we describe a simple method to increase nucleotide diversity using single-stranded DNA (ssDNA) as a starting template. An initial PCR amplification using phosphorylated primers with overlapping regions followed by treatment with lambda exonuclease generates ssDNA templates that can then be annealed via the overlap regions. Double-stranded DNA (dsDNA) is then generated through extension with Klenow fragment. To demonstrate the applicability of this methodology for directed evolution of nucleotide libraries, we generated both gene shuffled and regional mutagenesis synthetic antibody libraries with titers of 2×10^8 and 6×10^7, respectively. We conclude that our method is an efficient and convenient approach to generate diversity in nucleic acid based libraries, especially recombinant antibody libraries.

Directed evolution allows an in vitro mimicry of the natural in vivo evolution process to generate new or improved traits. The directed evolution process involves repetitive cycles of genotype diversification accompanied by a physical selection process to sieve out the best phenotype. The approach is synonymous with display techniques, such as phage, yeast, and ribosome display, that promote a physical linkage between genotype and phenotype. A basic requirement for any display technique is a diverse nucleotide library (1). Numerous methods have been developed to introduce diversity into these libraries, including random mutagenesis (2), in vitro or in vivo DNA shuffling (3–5) and site-specific recombination (6–8). Random mutagenesis can be carried out using several different methods, although most commonly through the use of a low fidelity DNA polymerase with $MnCl_2$ (9) and *Escherichia coli* mutator strains (10). Other methods include mutagenic polymerases (11), mixtures of triphosphate derivatives of nucleoside analogs (12), site directed mutagenesis (13), mutational hotspots (14), and parsimonious mutagenesis (15). Although random mutagenesis can generate randomization at any position in the sequence, the technique is mostly limited to short stretches of DNA. This prevents the alteration of entire binding sites, which are generally longer DNA regions. In addition, as some mutations may not be directed, errors such as frameshifts or stop codons could also occur during mutagenesis that may result in the disruption of the phenotype.

Antibody libraries are the mainstay in display approaches due to the nature of diversification present in antibody genes. Here, gene shuffling can be accomplished by numerous methods including chain shuffling (16), DNA shuffling (4) or staggered extension process (StEP) (17). Chain shuffling makes use of shuffling either the heavy or light chain variable regions of the antibody genes to generate new variants by conventional restriction digestion and ligation. In DNA shuffling, the antibody gene is digested with DNase I, randomly reassembled and amplified by PCR. StEP, which is also PCR-based, allows template switching by shortening extension times and, hence, will shuffle various portions of several parental antibody genes. DNA shuffling has been widely used for library generation, and a few recent modifications have been described that improve the efficiency of the technique including the use of single-stranded DNA (ssDNA) instead of double-stranded DNA (dsDNA) as template and use of restriction enzymes or endonuclease V instead of DNase I during DNAs fragmentation (18, 19). However, this method still presents limitations, including low frequency of chimeric genes due to preferred homoduplex formation, a limited distribution of restriction site, and a lack of high-resolution crossover (20, 21).

Lambda exonuclease is an enzyme that assists in the repair of dsDNA breaks in viral DNA (22). Lambda exonuclease is a highly processive 5′→3′ dsDNA exonuclease that selectively degrades a phosphorylated chain of the duplex to yield mononucleotides and ssDNA (23–25). ssDNA template used in directed evolution experiments is usually generated as the substrate for pairing enzymes that promote homologous recombination. The main characteristic of the enzyme is the requirement of a phosphate group at the 5′ dsDNA end. The use of lambda exonuclease to generate ssDNA has assisted in other method developments and technologies such as next-generation sequencing platforms (26), DNA-chips (27), SELEX (28), subtractive hybridization techniques (29), sample preparation for electrospray ionization mass spectrometry (30), and recombination methods (31).

PENJANAAN SEBUAH PERPUSTAKAAN ANTIBODI NAIF scFv MANUSIA BAGI MENGHASILKAN ANTIBODI MONOKLON MANUSIA DENGAN MENGGUNAKAN TEKNOLOGI PAMERAN FAJ

ABSTRAK

Perpustakaan antibodi naïf boleh mengatasi batasan kaedah tradisional hibridoma untuk menghasilkan antibodi monoklonal. Pendulangan pameran faj ialah sejenis teknik yang lebih kos efektif and menjimatkan masa untuk menghasilkan antibodi. Kepelbagaian perpustakaan antibodi naïf yang dibina amat penting bagi memastikan kejayaan dalam memperkaya antibodi terhadap pelbagai jenis sasaran antigen. Strategi baru untuk meningkatkan afiniti pengikatan antibodi diperkenalkan juga. Perpustakaan antibodi naïf scFv dibina daripada 90 orang penderma yang terdiri daripada kumpulan etnik yang berlainan (Melayu, Cina and India) dengan distribusi jantina yang sama. Semua kemungkinan V-gen antibodi diperkayakan dengan PCR konvensional kemudiannya diklonkan dalam vektor fajmid. Dua cara yang berbeza telah digunakan untuk menjana perpustakaan antibodi, iaitu perakitan PCR dan pengklonan dua langkah. Kedua-dua perpustakaan antibodi yang dibina mempunyai saiz anggaran $2x10^9$. Walau bagaimanapun, cara perakitan PCR menunjukkan kadar kemasukan sebanyak 67 % sahaja manakala pengklonan dua langkah mencapai 80 %. Perpustakaan antibodi disaring untuk mendapatkan antibodi monoklonal terhadap tiga jenis antigen, iaitu ubiquitin, hemolysin E dan HIV matriks protein p17 (MAp17). Penyaringan dijalankan dengan menggunakan dua jenis teknik, semi-automatik dan pendulangan konvensional. Antibodi monoklonal scFv manusia telah diperoleh bagi semua antigen. Penggunaan gen-V pada rantai berat dan rantai ringan bagi semua antibodi monoklonal telah dikaji. Rantai berat untuk antibodi

terhadap ubiquitin dan hemolysin E adalah terdiri daripada IGHV3 manakala rantai berat untuk MAp17 adalah IGHV1. Sementara itu, rantai ringan untuk antigen-antigen tersebut adalah IGKV1, IGLV2 and IGLV1 masing-masing. Aktiviti pengikat untuk antibodi yang diperolehi juga dinilai dalam bentuk larut. Di samping itu, satu teknik baru bagi meningkatkan kepelbagaian antibodi telah diperkenalkan dengan menggunakan exonuklease lambda untuk menghasilkan DNA untai tunggal (ssDNA) sebagai templat untuk tujuan mutagenesis. Dua jenis randomisasi telah ditunjukkan dengan menggunakan teknik ini, iaitu pangacakan rantai dan mutagenesis rawak CDR3. Keadaan yang terbaik untuk meghasilkan ssDNA adalah meggunakan 1 μg DNA dengan 10 U lambda exonuklease pada 37 ^{o}C selama 30 mins. Kesimpulannya, kepelbagaian perpustakaan antibodi naïf scFv yang dibina menghasilkan antibodi monoklonal larut terhadap antigen.

GENERATION OF A NAÏVE HUMAN scFv ANTIBODY LIBRARY FOR THE PRODUCTION OF HUMAN MONOCLONAL ANTIBODIES BY USING PHAGE DISPLAY TECHNOLOGY

ABSTRACT

Naive antibody libraries are able to overcome the limitation of traditional hybridoma method in producing monoclonal antibodies. Phage display biopanning is a more cost effective and less time consuming technology for antibody production. Construction of a naive antibody library with a large diversity is crucial to ensure successful enrichment of antibodies against various target antigens. A naïve scFv antibody library was constructed from 90 donors of different ethnic groups (Malay, Chinese and Indian) with an equal gender distribution. All possible antibody V-genes were amplified by conventional PCR method and cloned into a phagemid vector. Two different approaches were used to generate the antibody library, namely PCR assembly and two step cloning. Both antibody libraries constructed had an estimated size of $2x10^9$. However, the PCR assembly method showed an insert rate of 67 % only while two step cloning reached 80 %. The antibody library was used to screen for monoclonal antibodies against three different types of antigens, ubiquitin, hemolysin E and HIV matrix protein p17 (MAp17). Selection of monoclonal antibody was carried out using two different techniques, semi-automated and conventional plate biopanning. Human monoclonal scFv antibody was successfully obtained for all the antigens. V-gene usage of heavy chain and light chain of all the antibody clones were studied. Heavy chain of antibody against ubiquitin and hemolysin E were made up from IGHV3 whereas the heavy chain for MAp17 was from IGHV1. Meanwhile the light chain for the antigen was from IGKV1, IGLV2

and IGLV1 respectively. The antibodies enriched were also evaluated in soluble form for binding activities. In addition, a novel method to increase antibody diversity was introduced utilizing lambda exonuclease to create single-stranded DNA (ssDNA) as template for mutagenesis. Two types of randomization were demonstrated using this method, chain shuffling and CDR3 random mutagenesis. The best condition for ssDNA generation was incubating 1 μg of DNA with 10 U lambda exonuclease at 37 ^{o}C for 30 mins. In conclusion, the diverse naïve scFv library generated was able to generate soluble monoclonal scFv antibodies against the target antigens.

CHAPTER 1

INTRODUCTION

1.1 Research background

In the immune system, plasma B-cells secrete Y shape proteins known as antibodies to identify and remove foreign molecules from the body. Antibodies have been identified as the key protein in target recognition which can be exploited for many applications. This is because antibodies have an unique structure with good specificity and affinity to target antigens (Padlan, 1994). It is with this mechanism that antibodies are able to specifically identify target antigens and neutralize them. The human body is capable of producing a mixture of antibodies that is capable of binding to an array of different antigens. This collection of various target specific antibodies is called a polyclonal antibody. However, polyclonal antibodies lack in the specificity of a monoclonal antibody whereby it targets only a specific site on the target protein (Schirrmann et al., 2011).

Monoclonal antibodies were first described by Kohler and Milstein (Köhler and Milstein, 1975). The conventional method to produce monoclonal antibodies is by using the hybridoma technique. Production of hybridomas involves the fusion between spleen cells of an immunized animal to myeloma cells. This will allow the hybridoma to present a hybrid characteristic of both the myeloma and spleen cells. However, since hybridoma antibodies require the use of animals as a host and host specific myeloma cells, this method cannot be applied for human antibody production.

Recent advancement in recombinant antibody engineering has helped resolve a number of challenges associated with animal derived antibodies in order to avoid possible cytotoxic side effects when used in humans. For instance, the introduction of humanized and chimeric antibodies (Smith et al., 2004). These antibodies were developed with both murine and human DNA sequences. The past decade has witnessed the introduction of antibody phage display as a new route for generating monoclonal antibodies. Phage display derived antibodies has become a major approach as it allows the antibodies selection of highly specific binders from a highly diverse antibody library (Boel et al., 2000).

Furthermore, recombinant antibody technology allows customization of screening conditions and additional genetic manipulation for selection of high affinity monoclonal antibodies. As antibodies have high specificity and affinity towards specific target antigens, the potential of antibodies in diagnostic and therapeutic applications is enormous (Dimitrov, 2010).

1.2 Literature review

1.2.1 Immune system

The immune system is mainly involved in the defence against invading pathogenic organisms to the body (Boyden, 1966). The two main aims of the immune system that are triggered when the body encounters a pathogen is the innate and adaptive immunity. The innate immune system that serves as the first line of defence will immediately eliminate pathogens, while the adaptive immune system is initiated as a secondary immune response upon failure of the innate system to kill the pathogens. Adaptive immune system is made up of the humoral immune system (B

lymphocyte) and cell-mediated immune system (T lymphocyte) (McCullough and Summerfield, 2005).

In the humoral immune system, B lymphocytes are responsible for the production of antibodies which recognize specific target antigens. Plasma B-cells secreting antibodies are derived from B lymphocytes. B lymphocytes develop and mature into plasma B-cells when activated by the binding of antigens on its surface through the B-cell receptor (Yancopoulos and Alt, 1986). When antibodies recognize target antigens, the antibody-antigen complex is formed and agglutination will take place. This will lead to phagocytosis by macrophages or other complement immune response to get rid of the pathogens.

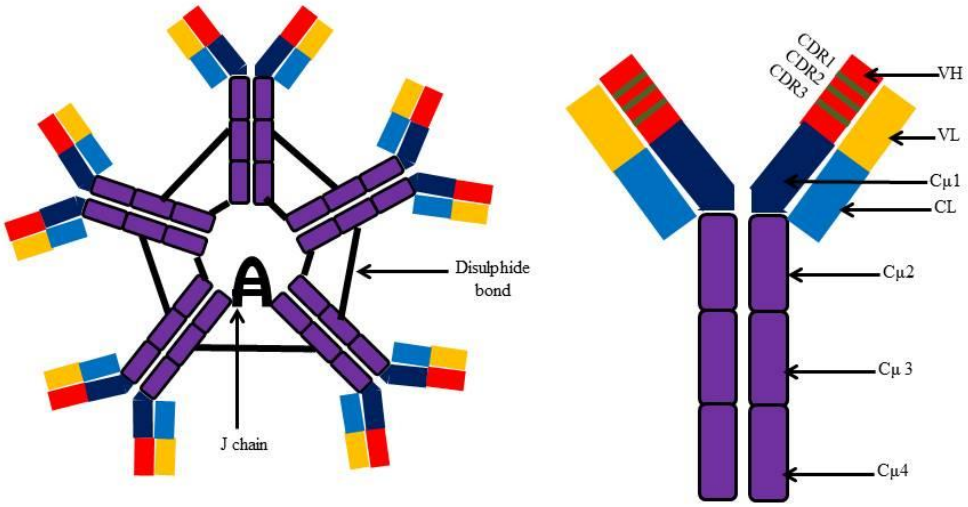
1.2.2 Antibody and its classes

Antibodies or immunoglobulins (Ig) are globular proteins produced by B lymphocytes as part of the immune system. They are secreted by plasma B-cells as soluble protein into the cellular fluid environment where it will move in circulation to identify, capture and remove pathogenic proteins (Boes, 2000).

There are five classes of heavy chains and two classes of light chains. Although antibodies comprise of both heavy and light chains, antibodies are categorized according to their heavy chain only. There are five different classes of antibodies, which are alpha (α), delta (δ), epsilon (ε), gamma (γ), mu (μ) corresponding to IgA, IgD, IgE, IgG and IgM respectively (Davies and Metzger, 1983). Depending on the class of the antibody, the heavy chain constant region can be further divided to either three or four variable domains (C_H1, C_H2, C_H3, C_H4). On the other hand, light chains can only be grouped as either lambda (λ) or kappa (κ).

All classes of antibodies have different biological functions with a distinct location in the body.

IgM is the first antibody to respond during initial exposure to an antigen. The heavy chain of the IgM monomer has one variable and four constant region domains without the hinge region (Figure 1.1). The hinge region is the area between the first and second constant region domain of the heavy chain that creates the Y shape of an antibody and provides antibody flexibility. The hinge region can be found in IgG, IgA and IgD only. Although IgM exists as a monomer, it can also form pentamers as well with a J chain. The J chain joins five IgM monomers together to form a pentameric structure and creates a total of 10 antigen binding sites (Randall et al., 1992). The pentameric structure of IgM makes it the largest antibody with approximately 970kDa in size. As the result of the unique structure, IgM possesses higher avidity compared to other classes of antibodies. As IgM is mainly produced by pre-B cell that would not have had contact with a target antigen, the repertoire of the IgM antibodies is naive. The naive nature of the repertoire would allow for the generation of highly diverse antibodies against an array of target antigens making it ideal as the first line responder during an infection. Thus, IgM plays an important role to immediately eliminate antigens in the early stage of exposure.



CDR: complementary determining regions

VH: variable heavy

VL: variable light

CL: constant light

Cμ: constant mu

Figure 1.1: Structure of an antibody molecule. The structure of immunoglobulin M (IgM) pentamer on left and monomer on right

1.2.3 The structure of antibody

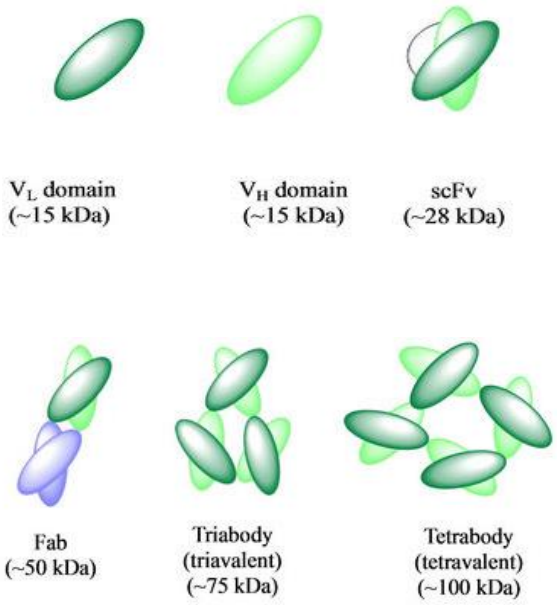
All antibodies have a basic Y shape structure consisting of an identical pair of heavy chain and light chain polypeptides. Both chains are held together by disulphide bonds and non-covalent bonds (Brinkmann et al., 1995). Each heavy chain and light chain consists of two regions; variable region (V) and constant region (C) which reflects the V_H and C_H for the heavy chain and V_L and C_L for the light chain (Edelman, 1973).

Fv (fragment variable) region of an antibody is the variable region of both the heavy chain (V_H) and light chain (V_L) that creates the antigen binding site (Skerra and Pluckthun, 1988). Fc (fragment crystallizable) region is made up of a pair of heavy chain constant domains. In short, the Fv region contains the antigen binding site while the Fc region is responsible for the recruitment of cellular actions such as complement activation.

Antibodies bind target antigens through the antigen binding site or paratope while the paratope recognizes the epitope of an antigen (Hoogenboom and Chames, 2000). The antigen binding site is made up of the combination of both V_H and V_L. Each V_H and V_L has their own alternating conserved frameworks (FR) and diverse complementary determining regions (CDR) whereby three CDRs are situated in between four FRs. These highly diverse CDRs are crucial to generate antibody's diversity.

1.2.4 Recombinant antibody

Recombinant antibodies can be presented in several different formats as shown in Figure 1.2 (Hudson, 1998, Bird et al., 1988). Fragment antibody (Fab) is a larger format of antibody consisting of the variable and constant region, where V_H-C_H (variable heavy and constant heavy) and V_L-C_L (variable light and constant light) are linked together by an interchain disulphide bond. The disulphide bond is formed from the C-terminus of light chain to a cysteine peptide on the CH1 domain or hinge between CH1 and CH2 of heavy chain (Williams and Barclay, 1988). A smaller size of the recombinant form of an antibody is termed as the single-chain variable fragment (scFv). This format of antibody only consists of variable regions, V_H and $V_{L.}$ A short repetitive stretch of glycine and serine peptide sequence is used as a linker to hold both the V_H and V_L regions together. Depending on the length of the linker, scFv molecules could form dimer, trimer or even tetramers (Hudson and Kortt, 1999). The smallest antibody fragment is single domain antibody (sdAb) or nanobody (Muyldermans, 2001). This format of antibody contains only the single variable domain either derived from the heavy chain or light chain.



VL: variable light

VH: variable heavy

scFv: single chain fragment variable

Fab: fragment antigen binding

Figure 1.2: Different formats of antibodies. Various recombinant **a**ntibody formats (Barthelemy et al., 2008)

1.2.5 Diversity of antibody

Generally, the diversity of an antibody is created at the germline level. During B-cell development, a germline repertoire of no fewer than 250 gene segments is used to generate a much larger repertoire of variable domain structures. Combinatorial rearrangement of different variable (V), diversity (D) and joining (J) gene segment takes place in the bone marrow to produce diverse primary antibody repertoires (Tonegawa, 1983). The heavy chain is generated by pairing of the V, D and J segments while the light chain has only the V and J segments. This combination creates large diversities that contribute to the unique antigen binding site or paratope of an antibody. Furthermore, somatic hypermutation, a mechanism to increase antibody diversity also takes place during B-cell differentiation (Di Noia and Neuberger, 2007). Somatic hypermutation leads to the changes of the antibody gene to create further variation and increase antibody binding affinity.

1.2.6 Antibody library

The human immune system can potentially generate 10^8-10^{12} different antibodies. To replicate this *in vitro*, a highly diverse collection of antibodies would be required to generate a representation of the *in vivo* repertoire. This is only feasible with the generation of diverse phage antibody library. There are three types of antibody libraries depending on the origin of the repertoire, which is the naïve, immunized and synthetic antibody library (Figure 1.3).

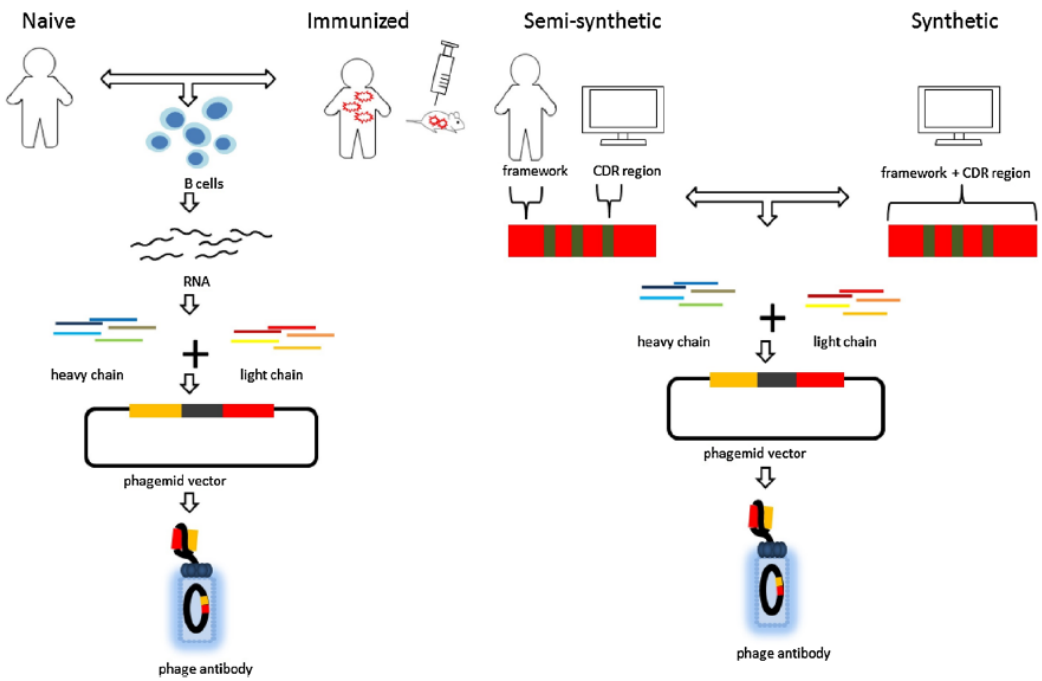


Figure 1.3: Schematic diagram of antibody library generation. Different types of antibody based libraries, naive, immunized and synthetic (Lim et al., 2014). The repertoires for naïve antibody library derived from healthy donor, the host for immunized antibody library had been exposed to disease and repertoires for semi-symthetic and symthetic library are synthetically synthesis

1.2.6.1 Naïve library

Naïve or non-immunised libraries can be constructed by amplifying IgM repertoires from B-cells of an unimmunized pool. B-cells can be harvested from the peripheral blood lymphocytes, bone marrow or spleen cell (Marks et al., 1991). In many cases, the lymphocytes can be derived from either a human or animal origin (Clackson et al., 1991). Several naive antibody libraries have been reported to be generated from human, mouse, rat, sharks and even camels. The main similarity between these different libraries is that the hosts or donors are healthy and are not infected with any diseases. The term naive reflects the nature of the antibody repertoire available or used. The repertoire would have to be unbiased and is diverse to cater for the generation of a large collection of antibodies. Naive antibody libraries are usually generated using the antibody repertoire from the IgM isotype. The IgM family has an immature repertoire that provides an unbiased characteristic which is ideal to create naïve repertoire (Dobson et al., 2005). This is because IgM is produced at the early stages of B-cell development where exposure to any foreign antigen has not taken place yet. It is this characteristic that makes IgM repertoires highly suitable for use in naive library generation. This library is very useful as large repertoires are available from a large collection of human V-genes and is rearranged *in vitro* to further enhance the diversity (Marks, 2003). Although the construction of the library is time consuming and challenging as the size of the library required must be large in diversity, the library is favourable solution for antibody development as it can be used to generate antibodies against all types of antigens.

1.2.6.2 Immunized library

Immunized libraries can also be derived from both human or animal origin, which experienced earlier exposure to a certain disease. The size of the library can be smaller in comparison to naïve libraries as this type of library consists of antibodies that are more specific against a certain target. For humans, the V-gene used to construct this library is amplified from IgG repertoires from B-cells (Azzazy and Highsmith, 2002). The library repertoires will consist of enriched antigen-specific antibodies that would have undergone affinity maturation *in vivo*. The antibodies produced would generally have a higher affinity compared to naive libraries. Even so, a major disadvantage of immunized libraries is the constant need to create a new library for every new antigen as the repertoire is biased towards a certain disease or infection (Hoogenboom, 1997). This is not efficient in terms of time and cost for general antibody production processes.

1.2.6.3 Synthetic library

This type of library is constructed *in vitro* based on a fully synthetic construct. The construction of a synthetic library is generally made up of randomized sequences in the antibody binding site of a fixed selected framework. A synthetic repertoire can be generated easily with the advancements of bioinformatic to identify stable and suitable antibody structures (Ponomarenko et al., 2008). Synthetic antibodies are made up of a fixed framework, generally the most stable ones and the diversity is introduced by synthetic oligonucleotides (Sidhu and Fellouse, 2006). The diversity generated by synthetic libraries is rather similar to that of naive libraries. This is because there is no preference in the repertoire as the diversity is generated randomly

mainly by combinatorial mixing. *In vivo*, antibody CDR segments show high variation not only in terms of amino acid propensities but also in terms of length. The variation of lengths in synthetic libraries is very challenging as opposed to a natural naive library. In certain instances, the use of fixed framework sequences and CDR randomization could be detrimental to the library quality as the diversity could be reduced in this manner. However, the process to create highly diverse synthetic libraries is possible with various advancements in molecular based methods. The HuCAL library is a good example of such a diverse library but the eventual cost to generate such a library is high and mostly unaffordable by most laboratories (Knappik et al., 2000). Therefore, naive antibody libraries are still the preferred choice for an all rounded antibody library.

1.2.7 Phage display

Phage display is a technology which was first described by George P. Smith demonstrating presentation of peptides as a fusion with the pIII protein of filamentous phage (Smith, 1985). Using phage as a host, a repertoire of random peptides can be expressed on its surface allowing a physical interaction between the phenotype and genotype. The phage presenting peptides can then be used for screening or selection to identify interaction partners. Phage display is proven to be a very powerful technique to display libraries containing millions or even billions of different peptides or proteins with their functionality still intact (Paschke, 2006). Phage display system is becoming a popular approach used to create possible binders against any target rapidly by *in vitro* selection. Several applications of phage display

includes antibody generation, enzyme evolution and interaction partner identification (Li and Caberoy, 2010, Pansri et al., 2009).

1.2.7.1 Phage biology

There are many different types of bacteriophage being used but filamentous phage is most commonly used for phage display. The filamentous phage particles mostly used for phage display is M13 which is a non-lytic filamentous phage (Barbas and Barbas, 1994). M13 infects *E. coli* containing the F pilus, for instance XL1 blue, TG1 and ER 2738 bacteria strains.

M13 phage contains a single stranded DNA genome with 6407 nucleotides coding 11 proteins. p1, p4 and p11 are involved in phage assembly and exportation. p2, p5 and p10 are important for DNA replication while p3, p6, p7, p8 and p9 make up the five types of phage coat proteins needed for phage packaging (Lopez and Webster, 1983). The major coat protein of a phage particle is p8 displaying up to 2700 copies. p7 and p9 each has 5 copies of protein as phage head protein while p3 and p6 make up the phage tail protein with each having 5 copies of proteins as well. In most phage display applications, either p3 or p8 are used for displaying peptide or protein (Rodi and Makowski, 1999).

Life cycle of M13 phage can be divided into three stages; attachment, DNA replication and exportation. Infection starts when p3 of M13 phage attaches to the F pilus of *E. coli*. Then, single-stranded DNA of M13 enters the bacteria and converts into the replicative form. Using the replicative form as template, DNA is replicated by rolling circle replication mechanism. Then, phage particles are assembled and

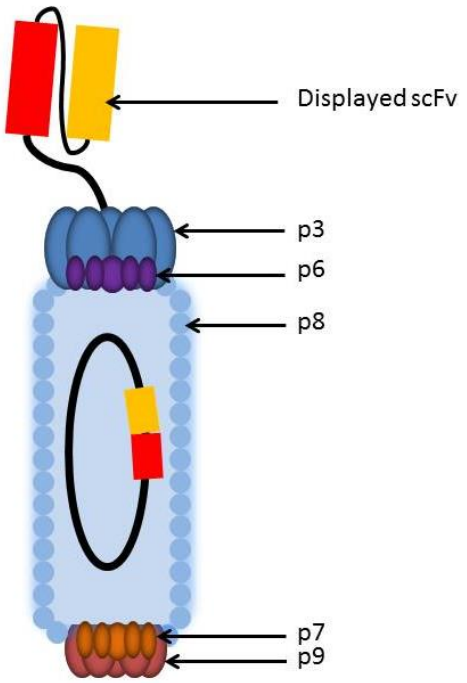
extruded out bacterial membrane without cell lysis (Marvin, 1998, Ebrahimizadeh and Rajabibazl, 2014).

The P3 is made up of two domains, the N1 and N2 domains. During infection, domain N2 will attach itself to the F pilus of the bacterial cell. Contraction of the F pilus will allow the N1 domain to bind to the TolA protein on the bacterial cell membrane. This will facilitate the entry of the phage DNA into the bacterial cytoplasm. Phage single-stranded DNA will first be converted into its replicative form by host polymerases and enzymes. This replicative form acts as template for phage gene expression and replication. Phage single-stranded DNA is then replicated using rolling circle amplification technique while phage coat proteins are expressed. Phage coat proteins will assemble at the cytoplasmic membrane when phage single-stranded DNA is ready. Finally, new phage particle is packaged and secreted from the bacterial cell without cell lysis (Russel, 1993, Rakonjac and Model, 1998).

1.2.7.2 Antibody phage display

The antibody genes encoded in the phage allows for the gene products to be displayed on the surface of filamentous phage as fusion proteins. This collection of phage is called a phage display antibody library, where each phage particle displays a single unique antibody in accordance to the antibody genotype encoded by the phage particle (Winter et al., 1994). In order to construct an antibody phage library, antibody genes are fused to phage genes mainly gIII, thus creating a link between antibody phenotype and its encoded genotype (Barbas et al., 1991) (Figure 1. 4). So, it is easier for target antibody enrichment, duplication and identification after selection.

Due to the difficulty in assembling and displaying a complete antibody fragment that has a large molecular size around 150 kDa on the phage surface, smaller antibody fragments are normally preferred. Both scFv (25-30 kDa) and Fab (~60 kDa) formats have been used successfully in constructing phage antibody library. scFv is small in size and has huge advantages for display as it is easier to present with a better production rate as a fusion protein in an *E. coli* cell. A potential problem with such fragments is the ability to form multimers which may disrupt binding by steric hindrance (Kortt et al., 1997). On the other hand, Fab is more stable since it consists of constant domains and disulphide bonds to hold the molecule together. Fab is normally chosen over the scFv format when a complete antibody structure is needed as the final product (Feldhaus et al., 2003). This is because removal of the scFv linker might result in the change of antigen binding properties. The choice of whether to produce scFv or Fab antibody library would mainly depend on downstream applications (Carmen and Jermutus, 2002).



p3: protein 3

p6: protein 6

p7: protein 7

p8: protein 8

p9: protein 9

Figure 1.4: Structure of phage particle. Phage particle has five major coat proteins (p3, p6, p7, p8 and p9). Phage particle displayed scFv antibody fragment at p3

1.2.8 Biopanning

Biopanning is a process in phage display where unique phage clones that bind to the target of interest are identified (Willats, 2002). This process usually involves consecutive rounds of binding, washing, elution and amplification. After several rounds of biopanning, affinity selection of the libraries will result in the enrichment of only antigen bound phages (Figure 1.5).

In the binding step, the antibody phage library is incubated with the immobilized target antigen to screen for possible antibody displaying phage. Target antigen is first immobilized on a solid phase, for instance immunotube, polysterene microplate and magnetic nanoparticle (Steiniger et al., 2007, Konthur et al., 2010). Target phage is captured during incubation through affinity binding. After the binding step, the wash step is carried out so that unbound phages will be washed away. Washing is very important to wash off phage that binds non-specifically to the surface of the solid support. To ensure strong binders are obtained, more washing steps and stringency in washing condition are introduced. After washing, only phage displaying antibody that bind specifically to the antigen is remained. This phage particle is then eluted. Elution of bound phage can be done by changing pH or using enzymatic cleavage (D'Mello and Howard, 2001, Corey et al., 1993). Lastly, eluted phages are amplified in *E. coli* for subsequent round of panning. Phages with higher affinity and specificity are enriched at the end of the selection round.

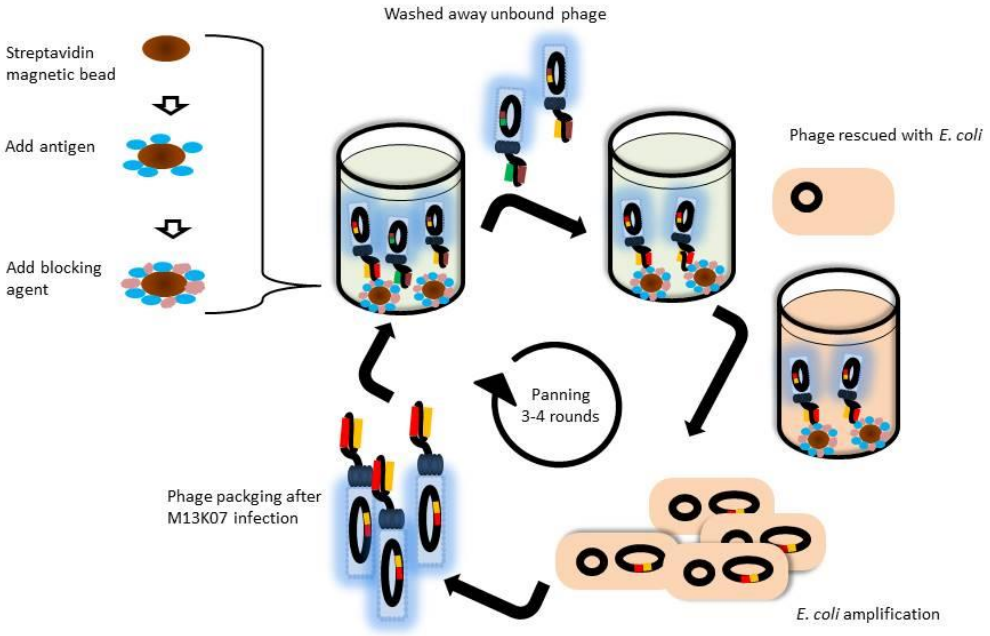


Figure 1.5: Schematic diagram of biopanning. Biopanning is a process involves four consecutive rounds of binding, washing, elution and amplification steps to enrich unique antibody clone through affinity selection

1.2.9 Antibody mutagenesis

Panning of phage displayed antibody is targeted to isolate antibody with high specificity and affinity. However, antibodies selected from naïve antibody library generally have relatively low affinities. Antibodies with low affinity may suffer from certain limitations in terms of binding capacity and specificity in diagnostic or therapeutic application. To overcome this, *in vitro* affinity maturation can be done to improve antibody binding affinity (Wang et al., 2006). Various methods have been developed to increase antibody affinity through mutagenesis such as chain shuffling and CDR mutagenesis (Lou and Marks, 2010, Drummond et al., 2005) (Figure 1.6).

In general, the heavy chain and light chain of an antibody determine its binding affinity and specificity. By shuffling between heavy and light chains or making new chain combinations, the affinity and specificity of the antibody can be altered (Christensen et al., 2009). Through *in vitro* affinity maturation, new chain combinations are useful to generate higher antibody affinity. CDR region of the antibody creates the binding site for antigen capture. The binding site is formed by a total of six CDRs, three CDRs each in the heavy chain and light chain respectively. Mutations on CDRs can change the structure of the antigen-binding site thus modifying the antibody affinity as well (De Kruif et al., 1995). The most significant affinity changes often result from mutations in the CDR3 of the heavy chain (Xu and Davis, 2000). Therefore a directed process for affinity modification is a critical step in the production and improvement of antibodies for downstream applications.

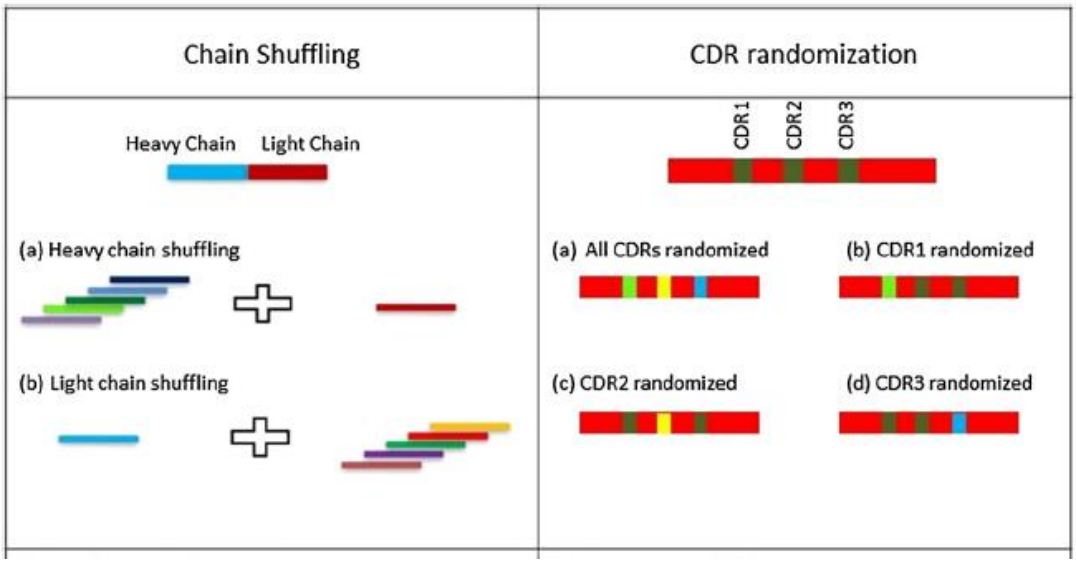


Figure 1.6: Methods of antibody mutagenesis. Two different strategies for antibody mutagenesis to increase antibody diversity (Lim et al., 2014), shuffling between heavy chain and light chain and randomization in CDR region

1.3 Rationale of study

The generation of monoclonal antibodies has traditionally involve the use of the hybridoma method. However, conventional hybridoma technique has several limitations and requires tedious animal immunisation for each desired antigen targets A more cost effective alternative for antibody production would be beneficial for long term applications. Phage display offers the possibility of producing recombinant antibodies at any given time with the generation of a naive antibody library. The method has been proven to be a rapid and robust method to generate antibodies against various antigens. By using phage display, selection of novel antibodies that bind against various types of antigens especially pathogen targets is feasible at a faster and more cost effective rate. Thus, the use of antibody phage display to identify specific antibodies for diagnostic and therapeutic applications has been widely exploited.

The success of any antibody library is largely dependent on the diversity of the repertoire generated. As genetic diversification and immune tolerance plays a large role in antibody diversity, the ability to generate a large and diverse library would require a large collection pool as well as genetically diverse backgrounds. To this extent the ability to collect samples from healthy donors of different ethnicity would provide the necessary repertoire needed for a diverse library. A naïve library would greatly benefit from this as the combinatorial mixing of antibody genes derived from different ethnic background would produce a higher diversity for a better quality library.

In order to evaluate the diversity of a naïve library, the application of a single target antigen would not suffice. Theoretical diversity stipulates the possibility of

enriching antibodies against a variety of antigens with the increased diversity. Therefore three different antigens from different categories including human, bacteria and viruses were investigated. The antigens are ubiquitin, *Salmonella enterica* serovar Typhi hemolysin E protein and HIV matrix protein p17.

Ubiquitin is a small regulatory protein with only 76 amino acids that is highly conserved in all eukaryotes. Ubiquitin functions by tagging to specific proteins for degradation. In regulatory systems, proteasome will recognize the ubiquitin tagged protein and proceed for degradation. As this protein is commonly found in eukaryotes, it appears as a good candidate to use for antibody library validation. Typhoid fever is a life-threatening disease caused by bacteria infection of *Salmonella enterica* serovar Typhi (*S.* Typhi). *S.* Typhi is a human pathogen producing pore-forming toxin, hemolysin E (hlyE). HlyE is also known as cytolysin A (ClyA) and silent hemolysin, locus A (SheA). During infection, *S.* Typhi releases hlyE leading to haemolysis by lysing erythrocytes. Since hlyE is reported to be antigenic, it is a promising target for diagnostic and therapeutic. HIV *gag* gene produces three types of polyprotein, matrix protein (p17), capsid domain (p24) and nucleocapsid domain (p15). Matrix protein (MA) p17 is a structural protein that plays an important part in the early and late stages of the HIV life cycle. MAp17 is involve in the pathogenesis of HIV disease during virus infection and replication. Thus, MAp17 is an attractive target for antibody recognition as a potential solution to combat the disease.

Antibodies with high specificity and affinity are necessary for diagnostic and therapeutic applications. However, antibodies obtained from phage antibody library may not necessary produce antibodies with high affinity. Therefore affinity maturation processes are also required to improve antibody quality much like the process that occurs naturally. Recent advancements in molecular biology methods

have made *in vitro* antibody affinity maturation faster and easier. Through affinity maturation, antibody binding efficiencies can be modified. The general strategy involving antibody maturation can be accomplished either by CDR mutagenesis of chain shuffling. A novel strategy for affinity maturation that would allow for rapid CDR mutagenesis and chain shuffling is crucial to fine tune antibody affinities after identification. Coupled together with a naive antibody library, a one-stop solution for antibody generation and modification would then be possible.

In order to generate antibodies applicable for diagnostic applications, the generation of a naive antibody library is crucial. This is because the naive nature of the repertoire would allow the use of the library for many different antigens. As such, the development of a naive antibody library with a novel method for mutagenesis is critical in the process of generating recombinant antibodies for diagnostic applications.

1.4 Objectives

The main objectives of this project were as follows:

i. To generate a naïve human scFv antibody library by using phage display technology
ii. To select phage displayed monoclonal antibodies against ubiquitin, HIV matrix protein p17 and *Salmonella* Typhi hemolysin E
iii. To produce soluble single chain fragment variable (scFv) against antigens
iv. To develop a novel mutagenesis method for antibody maturation