View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repository@USM

A COMPARATIVE STUDY OF ANTIBODY RESPONSES IN TST+ AND TST- GROUPS WITH LONG TERM EXPOSURE TO *Mycobacterium tuberculosis*

by

CHIN SIANG TEAN

Thesis submitted in fulfillment of the requirements

for the Degree of

Doctor of Philosophy

May 2017

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude and appreciation to my project supervisor, Dr. Lim Theam Soon, for his kind supervision, guidance, constructive advice and suggestions. His accessibility, knowledge, experience, encouragement and practical criticism have helped me a lot. I am also very grateful to my co-supervisors, Prof. Mohd Nor Norazmi, Prof. Armando Acosta and Prof. Maria Elena Sarmiento for their guidance, understanding, and patience in my project and paper submission.

I owe my thanks to my husband, Wang Jian Zhong for his support, love and patience. He has taken good care of me and our beloved cats, Snowy and Daisy. He companied me through both the joyful and difficult times. Without him, I can't possibly complete my study.

I would like to thank my lab mates especially Anizah and Hamizah for their kind assistance, opinion, fruitful discussion, companion and sharing good laughter. I would also like to acknowledge my friends in the Postgrad room, Mooi Kwai and Priscilla Saw, for their charming characters that enabled me to write and read pleasantly in the Postgrad room.

I would like to utter my great acknowledgement to Malaysian Ministry for Education for the funding to my project through Higher Institution Centre of Excellence (HiCOE) Grant and the financial support through Mybrain15 scholarship.

Last but not least, I wish to express my thanks and appreciation to Institute for Research in Molecular Medicine (INFORMM) for providing a nice environment and adequate equipment needed to complete my study. My gratitude also goes to the laboratory and administration staffs in INFORMM for being helpful in many ways during the progress of my study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

Ab	Antibody
ABTS	2,2'-azino-bis (3-ethylbenzthiazoline 6-sulfonic acid) diammonium
ADCC	Antibody dependent cell mediated cytotoxicity
Amp	Ampicillin
ampR	Ampicillin resistant gene (bla)
APS	Ammonium Persulphate
AVI-tag	Avidin-tag
bp	Base pair
BCG	Bacille Calmette–Guérin
BSA	Bovine serum albumin
cDNA	Complementary DNA
CDR	Complementarity determining region
cfu	Colony forming unit
CH	Constant heavy region
CL	Constant light
CSR	Class switch recombination
C region	Constant region
C-terminus	Carboxy-terminus
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleosid-5'-triphosphate
D segment	Diversity segment
E. coli	*Escherichia coli*
EDTA	Ethylendiaminotetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Fab	Fragment antigen binding
Fc	Fragment crystalline
FR	Framework
Fw	Forward (primer)
g	Gram
G	G-force
His-tag	Histidine affinity tag
hr	Hour
HRP	Horseradish peroxidase
Ig	Immunoglobulin
IPTG	Isopropyl-ß-D-thiogalactoside
J segment	Joining segment
kb	Kilo base pair
kDa	Kilo Dalton
M	Molar

min	Minute(s)
mL	Milliliter
mm	Millimeter
mM	Millimolar
mRNA	Messenger RNA
ng	Nanogram
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
OD	Optical Density
OD_{600}	OD at 600 nm wavelength
o/n	Over night
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with Tween 20
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PTM	Milk powder in Phosphate buffered saline with Tween 20
RNA	Ribonucleic acid
rpm	Revolutions per minute
Rv	Reverse (primer)
scFv	Single chain fragment variable
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate- polyacrylamide gel-electrophoresis
s	Second(s)
SSB	Single-strand DNA binding protein
SHM	Somatic hypermutation
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)-aminomethane
TST	Tuberculin skin test
Tween 20	Polyoxyethylenesorbitan monolaurate
U	Enzyme units
UV	Ultra violet
V	Volt(s)
V genes	Variable genes
V_H	Variable domain of the immunoglobulin heavy chain
V_κ	Variable domain of the immunoglobulin light chain
V_λ	Variable domain of the immunoglobulin light chain
V segment	Variable gene segment
^oC	Degree Celsius
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometer

PERBANDINGAN TINDAK BALAS ANTIBODI OLEH KUMPULAN TST+ DAN TST- YANG TERDEDAH DALAM JANGKA MASA PANJANG DENGAN *Mycobacterium tuberculosis*

ABSTRAK

Tuberkulosis (TB) merupakan penyakit yang disebar oleh *Mycobacterium tuberculosis* (MTB) melalui udara. Secara tradisinya T-sel diberikan penekanan yang utama dalam pengantaraan jangkitan MTB dengan sedikit penekanan diberikan kepada B-sel dan antibodi. Kebolehan untuk menghubungkaitkan kegunaan DNA V dan tindak balas antibodi dalam kedua-dua kumpulan TST+ dan TST- yang sihat yang terdedah kepada MTB dalam jangka masa panjang dapat memberi penerangan yang lebih teliti terhadap perlindungan TB melalui B-sel dan antibodi. TB sebagai jangkitan melalui udara, sekretori IgA yang hadir terutamanya di kawasan mukosa merupakan isotype penting yang akan memberi perlindungan berantaraan antibodi. Kumpulan TST- adalah kumpulan yang paling penting kerana ia menunjukkan tindak balas T-sel yang lemah bagi ujian TST walaupun selepas sepuluh tahun pendedahan kepada TB. Seperti yang dilihat dalam penyakit-penyakit lain, toleransi unik terhadap TB dalam kumpul-kumpulan TST mungkin disebabkan oleh himpunan antibodi tertentu yang terbentuk akibat penggabungan semula somatik VDJ. Oleh yang demikian, kajian penggunaan jujukan DNA V dilakukan untuk mengenal pasti jujukan DNA V yang paling kerap digunakan. Keputusannya, jujukan DNA V_H3-23 dan V_K1 nenunjuk kekerapan penggunaan yang paling tinggi. Jujukan DNA V_H3-23 kemudian dianalisasikan oleh penjujukan 454 amplicon untuk memberi permahaman tentang penggunaan kegemaran D dan J segmen dan perpasangan D-J. Segmen D3-10 and D3-22 masing-masing paling banyak digunakan dalam kumpulan TST- dan TST+. Segmen J4 paling kerap digunakan dalam kedua-dua kumpulan TST. Perpasangan J4 dengan D3-3 adalah disukai oleh kumpulan TST-. Perbezaan dalam genetik antibodi di antara kumpulan mencadangkan kegemaran kumpulan TST- terhadap antibodi yang menpunyai jujukan DNA V_H3-23, D3-3 dan J_H4. Penjujukan 454 hanya mampu menunjukkan kegemaran jujukan

DNA antibodi di peringkat genetik tanpa menghubungkaitkan ciri-ciri antibodi di peringkat fizikal. Oleh itu, perpustakaan IgA antibodi telah dibina dengan kaedah paparan faj. Perpustakaan antibodi ini dapat membentangkan tindak balas antibodi dalam kumpul-kumpulanTST secara keseluruhan. MTB α-kristal antigen telah dihasil untuk kegunaan dalam pemeriksaan biopanning bagi mempelajari tindak balas spesifik antibodi-antigen. TST+ E3, C5 and E12 monoklonal domain tunggal telah dikenal pasti dari perpustakaan antibodi kumpulan TST+ dan masing-masingnya mengandungi combinasi V_H5-D3-9-J4, V_H1-D1-26-J4 dan V_H3-D2-2-J3. Manakala TST- B1, B2, F5 and B8 monoklonal domain tunggal telah dikenal pasti dari perpustakaan antibodi kumpulan TST- dan masing-masingnya mengandungi combinasi V_H1-D2-2-J4, V_L1-J1, V_L2-J3 dan V_H4-D4-23-J4. Perbandingan dalam penggunanan jujukan DNA V antibodi dan tindak balas spesifik antibodi antara kumpulan TST+ dan TST- dapat memberi maklumat yang penting tentang mekanisme B-sel dalam MTB imunologi.

A COMPARATIVE STUDY OF ANTIBODY RESPONSES IN TST+ AND TST- GROUPS WITH LONG TERM EXPOSURE TO *Mycobacterium tuberculosis*

ABSTRACT

Tuberculosis (TB) is an air-borne disease caused by *Mycobacterium tuberculosis* (MTB). MTB infection is traditionally conceived as entirely T-cell mediated with little emphasis given to B-cell and antibody. The ability to correlate the V gene usage and antibody responses in both TST+ and TST- healthy groups that have prolonged exposure to MTB can provide deeper insight into the roles of B-cell and antibody in TB protection. As TB is an air-transmitted infection, the secretory IgA that exists mainly at mucosal area would be the most prominent isotype that might deliver the antibody-mediated protection. The TST- group is of particular interest because it shows weak T-cell response for TST test even after 10 years exposure to TB. An important potential protection mechanism in the TST groups could be due to the specified antibody repertoire shaped by somatic VDJ recombination as seen for other diseases. Therefore, V gene usage study was carried out to identify differences in usage of V gene segments. As a result, the V_H3-23 and V_K1 genes were shown to be highly used in TST- group. V_H3-23 gene was further analysed by 454 amplicon sequencing to provide an insight into preferential D and J segments usage and D-J pairing. D3-10 and D3-22 segments were mostly used in TST- and TST+ groups respectively. J4 segment was most highly used in both TST groups. Segment pairing of J4 to D3-3 was favoured in the TST- group. The antibodyome difference between the TST groups suggests a preference for antibody with V_H3-23, D3-3 and J4 gene usage by the TST- group. 454 sequencing can only show genotypic preference of antibody gene without correlating the phenotypic feature of antibody. Hence, IgA scFv library was constructed for TST- and TST+ groups by using phage display method. The antibody library will be the representation of overall antibody responses of the TST groups. The MTB α-crystalline antigen was generated for the use in biopanning screening to study specific antibody response. TST+ E3, C5 and E12 single domain monoclonals were identified from TST+ library and were found to use

V_H5-D3-9-J4, V_H1-D1-26-J4 with V_H3-D2-2-J3 recombination respectively. Whereas TST-B1, B2, F5 and B8 single domain monoclonals were identified from TST- group and were found to use V_H1-D2-2-J4, V_L1-J1, V_L2-J3 and V_H4-D4-23-J4 recombination respectively. The V gene usage and specific antibody response comparison between the TST+ and TST- groups can therefore provide important insights into B-cell mechanisms in MTB immunity.

INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (MTB). It remains as a major public health problem that kills almost 2 million people annually and leaves at least one third of the world's population latently infected. The Bacillus Calmettee-Guérin (BCG) is the only licensed vaccine against TB, despite widespread coverage, has failed to control the spread of TB in high burden areas (Brewer, 2000).

The control of MTB infection in a state of clinical latency indicates the existence of human natural immunity against TB. However, the precise immunological components and their interactions necessary to control or prevent MTB infection in human remain incompletely understood. Although it is established that the cellular immunity mediated by T-cell is necessary for the control of latent infection, the presupposition that such immunity is sufficient to induce long term protection has recently been challenged. In contrast to the T-cell response, the B-cell or antibody response is conventionally thought to exert little immune control over the course of MTB infection. Until recently, the protective effects in mouse challenge models with MTB suggest that specific antibodies may limit the dissemination of MTB and could potentially play a role in prevention of MTB infection via the mucosal immunity (Maglione and Chan, 2009, Acosta et al., 2013). Furthermore, antibodies are now understood to confer protection against a range of intracellular pathogens by modulating immunity via Fc-receptor mediated phagocytosis (McEwan et al., 2013, Joosten et al., 2013). However, the detail mechanisms involving human B-cell responses remain not clear, in particular on the aspect of antibody gene preferences and specific antibody response for MTB antigen. A greater understanding in these unexplored areas may provide important insights on how B-cell mechanisms are used for MTB protection.

As TB is an airborne disease, the secretory IgA antibody should be one of the key immunological elements to interact with MTB at the mucosal area of the respiratory tract to

prevent MTB infection. It has been shown that specific IgA reduces bacterial load in infected lungs of mouse models when administered intranasally (López et al., 2009, Reljic et al., 2006). Using IgA antibodies for immunotherapy of TB is promising with a report of the successful use of a novel human IgA monoclonal antibody and IgA purified from human colostrum for passive immunotherapy of TB (Balu et al., 2011, Williams et al., 2004, Alvarez et al., 2013). Even so, not much is understood about the details of the antibody gene diversification process of the IgA isotype with relation to TB.

The adaptive immune system recognizes a vast variety of antigens by its ability to boost a diverse repertoire of antibodies. The advances in high throughput sequencing technologies has enabled characterization and analysis of large antibody repertoires. This in turn helps to provide a proper representation of the abundance of antibodies and their genes that occur in the immune system. The antibody repertoire is created by three main genetic modification processes, the somatic V-D-J recombination, somatic hypernutation and affinity maturation. The patterns of somatic V-D-J rearrangement is principally important in the study of antibody variable region gene (V gene) usage. Somatic V-D-J recombination is the earliest antibody diversification process that involves random combination of germline V, D and J gene segments to generate diverse variable region genes. Ideally, somatic V-D-J is a random process. However, if gene segments are preferably used, the resultant antibody repertoire will contain a skewed V-D-J number. This gives different immune responses towards infections among individuals whereby some individuals exhibit stronger resistance toward certain infections and vice versa. Analyses of the use of immunoglobulin V genes in a number of studies have revealed a preferential antibody V gene repertoire for patients with ankylosing spondylitis (Kim et al., 2010), systemic lupus erythematosus (Kwon et al., 2005), chronic lymphocytic leukaemia (Pritsch et al., 1999) and other diseases.

There have been documented cases whereby individuals remain unresponsive to Tuberculin Skin Test (TST) despite a high and prolonged exposure to MTB (Rose et al., 1995, Rieder,

1999). The interpretation of this phenomenon has been associated more with an intrinsic resistance to infection rather than an incapability of reacting to the TST (Rose et al., 1995, Cobat et al., 2009). Healthy subjects with more than ten years of exposure to MTB were used as study subjects in the present study. The subjects were classified in the TST positive (TST+) or TST negative (TST-) group. The TST+ individuals could represent subjects susceptible to latent MTB infection but with resistance for progression to clinical TB, whereas the persistence of TST negativity in the TST- individuals could be associated with resistance to infection, with both populations exhibiting different mechanisms of protection.

To understand the B-cell mechanisms for MTB protection, this project involved systematic study of antibody genes on two different levels; the V gene usage level and the specific antibody response level. In this study, it was hypothesized that the unique TB tolerance in the TST individuals could be due to the specified antibody repertoire shaped by antibody diversification process, especially by the somatic V-D-J recombination. Therefore, V gene usage study was carried out to compare the usage of 22 different V genes between the TST+ and TST- groups. The V gene that was found to be highly used was sent for 454 amplicon sequencing to give insights into preferential D and J segments usage and D-J pairing. V gene usage study enables identification of genetic determinants that are probably associated with resistance to MTB infection. However, it remains unknown whether the genotypic preferences seen from the V gene usage study would actually correlate with the generation of protective antibodies for TB. Thus, specific antibody response study was carried out and was aimed to correlate the antibody genetic preferences seen in somatic V-D-J recombination to generation of specific antibodies that are able to bind particular MTB antigen.

The specific antibody response study involved antibody library generation and MTB antigen generation. The IgA library was constructed for each the TST+ and TST- group by using phage display technology. The concept of phage display was introduced by G. Smith based on the ability to express foreign peptides such as antibody fragment as fusion to phage coat

protein on the surface of bacteriophage (Cabezas et al., 2008). Phage display technique first involves generation of large and diverse antibody library that could mimic the natural antibody repertoire of human immune system, followed by isolation of antigen binding antibody via biopanning selection. The library represents the overall antibody responses of the TST individuals as the antibody genes were derived from the B-cell of the healthy subjects. Due to limitations in *E. coli* folding machinery, only antibody fragments such as fragment antigen binding (Fab), single chain fragment variable (scFv) and domain antibody are used for phage display library generation (Nelson, 2010). Moreover, the size of full antibody is too large to be displayed on bacteriophage. In this study, scFv was used as the antibody format for phage display library generation.

Due to its biological and immunological roles in MTB pathogenesis, the MTB α-crystalline antigen was generated in this study as a target antigen. The antibody library was screened against the MTB α-crystalline antigen for specific antibody-antigen response study with the aim to isolate different clones of MTB α-crystalline binding antibodies from each library. As phage display allows direct link of phenotypes with the encapsulated antibody genotypes, the screening of antibody library thus enables isolation of antibodies with predefined antigen specificity. To determine whether or not the antigen specific antibodies contain the genetic preferences seen from the V gene usage study, the antibodies retrieved from the biopanning selection were sent for DNA sequencing to identify their V, D and J segments usage. The isolation of MTB antigen binding antibodies encoded by preferred V-D-J recombination seen from the V gene usage study could therefore indicate protective IgA immune responses in resistant individuals by promoting the blockage of the entrance of MTB.

CHAPTER 1: LITERATURE REVIEW

1.1 Tuberculosis

Tuberculosis (TB) is an air-transmitted disease that causes serious respiratory disability in nearly one third of the world population. TB fueled by HIV infection (Law et al., 1996), kills millions of people yearly in South East Asia, Africa and other developing countries. Human TB infection is mainly caused by *Mycobacterium tuberculosis* (MTB) via inhalation of MTB droplets. MTB was first described by Robert Koch in 1882 (Smith, 1985) who received the Nobel Prize in 1905 for this discovery. MTB together with seven closely related mycobacterial species that includes *M. africanum, M. bovis, M. canetti, M. caprae, M. microti, M. mungi, and M. pinnipedii* makes up the MTB complex. MTB is a slow growing species that divides every 15 to 20 hours (hrs). It is a gram-positive bacterium as its cell wall contains peptidoglycans. It is also acid-fast due to the absence of the outer cell membrane. The high lipid content in the cell wall especially the mycolic acids (Dubel and ed, 2007) plays an important role in MTB virulence and resistance such as determining the permeability of the MTB cell surface (Holt et al., 2003).

The clinical symptoms of TB includes chronic cough with blood-stinged sputum, fever, chest pain, weight loss, loss of appetite, and night sweat. The environmental exposure to MTB does not always result in TB transmission (Nissim et al., 1994). In fact, it is estimated that only 20-30% of individuals that have come into contact with someone with active TB infection will be infected (Sharma and Mohan, 2013). In most of the infected individuals, the disease remains in a latent state where the MTB survives in balance with the immune response of the host, with just 5-10% of infection that will eventually lead to primary TB infection (Andris-Widhopf et al., 2011). The latent infection might last for the lifespan or progress to active TB disease with the risk of 5-10% of infection reactivation (Sripa et al., 2005). The reactivation of MTB infection is primary due to immunosuppressive medication (Summers, 1987) and host genetic factors (Stein, 2011). Most of the active TB infections are

pulmonary TB in which the disease develops in the lung of patients. The other minority form of the active TB infection (10%) is known as extra-pulmonary TB, in which the infection moves from the lung to other organs.

Many diagnostics are available for TB infection. For latent TB diagnosis, the tuberculin skin test (TST) is one of the generally use diagnostic tests. TST measures the size of induration caused by delayed-type hypersensitivity reaction of the memory T-cell sensitized by prior TB infection to the crude mixture of mycobacterial antigens. This TST antigen mixture contains the Bacille Calmette–Guérin (BCG) vaccine strains and non-tuberculous mycobacterial (Bartu et al., 2008). A more recent diagnostic test for latent TB infection that measures interferon Gama (IFN-ϒ) secreted from pre-exposed effector T-cell was developed. The T-cell response to the early secretory antigenic target-6 (ESAT-6), culture filtrate protein-10 (CFP-10) and TB7.7 MTB specific antigens (Rothel and Andersen, 2005) of interferon-ϒ release assay (IGRA) has been reported to give a higher degree of sensitivity and specificity (Lalvani, 2007) in low TB burden countries (Dheda et al., 2009). However, the use of IGRA assay has been controversial with the variations in gene polymorphism of human leukocyte antigen (HLA) (Hang et al., 2011, Duarte et al., 2011) and IFN-ϒ receptor (Rosenzweig et al., 2004, Henao et al., 2006, Tso et al., 2005, Hwang et al., 2007) due to ethic and/ or geographic distinctions.

1.2 MTB α-crystalline

Two small heat shock proteins (HSPs), Acr1 (α-crystalline-related protein 1) and Acr2 were reported to be highly expressed in MTB under different extreme conditions (Stewart et al., 2005, Kennaway et al., 2005). Due to its high consistency and contribution to the humoral and cellular immune responses, Acr1 (accession number Rv2031c) is a prominent antigen to be investigated for its role in MTB infection. This MTB α-crystalline protein is a useful antigen with many potential. As a heat shock protein, α-crystalline functions as a molecular chaperone that aids assembly and disassembly of protein complexes. The synthesis of α-

crystalline increases during stressful conditions to protect cells from hostile environment (Zügel and Kaufmann, 1999, Kaul and Thippeswamy, 2011). The α-crystalline also involves in the formation of an oligomeric structure that prevents aggregation of denatured proteins and subsequently aids in the refolding of other chaperones for MTB survival in the adverse environment of macrophages (Chang et al., 1996, Siddiqui et al., 2011). More importantly, α-crystalline is highly expressed during latency but not in exponential phase (Demissie et al., 2006). Thus, it is a potential candidate that could facilitate MTB survival during latent infection. Due to its biological and immunological roles, MTB α-crystalline could potentially be used for early diagnosis, vaccine candidate and for drug development.

1.3 Tuberculosis pathogenesis

TB pathogenesis is initiated by the entry of MTB to the respiratory route. MTB will travel through the mouth or nose to the nasal passage, upper respiratory tract and bronchi to enter the lung to encounter the macrophages. Interaction of MTB with macrophages drives the innate signaling pathway that leads to the recruitment of the immune cell groups by secretion of cytokines and chemokines for early granuloma formation. Neutrophils, natural killer (NK) cells and monocytes are among the early leucocytes that are recruited to the site of MTB infection. They engulf additional MTB and secrete more cytokines and chemokines. Dendritic cells then migrate to the infection site to present the MTB antigens to lymphocytes. These signaling events will lead to the hallmark of MTB infection which is the formation of granuloma structure (figure 1.3), an immune microenvironment that contains and controls the MTB infection (Nissim et al., 1994).

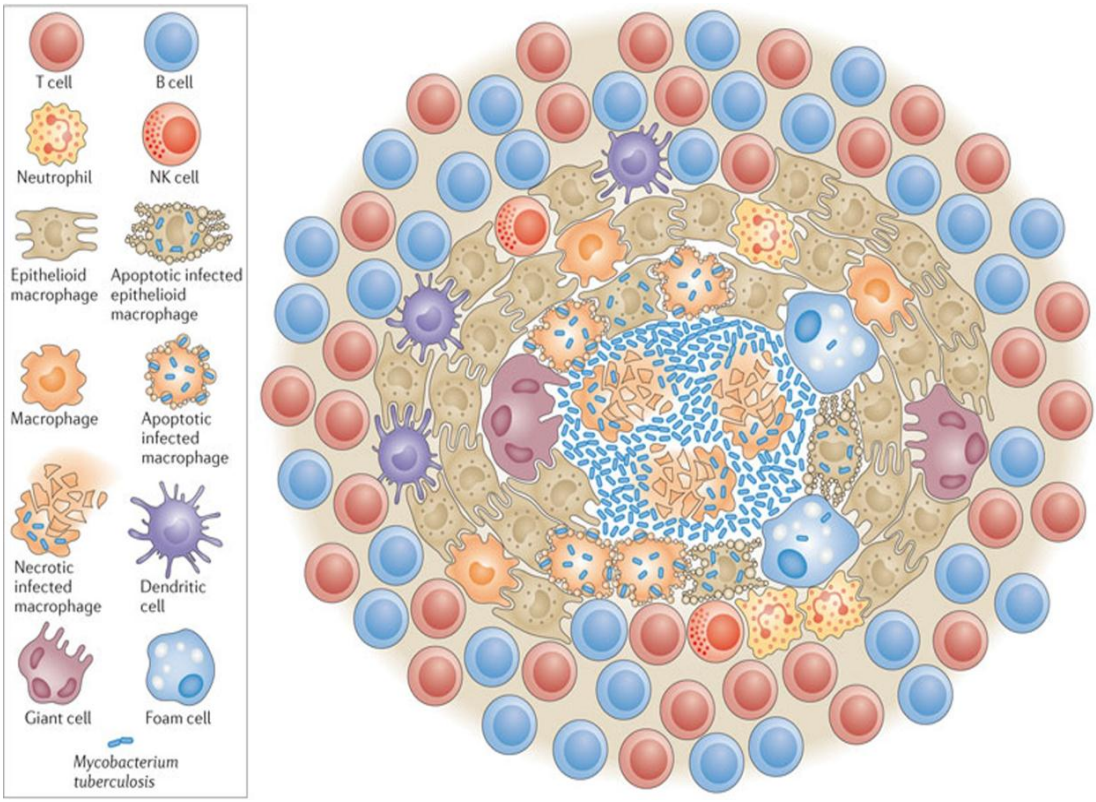

Figure 1.3. Structure and cellular components of tuberculous granuloma.

Macrophages are the principal cell types in granuloma. The differentiated macrophages are called epithelioid cells. Macrophages can fuse into giant cells or differentiate into foam cells that contain lipid bodies. Other cell types such as B and T-cells, neutrophils, NK cells, dendritic cells, fibroblasts and cells that secrete extracellular matrix components also constitute the granuloma. The death of the infected macrophages causes the formation of necrotic zone, a milky appearance in the center of granuloma with increased number of MTB. Figure was taken from (Marks and Marks, 1996).

1.4 Tuberculosis immunology

The immune system consists of both innate and adaptive mechanisms for recognition and elimination of pathogen. Early TB exposure is mediated by innate immunity, an antigen-nonspecific defense that host uses immediately after exposure to an antigen (Bradbury et al., 2003). As long as the infection is kept constant within the granuloma by adaptive immunity, TB infection will remain latent. Conversely, decline in the adaptive mechanism will lead to the failure of granuloma containment, followed by effective MTB replication, development of secondary TB disease and disease transmission to other individual (Schier et al., 1996).

Adaptive immunity in vertebrates such as humans has evolved to include both the cellular and humoral components. The cellular immunity is mediated by T-cells that target and kill infected cells directly or through activation of cytokines (Chowdhury and Pastan, 1999). Whereas the humoral immunity is mediated by B-cells that generate antibodies against a specific antigen (Hoogenboom, 1997). B-cell also helps in pathogen clearance by acting as antigen presenting cells (APCs), activating the complement system and immune cells such as neutrophils, NK cells and macrophages via antibody-dependent cell mediated cytotoxicity (ADCC) (Schier et al., 1996, Sheedy et al., 2007, Yang et al., 1995). The protection against intracellular pathogen such as MTB is generalized as entirely T-cell mediated; whereas B-cell and antibodies are perceived only for extracellular infections (Sripa et al., 2005). This follows the concept that antibodies which circularize through the serum, are dedicated for extracellular microbes. In contrast, the T-cell that only response to processed antigen can look within the cells and eliminate intracellular pathogens (Pansri et al., 2009). Therefore the T-cells and T-cell receptor (Casetti et al., 2008, Mahon et al., 2012, Plosker and Figgitt, 2003, Qiao et al., 2011) studies for MTB have been extensively illustrated with little emphasis given to B-cells and antibodies.

However, the idea of the immunological duality whereby intracellular infection is conferred by cellular immunity and extracellular infection is conferred by humoral immunity is

problematic. Studies have shown that priming of the humoral immunity is effective to prevent intracellular pathogen infection. For instance, viral infections such as influenza (Miller et al., 2005), measles (Cabezas et al., 2008) and varicella (Almquist et al., 2006) have shown to be significantly reduced or entirely eliminated after administration of antibody-based vaccine. This long term immunity against the viral pathogens is suggested to be mediated by the neutralizing antibodies that can prevent viral cellular entry and molecular interactions (Fermer et al., 2004). Other than viral infections, humoral response against intracellular bacteria and parasite infections such as MTB (Kay et al., 1996), *Plasmodium chabaudi* (Valjakka et al., 2002), and *Pneumocystis carinii* (Wu et al., 2004) were also demonstrated in B cell deficient mice studies. As such, the adaptive immunity is probably a more resourceful system that includes both the cellular and humoral responses to achieve maximally collaborative, operative, and versatile controls for pathogenic infection (Maglione and Chan, 2009). An example is the conjugate *Haemophilus influenzae* type b vaccine that combines the T-cell independent carbohydrate antigen with a T-cell dependent protein immunogen to induce long-term antibody immunity in young children (Hoogenboom, 1997, Schier et al., 1996). Another possible mechanism is by modulation of T-cell activation processes through generation of antibodies and cytokines, which can either improve or inhibit T-cell responses (Lund, 2008, Youinou et al., 2009, Winer et al., 2011, Fillatreau, 2012). Emerging evidences have also proposed the involvement of B-cell in priming memory T-cell immunity (Kennaway et al., 2005, Zügel and Kaufmann, 1999, Kaul and Thippeswamy, 2011).

1.5 Antibody basic structure

The recognition protein found in the serum and other body fluids that reacts specifically to an antigen is called an antibody (ab) or immunoglobulin (Ig). A host can produce a vast array of antibodies that are structurally similar (Y-shaped molecules) yet unique in term of specificity. The major contribution of B-cell for humoral immunity is the production of antibodies. The basic antibody structure is shown in figure 1.5. Generally, the difference in amino acids in the arms of an antibody confers its specificity towards an antigen. Whereas the base stem of an antibody is responsible for the biological activity of the antibody by recruiting immune cells and cytokines to eliminate pathogen.

Generally, each antibody consists of two light chains and two heavy chains. Each light chain pairs with a heavy chain, and each heavy chain pairs with another heavy chain by inter-chain disulfide bonds and non-covalent interactions. The two light chains (23 kDa each) and two heavy chains (50 kDa each) are identical. Each chain is divided to two regions, the variable (V) and constant (C) regions. The amino terminal contains the V region while the carboxyl terminal contains the C region. Variability of the amino acid sequences is localized within certain sections of the V region, and these sections have extensive sequence variation from antibody to antibody. These sections are called complementarity determining regions (CDRs). Moving from the amino terminal, the three CDRs are called CDR1, CDR2 and CDR3. When the light and heavy chains are joined, the CDRs will form a cleft that serves as the antigen-binding site. The amino acid sequences of the CDRs determine the shape and ionic properties of the antigen-binding site therefore directly defining the antibody specificity. Intervening sequences between the CDRs have restricted amino acid variability. These invariant segments constitute the four framework (FR) regions, which are called FR1, FR2, FR3 and FR4, and define the location of CDR. Both the light chain and heavy chain consist of multiple variable (V), numerous diversity (D) (for heavy chain only), several joining (J) gene segments and some constant region (C) exons (for heavy chain gene only). The V_κ and V_λ gene segments contain FR1, FR2, and FR3, CDR1, CDR2 and the amino

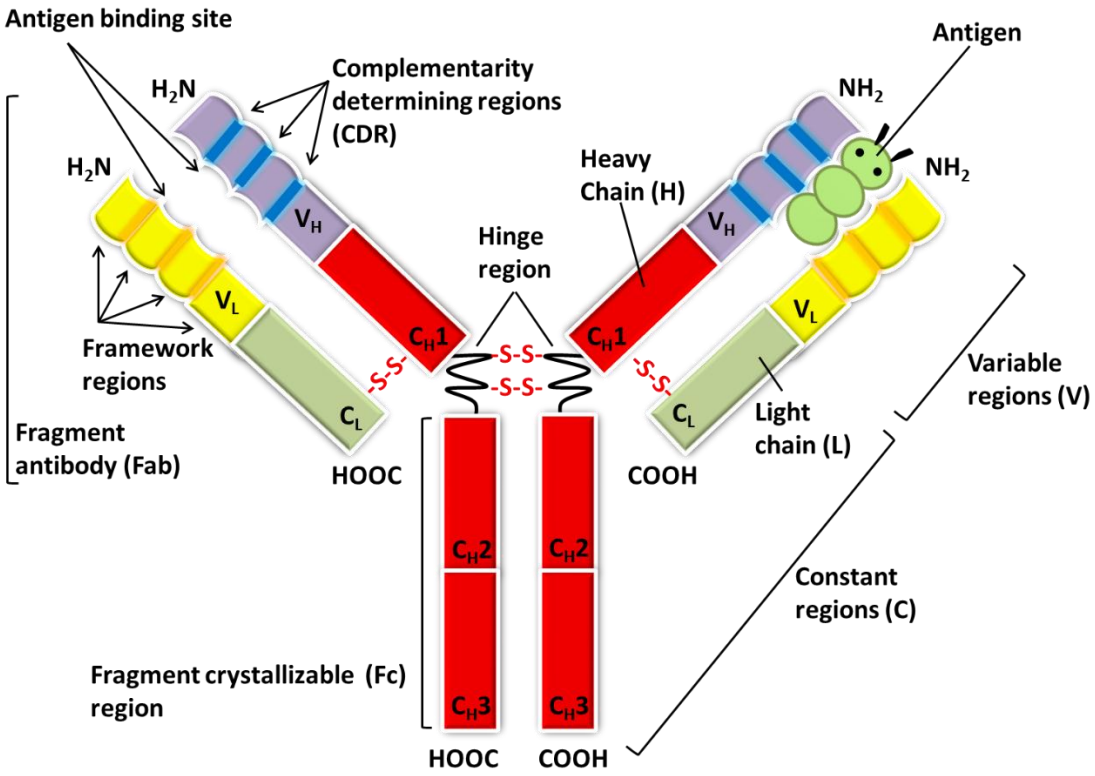

Figure 1.5. Basic IgG antibody structure.

IgG composes two identical heavy and light chains joined by disulfide bonds. Each chain has a variable (V) region that contributes to the antigen binding site, and a constant (C) region that determines the effector properties. The flexible Fab movement is due to the hinge region that joins the arms and the stem portion. Different antibody isotypes have different Fc fragment and thus different biological properties.

terminal portion of CDR3. The J_κ and J_λ gene segments contain the carboxyl terminal of CDR3 and FR4. For V_H gene segments, it contains FR1, FR2, FR3, CDR1, CDR2 and the amino terminal portion of CDR3. The D_H gene segment contains the middle portion of CDR3, and the J_H gene segment contains the carboxyl terminal of CDR3 and FR4 (Wang et al., 2005). The constant regions for both the λ and κ light chain genes are encoded as single exon. In contrast, the heavy chain gene contains exons that encode different constant regions for production of antibodies with different classes and subclasses (Choi et al., 2006).

The antibody stem region that engages biological functions such as activation of complement system, enhanced phagocytosis and ADCC is termed as fragment crystallizable (Fc) region. The IgG, IgA and IgD antibodies have a C region for each light chain and three C regions (C_H1, C_H2, and C_H3) for each heavy chain. The IgM and IgE antibodies have an additional C_H4 domain. The C_H1 domain is located within the fragment antibody (Fab), whereas the remaining C_H domains comprise the Fc region that defines the isotype of antibody. The heavy chain V region shows similarity to the V region of the light chain, while the three C-region units show strong homology to each other and to the C region of the light chain. Antibodies also demonstrate segmental flexibility, in which the two Fab portions can move relative to one another on antigen binding. The angle varies from 60 to 180 degrees. This flexible region is called hinge region and is located between the C_H1 and C_H2 domains. Only Ig G, Ig A and Ig D antibodies molecules have the hinge regions.

1.6 Antibody diversity creation

An enormous antibody repertoire must present in host to counteract with large diverse antigen that occurs naturally. Three genetic modification processes known as VDJ recombination, somatic hypermutation, and affinity maturation are responsible for the creation of antibody diversity. The earliest antibody diversification process is known as VDJ recombination that occurs at the pro-/pre-B-cell stage in which the recombinase enzyme complex catalyzes the fusion of one of the D_H gene to a J_H gene with the removal of the intervening DNA sequences (Boder et al., 2000). The recombinase then joins one of the V_H genes to the rearranged D_HJ_H gene. Terminal deoxynucleotidyl transferase (TdT) is expressed to increase the amino acid diversity of the rearranged $V_HD_HJ_H$ gene by random addition of nucleotides into the D_H-J_H and V_H-D_HJ_H joining sites. If $V_HD_HJ_H$ rearrangement produces an in frame and functional heavy chain transcript, an IgM heavy chain will be synthesized and paired with the surrogate light chain heterodimer (λ5 and VpreB proteins). Allelic exclusion is a process of expression of this pre-B-cell receptor in order to prevent V_H to D_HJ_H rearrangement on other chromosome, so that each B-cell has one unique specificity for an antigen. Once a functional heavy chain is generated, B-cell down regulates its TdT gene and initiates light chain rearrangement. Firstly, a V_κ gene segment fuses with a J_κ gene segment. If κ light chain assembly is successful, the κ light chain pairs with the heavy chain to form a functional IgM receptor on the naive B-cell surface. The rearrangement of the λ light chain proceeds if the κ rearrangement is unsuccessful (Choi et al., 2006).

Immature B-cell extends transcription of the heavy chain locus to include the exons encoding the Cδ chain for production of IgD that is downstream of exons encoding Cμ chain. IgM and IgD are co-expressed by alternative RNA splicing of the $V_HD_HJ_H$ exon to either the μ or δ exons. IgM and IgD on each of the newly mature B-cell share the same variable domains, thus have the same antigenic specificity. Mature IgM^+IgD^+ B-cell leaves bone marrow to enter the blood and migrate to the peripheral lymphoid tissues to encounter antigen. The fate and lifespan of mature B-cell are depend on antigen selection. Unstimulated B-cell lives for

short period, usually several days or few weeks. If B-cell is activated in the absence of T helper cells and without the assistance of T-cell costimulatory proteins, B-cell will differentiate into short-lived plasma cells with little B-cell memory (Wang et al., 2005). T-cell dependent stimulation usually occurs in the germinal center to increase the antibody gene diversification, including somatic hypermutation at the variable regions that leads to affinity maturation and class-switching (figure 1.6.1) in which T-cell cytokines control the antibody isotype produced (Larentis et al., 2011a). Unswitched B-cell retain the IgM and IgD expression. B-cell that undergone class switching and lost IgM and IgD expression is called switched memory B-cell (Larentis et al., 2012). Memory responses are characterized by faster production of high affinity IgG, IgA or IgE antibodies after subsequent exposure to the same antigen. Memory B-cell can be subsequently activated to become long-lived plasma cells (Boder et al., 2000). The B-cell differentiation and development is summarized in figure 1.6.2.

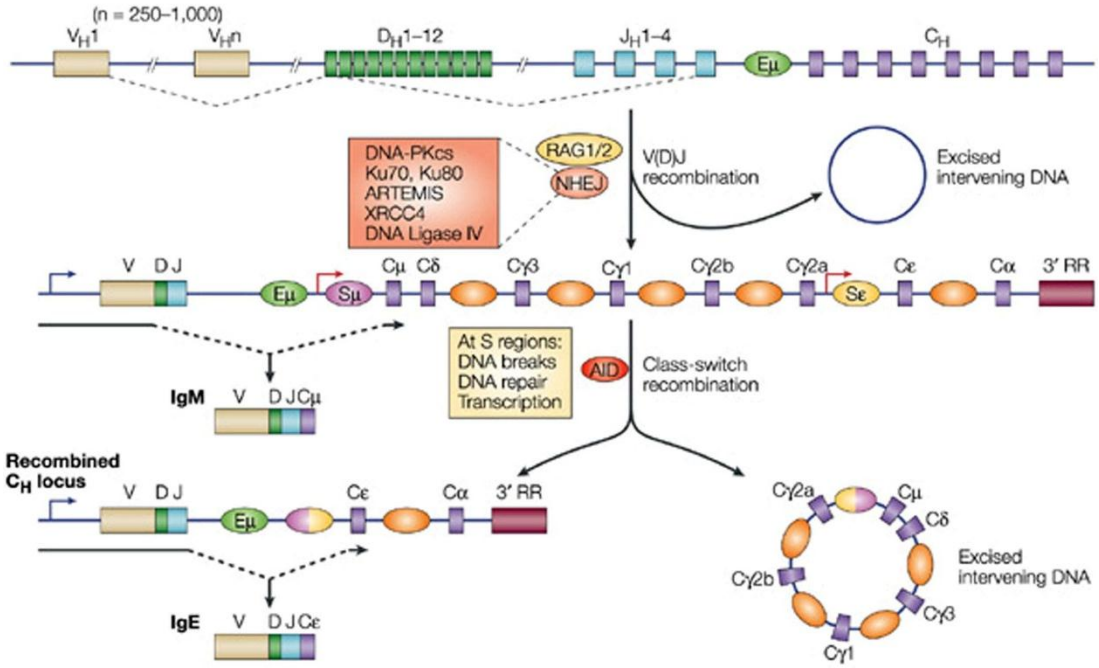

Figure 1.6.1. Somatic VDJ recombination and class-switch recombination.

Heavy chain variable domain is assembled from V, D, and J gene segments through VDJ recombination. It involves cleavage of the DNA sequences in the recombination signal by the recombination-activating gene 1 (RAG1)–RAG2 complex. Joining of the DNA breaks requires non-homologous end-joining (NHEJ) proteins, including Ku70, Ku80, ARTEMIS, X-ray repair cross-complementing protein 4 (XRCC4), DNA ligase IV and the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). Transcription is driven by a promoter upstream of the rearranged VDJ segment (blue arrow) to facilitate the synthesis of a μ heavy chain. Folowed by association with a light chain, an IgM is formed on the B-cell surface. Secondary isotypes are then produced by class-switch recombination (CSR) to exchange the constant region of the heavy chain with a set of downstream constant-region genes (CSR to IgE is shown). This deletional-recombination reaction requires the enzyme activation-induced cytidine deaminase (AID) to generate DNA breaks at switch (S) regions (located at downstream of the constant region genes), followed by DNA repair. This leads to a rearranged CH locus and deletion of the intervening sequence as an episomal circle. Cytokines stimulate transcription (red arrows) through the CH gene and determine the antibody isotype that the B-cell will switch to. The rearranged variable regions of both the heavy and light chains also undergo a high rate of point mutation through the process of somatic hypermutation (SHM) (not shown). The Emu and 3'-regulatory-region (3' RR) enhancers influence V(D)J recombination and CSR respectively. Figure was taken from (Einsfeldt et al., 2011).

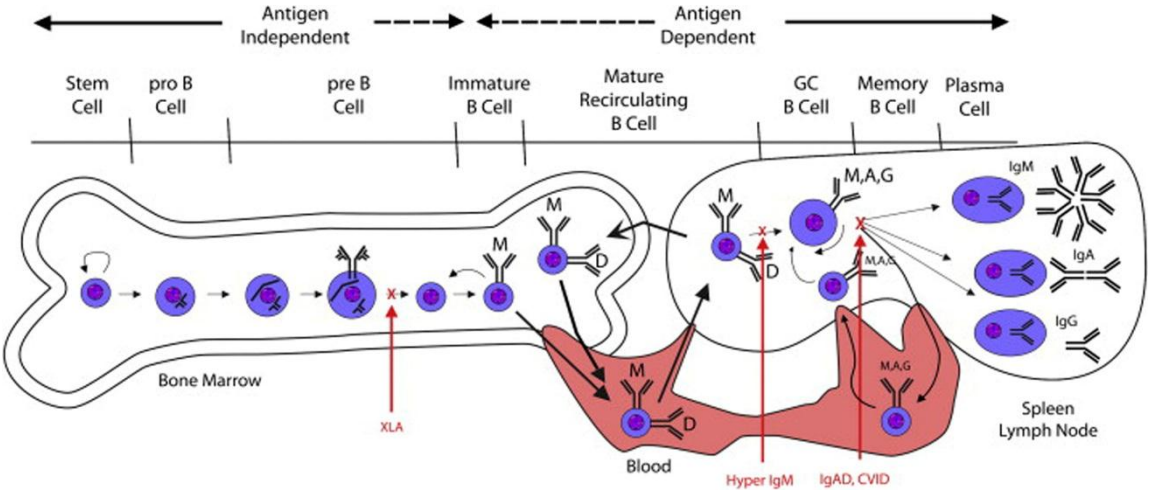

Figure 1.6.2. B-cell development and antibody diversification.

During antigen-independent B-cell development in the bone marrow, heavy chain gene rearrangement occurs first followed by light chain. The naïve B-cell expresses fully formed IgM and IgD receptors on the surface and leave the bone marrow. In peripheral lymphoid tissues, B-cell matures further under the influence of antigen and T-cell to undergo isotype switching and affinity maturation through somatic mutation. Activated B-cell enter follicles and establish germinal centers to form memory B-cell. Memory B-cell can be activated to become long-lived plasma cells in the future. Patients with X-linked agammaglobulinea (XLA) have difficulty to produce IgM and immature B-cell. Patients with hyper-IgM syndrome (Hyper IgM) cannot undergo class-switching process. Patients with selective IgA deficiency (IgAD) or common variable immune deficiency (CVID) can undergo class switch recombination but have problem to become memory B-cell or plasma cells. Figure was taken from (Wang et al., 2005).

1.7 Antibody isotypes

During early B-cell development, IgM and IgD are co-expressed on B-cell surface. During antigen stimulation and also cytokine regulation, the VDJ unit of the IgM and IgD can be associated with other isotype. This enables the antibody production with different Fc region (IgG, IgA and IgE) but the same antigenic specificity (same VDJ unit). There are five main classes of antibodies, the IgM, IgD, IgG1-4, IgA1-2 and IgE. The isotypes differ in many properties, such as size, structure, location, Fc binding and response to antigen.

IgM is the first antibody expressed on the B-cell surface. On maturation and antigen stimulation, B-cell secretes pentameric IgM with high avidity. Early expression during B-cell development allows IgM to act as natural antibody to eliminate pathogen during early infection before presence of sufficient amount of IgG to mediate of primary antibody response (Gronwall et al., 2012). IgD, similar with IgM, is expressed as antigen receptor on naïve B-cell surface. Circulating IgD is present at very low serum levels with short serum half-life (Vladutiu, 2000). As it is not known to participate in major antibody effector mechanisms, the fuction of circulating IgD is still unclear. IgG is the predominant antibody isotype and has the longest serum half-life. It confers the major antibody-based immunity against pathogen, including complement fixation, opsonisation and neutralization of toxins and viruses. IgG is the only antibody isotype that can cross placenta to give passise immunity to the fetus (Palmeira et al., 2012). Although IgE is found at the lowest serum concentration and has the shortest serum half-life, it is a very powerful antibody. It is associated with hypersensitivity and allergic reactions by binding to the FcεRIs on mast cells, eosinophils, Langerhans cells and basophils to trigger release of histamine. IgE also involves in protection against parasitic worms (Galli and Tsai, 2012).

IgA is found mainly in mucosal areas such as respiratory tract, gut and urogenital tract; and also in secreations including saliva, breast milk and tears. The serum level of IgA is higher than IgM but lower than IgG. However, IgA level is much higher than IgG at mucosal

surfaces and in secreations. IgA generally appears as monomer structure in the serum. When presents at mucosa, the IgA is termed as secretory IgA (sIgA) in a dimer structure that is accossicated with a J-chain and another polypeptide chain known as secretory component (SC). The transport of polymeric IgA (pIgA) within mucosal lymphoid tissue into mucosal secretion is through polymeric immunoglobulin receptor (pIgR). Dimeric IgA with the J-chain is bound by the pIgR on basolateral surface of lymphoid epithelial cells. Receptor-mediated endocytosis facilitates the transport of dimeric IgA through epithelial cell cytoplasm. The extracellular portion of the pIgR is cleaved to form the SC and covalently binds with the pIgA by disulfide bond. This complex of dimeric IgA with SC forms sIgA. IgA has critical role in protecting mucosal areas from toxins, bacteria and viruses by direct neutralization or prevention of pathogen binding to the mucosal surfaces. It is suggested that sIgA might act as potentiator of the immune response in intestinal tissue by mean by uptaking antigen to dendritic cells (Chew et al., 2012).

1.8 High throughput antibody sequencing

Antibody diversity is created by several mechanisms. The most fundamental one is through somatic VDJ recombination. There are 56 V, 23 D and 6 J gene segments identified in human genome (Sunitha et al., 2000). The antibody diversity is mostly focused at the site of a heavy chain VDJ gene ligation called CDR-H3. Due to the combinatorial and junctional modification mechanisms that generate CDR-H3, it is the most diverse component in terms of amino acid length and sequence and is the main determinant for antigen specificity (McBride et al., 2005, Nikerel et al., 2006, Ko et al., 2009). Comprehensive desriptions of the CDR-H3 diversity, VDJ recombination and B-cell maturation in response to infection may greatly enhance clinical diagnostic, antibody therapeutic and vaccine design.

Advance in sequencing has enabled valuable insights and provided a broader picture on the genetic basis of antibody repertoire. Recent studies used high throughput sequencing to describe antibody diversity in human (Félix et al., 2011, Lessa-Aquino et al., 2013, Arnaout

et al., 2011, Nabity et al., 2012) and zebra fish (Weinstein et al., 2009). The main purpose of high throughput sequencing is to explore the antibodyome (Lin et al., 2009) so that the antibodies expressed by different subset of B-cell in particular those that are elicited by antigen stimulation can be identified for functional characterization. Analyses of the V_κ, V_λ and V_H gene usage and recombination frequencies between particular V-J, V-D and D-J gene segment pairing in the antibody repertoire of healthy individuals (Arnaout et al., 2011, Lin et al., 2010), individuals with specific infection (Wu et al., 2011, Donovan et al., 1996, Larentis et al., 2011b), autoimmune diseases (Silva et al., 2007, Frachon et al., 2006) as well as in antibody maturation process in zebrafish (Ramos et al., 2004) have revealved remarkable skewings and variabilities that may in turn shapes the repertoire in the antigen-exposed B-cell and results in different immune responses among individuals.

Eperimental design for high throughput antibody sequencing involves several parameters (figure 1.8.1). The first consideration is the source of B-cell. Most studies have used B-cells from peripheral blood as it is readily accessible. The second consideration is to use either genomic DNA (gDNA) or mesenger RNA (mRNA) as template. Sequencing of both gDNA and mRNA enables estimation of the B-cell clonality (number of B-cell expressing a given antibody sequence), as the number of sequence read is usually proportional to the number of gene if no primer bias. The gDNA amplification commonly uses PCR primers that are complementary to the rearranged VDJ recombinants. The mRNA amplification is far more easier for rearranged V genes as the introns are removed by splicing. After reverse transcription of mRNA to cDNA, cDNA amplification can be performed either by using a 5' primer pool complementary to the FR1 of the V genes and a single 3' C_H1 (or C_κ/C_λ for light chain genes amplification) or alternatively by using the 5' rapid amplification of cDNA end (RACE) technique (Ozawa et al., 2006). The amplified sequences can then be sequenced by using suitable platform, depending on desired application and technology compatibility.