

DEVELOPMENT OF A NOVEL BEAD DISPLAY TECHNOLOGY TO IDENTIFY PROTEIN LIGANDS: APPLICATION TO IDENTIFICATION OF VIRAL ENTRY INHIBITORS AND T-CELL EPITOPES

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Trinity Term 2013

Li-Chieh Huang

St Cross College

Abstract

Li-Chieh Huang Thesis submitted for the Doctor of Philosophy in Clinical Medicine

St Cross College Hilary Term 2013

With the continued need for drug discovery and the quest to understand disease and treatment, there remains a requirement for improved methods to study protein-protein interactions and to identify potential drug leads for protein targets.

We sought to develop a new approach to directly link genotype and phenotype to use as a probe for the identification of binding partners of proteins. The method creates millions of water-in-oil emulsions, each of which functions as a micro-environment for the amplification of a library of random peptide-encoding DNA molecules, which covalently bind to a bead. Subsequently the emulsions are broken and the bead-DNA complexes are recovered, which subsequently form another emulsion with *in vitro* transcription and translation components and incubated under suitable protein synthesis conditions. The synthetic peptide is designed with tags that link to the same bead which it is translated from. In chapter 3, the detailed design and optimisation of the method will be discussed.

Cross-clade neutralising antibodies specific to HIV-1 are rare, partly because glycosylation restricts access to conserved backbone residues of gp120. In chapter 4, we hypothesized that peptides may have greater access than relatively large antibody structures, and so used our method to display random peptides on beads using a protein domain scaffold. Using a single round of selection, we identified 22 gp120-binding peptides, 4 of which were able to inhibit HIV-1 replication *in vitro*. One of the inhibitory peptides was found to bind the CCR5/CXCR4-binding site of gp120 and was able to inhibit clade B and C CCR5-tropic isolates of HIV-1. We have identified HIV-1 cross-clade neutralising peptides using a novel *in vitro* bead display library.

Comprehensive antigenic characterization of a T cell population of unknown specificity is challenging. Existing MHC class I expression systems are limited by the practical difficulty of probing cell populations with an MHC class I peptide library and the cross-reactivity of T cells that are able to recognise many variants of an index peptide. We reasoned that a bead-based display high-throughput approach may overcome these challenges. Using emulsion PCR and emulsion *in vitro* transcription/translation of a random library of peptides conjugated to β-2-microglobulin on beads, we refolded with exogenous wild-type HLA-A*0201 or CD8-null A*0201 (domains 1 and 2 of HLA-A*0201 and domain 3 of K^b with mutated residues 226A/227L). The HLA bead libraries were used to probe HLA-A*0201-restricted T cells with specificity for influenza, CMV and EBV. High-throughput sequencing was used to rank sequences of identified peptides. Compared to pre-selection sequences, we observed significant enrichment for sequences containing HLA-A2 anchors and correct viral fragments for all T cell populations. HLA bead display provides a novel approach to identify the specificity of T cells.

In summary, we combine the convenient handling of beads, the homogeneity of micro-environment in emulsion, and next-generation sequencing to create an attractive alternative to identify ligands of protein targets and antigenic peptides.

Acknowledge

I would like to take this opportunity to thank my DPhil supervisor, Professor Graham Ogg, for his continuous support in both my academic and personal life. He is undoubtedly the best mentor I have had in my life, and the role model I would strive to become. I wholeheartedly recommend anyone seeking to join his group with no reservation.

I would also like to thank all lab members in the Ogg group that I had met during my study, they were irreplaceable forces in my work and their invaluable experiences in life also helped me shape my future. I will cherish the memories of my time in Ogg group.

And I would like to give my special thanks to all members in Profess Paul Bowness group, whom I had spent 3 years with. Their talents and science expertise helped me become a better scientist and make my experiments successful. Also I would like to thank all members in Dr Tao Dong group for their excellent guidance and assistance in T-cell works; Dr Guillaume Steward-Jones group for their expertise in HIV-1 structural works; Dr Lucy Dorrell group for their HIV-1 T-cell assay works; and many others. Without their collaborations this work would not have been possible. I am forever in debt to these wonderful people. I wish them all their best in their career and life.

I would like to thank my parents, my brother and sister. Their continuous believe in me help me through countless challenges during my long journey of studies. I feel sorry that I could not physically be with them all the time.

Finally I would like to thank Buddhism. The teaching of Buddha has trained me to cope with any life obstacle that tested me. The more science knowledge I learned the more I grow to believe in Buddha's teaching. I wish that my work can be of a minute help to liberate people from pain and suffering, and I will continue to dedicate my life towards helping people.

TABLE OF CONTENTS

ABSTRACT 2

ACKNOWLEDGE 3

ABBREVIATIONS 7

CHAPTER 1 - INTRODUCTION 11

1.1.1 PHAGE DISPLAY 12
1.1.1A OVERVIEW OF PHAGE DISPLAY WORKFLOW 13
1.1.2 YEAST AND BACTERIAL DISPLAY 15
1.1.3 RIBOSOMAL DISPLAY 16
1.1.3A OVERVIEW OF RIBOSOMAL DISPLAY WORKFLOW 16
1.1.4 MRNA DISPLAY 19
1.1.5 CIS AND CAD DISPLAY 21
1.1.6 DIRECTED EVOLUTION OF PROTEINS IN EMULSION – *IN VITRO* COMPARTMENTALISATION 24
1.2 COMPARISON OF CURRENT METHODS 27
1.3 INTRINSIC PROBLEMS ENCOUNTERED IN DISPLAY METHODS 29
1.4 THERAPEUTIC PEPTIDES RE-EMERGING AS A PHARMACEUTICAL INTEREST 30
1.5 LIMITATIONS OF PEPTIDES AS THERAPEUTIC DRUG CANDIDATES 31
1.6 PEPTIDES ARE STILL A PHARMACEUTICAL INTEREST 32
1.7 SYNTHETIC PEPTIDES AS VACCINES AND DRUGS AGAINST INFECTIONS AGENTS AND CANCER 33
1.7.1 SYNTHETIC PEPTIDES AS CANCER, ALLERGY AND VIRAL VACCINES 33
1.8 THE AIM - DESIGNING A NEW APPROACH TO THE *IN VITRO* DISPLAY 34

CHAPTER 2 - MATERIALS AND METHODS 37

2.1 SYNTHESIS OF DNA TEMPLATE CONTAINING RANDOMISED PEPTIDES 38
2.2 PROTEIN EXPRESSION - HLA-A2/B8 HEAVY CHAINS, B2M LIGHT CHAIN 38
2.3 EMULSION PCR LINKING EMULSION *IN VITRO* TRANSCRIPTION/TRANSLATION 39
2.4 TOPO-CLONING – EXTRACTING SEQUENCE INFORMATION OF THE EPITOPES THAT BIND TO T-CELLS 41
2.5 WESTERN BLOT 42
2.6 GP120 SYNTHESIS 43
2.7 BEADS SELECTION 43
2.8 PEPTIDE ANALYSIS AND SYNTHESIS 44
2.9 PEPTIDE SELECTION VIA AMINO ACID SEQUENCE ANALYSIS 44
2.10 ELISA (ANTIBODY BLOCKING) 44
2.11 ELISA (PEPTIDE BLOCKING) 45
2.12 SURFACE PLASMON RESONANCE (BIACORE) 45
2.13 HIV-1 INHIBITION ASSAY 46
2.14 HLA-PEPTIDE LIBRARY 47
2.15 PBMC SEPARATION AND SETTING UP ANTIGEN-SPECIFIC T-CELL LINE 49
2.16 ENRICHMENT IN THE SEQUENCES CONTAINING HLA-A2 ANCHORS 49
2.17 ENRICHMENT OF THE PEPTIDE FRAGMENTS AFTER T CELL SELECTION 51

CHAPTER 3 - DESIGN AND OPTIMISATION OF ORBIT TECHNOLOGY 53

3.1 Introduction 54
3.1.1 The optimisation of emulsion PCR linking emulsion *IN VITRO* transcription translation 54
3.1.2 Peptide library generation: 1. Creating DNA-bead library by emulsion PCR 54
3.1.3 Peptide library generation: 2. Synthesis of peptide library by emulsion In vitro transcription and translation 56
3.1.4 Peptide library generation: 3. Use the bead library directly: selection of high affinity binding peptides from the bead library 58
3.1.5 Peptide library generation: 3. Use the bead library directly: generation of refolded MHC class I molecules on the bead and the selection of antigens recognised by antigen-specific T-cell clones 59
3.1.6 Construction of randomised peptide library with T7 promoter systems 61
3.2 Results 64
3.2.1 Emulsion PCR – making water-in-oil emulsions 64
3.2.2 Emulsion PCR – dual labelled biotinylated primer 69
3.2.3 Emulsion PCR – Dynabead C1, M270, MACS for emPCR 71
3.2.4 Emulsion PCR – all templates designed were capable of attaching on the beads 72
3.2.5 Emulsion PCR – Restriction digestion showed C1 bead had most DNA templates attaching on the beads. 73
3.2.6 Emulsion PCR – prolonged annealing time in PCR cycle is required for emPCR amplification 75
3.2.7 Emulsion PCR under various conditions 77
3.2.8 *IN VITRO* transcription and translation 79
3.2.9 emIVTT - Translating proteins from DNA templates linking with beads 80
3.2.10 Generation of BZLF1 epitope-specific HLA-B8 refolded MHC tetramers 82
3.2.11 the refolding of HLA-B8 MHC molecules on the beads after emIVTT - FACS 84
3.2.12 The refolding of HLA-B8 MHC molecules on the beads after emIVTT – FACS 85
3.2.13 Beads linking refolded MHC class I complexes created antigen-presenting multimers 87
3.3 Discussion 88
3.3.1 Emulsion bead library as an option for next generation peptide selection 88
3.3.2 Potential problems encountered by emulsion PCR linking emulsion IVTT methods 89

CHAPTER 4 - SELECTION OF HIGH AFFINITY PEPTIDES WITH EMULSION TECHNOLOGY – TARGETING OF HIV-1 COAT PROTEIN GP120..... 91

4.1 Introduction 92
4.2 Results 97
4.2.1 Validating the binding of predicted peptides to bovine serum albumin 97
4.2.2 Validating the binding of predicted peptides to gp120 - ELISA 99
4.2.3 Validating the binding affinity of predicted peptides to gp120 – surface plasmon resonance 102
4.2.4 HIV-1 viral inhibition – determining the optimum HIV infection 104
4.2.5 HIV-1 viral inhibition with streptavidin-conjugated peptides – HIV-1 BaL 105
4.2.6 HIV-1 viral inhibition with streptavidin-conjugated peptides – HIV-1 clade C 106
4.2.7 HIV-1 viral inhibition with streptavidin-conjugated peptides – BaL, clade C, IIIB 107
4.3 Discussion 112

CHAPTER 5 - HIGH-THROUGHPUT IDENTIFICATION OF CD8+ T CELL SPECIFICITY USING HLA-PEPTIDE BEAD LIBRARIES..... 115

5.1 INTRODUCTION 116
5.2 RESULTS 119
5.2.1 HLA-PEPTIDE LIBRARY 119
5.2.2 ENRICHMENT IN THE SEQUENCES CONTAINING HLA-A2 ANCHORS 122
5.2.3 ENRICHMENT OF THE PEPTIDE FRAGMENTS AFTER T CELL SELECTION 127
5.3 DISCUSSION 129

CHAPTER 6 - GENERAL DISCUSSION..... 131

6.1 ORBIT AS A PROMISING ALTERNATIVE TECHNIQUE FOR LIGAND SELECTION 132
6.2 TECHNICAL DIFFICULTIES ENCOUNTERED BY EMULSION TECHNOLOGY 133
6.3 FUTURE PERSPECTIVE OF ORBIT 135

REFERENCES 136

APPENDIX 142

Abbreviations

Aa – amino acid

Ab – antibody

Ag – antigen

ALP – alkaline phosphatase

APC – Allophycocyanin

APC – antigen presenting cell

Beta-2-m (β2m) – βeta-2 microglobulin (MHC class I light chain)

Beads – general reference to Dynabead MyOne C1, Dynabead M270, Mitenyl MACS magnetic beads

CCR5 – chemokine receptor type 5

CXCR4 – chemokine receptor type 4

CMV - Cytomegalovirus

CTL – Cytotoxic T Lymphocyte

DMSO – Dimethyl sulphoxide

DNA –deoxyribonucleic acid

EBV – Epstein-Barr virus

EDTA – Ethylenediamine tetraacetic acid

ELISA – Emzyme-linked Immunosorbent Assays

ELISpot – Enzyme-linked immunospot assay

ER – Endoplasmic reticulum

EtBr – Ethidium bromide

Emulsion – oil mixture made up of mineral oil, surfactants (e.g. Span 80, Tween 80)

emPCR – performing PCR reactions under water-in-oil conditions

emIVTT –coupled *in vitro* transcription/translation reactions under water-in-oil conditions

FACS –Fluorescence Activated Cell Sorting

FCS – Foetal Calf Serum

FITC – Fluorescine isothiocyanate

Gp120 – HIV coat protein

HA - hemagglutinin

HIV – Human Immunodeficiency Virus

HLA – Human Lymphocyte Antigen, HLA- subtypes e.g. HLA-B8, HLA-A2

ICS – intracellular cytokine staining

IFN - interferon

Ig - Immunoglobulin

IL – interleukin e.g. IL-2

mAb – Monoclonal Antibody

MFI – mean fluorescent intensity

MHC I– Major Histocompatibility Complex molecule class I

NK – natural killer cell

ORBIT – Oxford Bead *in vitro* display system

PBMC – peripheral blood mononuclear cells

PB – Pacific Blue

PBS – phosphate buffered saline

PE – phycoerythrin

PerCP – Peridin chlorophyll protein

PHA - Phytohemagglutinin

Phage – bacteriophage, viruses that infect and replicate within bacteria

R10 – RPMI with 10% heat inactivated FCS

RNA – ribonucleic acid

RT – room temperature

SBP – streptavidin binding protein

TCR – T-cell receptor

Amino acid abbreviations

A Alanine

C Cysteine

D Aspartic acid

E Glutamic acid

F Phenylalanine

G Glycine

H Histidine

I Isoleucine

K Lysine

L Leucine

M Methionine

N Asparagine

P Proline

Q Glutamine

R Arginine

S Serine

T Threonine

V Valine

W Tryptophan

Y Tyrosine

Publications

1. **L Huang***, X Pan*, H Yang, L Wan, G Stewart-Jones, L Dorrell, G Ogg. Linking genotype to phenotype on beads: high throughput selection of peptides with biological function. Scientific Reports, 2013. In Press.

2. X Pan*, **L Huang***, T Dong, Y Peng, V Cerundolo, S McGowan, G Ogg. High-throughput identification of CD8+ T cell specificity using HLA-peptide bead libraries. Journal of Immunology Methods, 2013. In Press.

3. G Malavige, **L Huang**, M Salimi, L Gomes, S Jayaratne, G Ogg. Cellular and cytokine correlates of severe dengue infection. PLoS One, 2012, 7(11): e50387

4. M Salimi, J Barlow, S Saunders, L Xue, D Gutowska-Owsiak, X Wang, **L Huang**, D Johnson, S Scanlon, A McKenzie, P Fallon and G Ogg. A role for IL-25 and IL-33-driven type-2 innate lymphoid cells in atopic dermatitis. Journal of Experimental Medicine, 2013. In Press.

*Contribute equally

Chapter 1 - Introduction

1.1 *In vitro* display technologies

In vitro display techniques such as phage display provide a large protein library from DNA random inserts which allow the rapid selection of high affinity peptide/protein to a target of interest (Smith 1985). Since the method was introduced more than 20 years ago, significant improvements have been made and have led to the development of a wide array of proteins expressed on phage such as antibody-fragment-based reagents specifically targeted to disease relevant proteins (Holliger and Hudson 2005; Hoogenboom 2005). The advantage of linking the peptide to the encoding DNA is that methods available to sequence the DNA are far more advanced than those available to sequence protein. In particular, DNA sequencing is much more rapid and is cheaper than protein sequencing. DNA sequencing can be successful on very small samples with short length and low concentration. Furthermore, DNA samples can be easily amplified to provide more DNA if needed.

To date, a number of approaches have been developed that link genotype to phenotype in an analogous way to phage. The main techniques utilising *in vitro* display technology can be categorised briefly as follows: the selection of peptide/protein fragments by phage display (Hoet, Cohen et al. 2005), the display of nonimmunoglobulin domains by ribosome display (Binz, Amstutz et al. 2004), the identification of protein-protein interactions by mRNA display (Horisawa, Tateyama et al. 2004), the generation of stable DNA-protein complexes by covalent DNA display (Sumida, Doi et al. 2009),. These *in vitro* display techniques may differ in size, vectors or reaction environments, but they all share two common features: compartmentalisation, i.e. the ability to generate unique proteins from its encoding DNA in an isolated environment; and the ability to link phenotype (translated proteins) and genotype (encoding DNA).

Phage, yeast, ribosome, mRNA, and DNA displays achieve the linkage through covalent bonds. Exclusively, *in vitro* compartmentalisation achieves a cell-like micro-environment by creating a water-oil emulsion (Aharoni, Griffiths et al. 2005). The aqueous phase mimics the cellular cytoplasm whereas the oil phase acts as the cell plasma membrane. *In vitro* display technology effectively supersedes natural in-cell selection, which relies on the association of the fundamental diversification to the mutated gene, which is passed on to the following cell generations.

1.1.1 Phage display

The most widely used display technology is phage display (McCafferty, Griffiths et al. 1990; Smith 1991), which in essence links the displayed protein (binding activity) and its encoded sequence (genotype) within the same bacteriophage particle. Each particle displays one or more copies of a unique protein fused to either its coat proteins on its surface, allowing the enrichment of ligands with high specificity and affinity *in vitro*. Phage display is widely used for affinity maturation as discussed by Fernadez-Gacio et al (Fernandez-Gacio, Uguen et al. 2003). Phage display has been used to isolate very high-affinity antibodies from human Fab libraries with synthetic diversification in the key antigen binding regions even without the need for affinity maturation (Hoet, Cohen et al. 2005). Synthetic diversity in heavy chain complementarity determining regions 1 and 2 (CDR1 and 2) was combined with naturally occurring diversity in the heavy chain CDR3 and light chains. Affinity could be obtained in the low nanomolar range, which improved after conversion from Fab to whole IgG, possibly due to avidity. The requirement of a bacterial transformation step in phage display library construction imposes a technical limit on the actual size of library generated. No such limit is placed on the other *in vitro* display approaches. In theory, the ability to search through a

larger amount of sequence variations should require fewer rounds of selection to access very high-affinity ligands without resorting to molecular evolution and affinity maturation.

1.1.1a Overview of phage display workflow

Phage display involves the display of a protein library on filamentous phage, with the selection based on binding (Smith 1985). A brief cycle of phage display is shown in **Figure 1.1**. Phages are viruses that infect bacteria and utilise the bacterial replication machinery to multiply. To do so, bacteriophage such as phage M13 penetrates the bacterial outer membrane using the minor coat protein P3 or the major coat protein P4 to attach the bacterial filamentous pilus. The interaction facilitates the opening of the bacterial cell membrane, allowing the viral single-stranded DNA to enter the cytoplasm. The complementary cDNA strand is synthesised, and the protein-coding genes are expressed and directed to the surface of the transfected cell. The phage library (approximately 10^6-10^{11}) is designed as a synthetic protein carrying all or part of the major or minor coat proteins. The expressed proteins on the surface will be selected by affinity purification. With enough rounds of purification (approximately 2-3 rounds of panning), only proteins with the highest affinity will remain, and the genes encoding this protein can be studied.

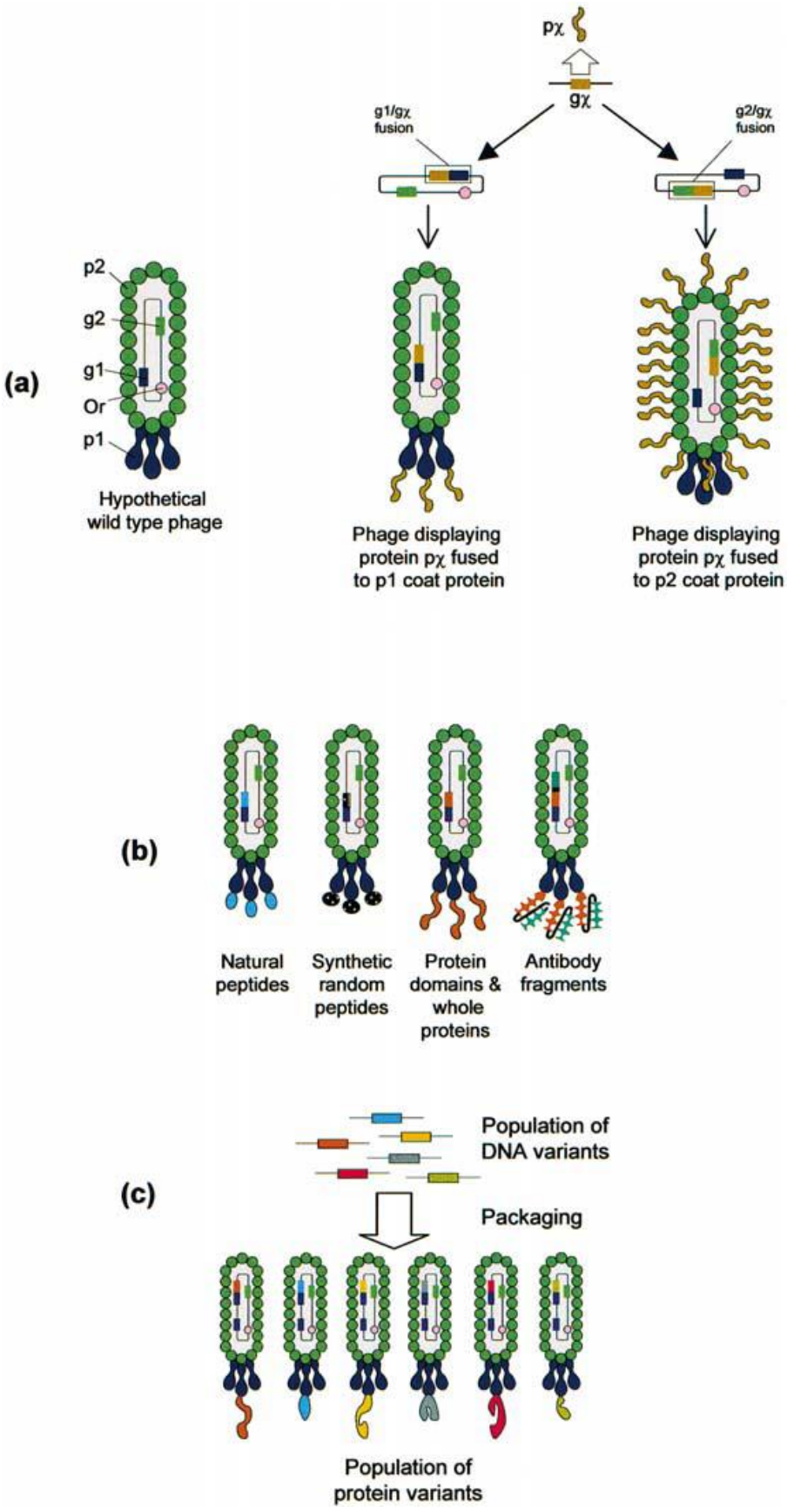

(a)
pχ
gχ
g1/gχ fusion
g2/gχ fusion
p2
g2
g1
Or
p1
Hypothetical wild type phage
Phage displaying protein pχ fused to p1 coat protein
Phage displaying protein pχ fused to p2 coat protein
(b)
Natural peptides
Synthetic random peptides
Protein domains & whole proteins
Antibody fragments
(c)
Population of DNA variants
Packaging
Population of protein variants

Figure 1.1 The principle of phage display. (a) the simplified hypothetical bacteriophage shown has a genome that contains an origin of replication (ori) and genes (g1 and g2) that encode two types of coat protein p1 and p2 respectively. A foreign protein px, that is encoded by gene gx, may be displayed at the phage surface by the fusion of gx to one of the phage coat protein genes. The number of copies of px displayed is related to which phage coat protein (p1 or p2) is chosen as a fusion partner. (b) this principle can be appied to the expression of natural or random peptides, protein domains or antibody fragments. (c) using phage display, a library of variant nucleotide sequences can be converted into a library of variant peptides or proteins. The phage display library may then be screened in order to isolate phage displaying peptides or proteins with desired properties. Adapted from ((Willats 2002)

1.1.2 Yeast and bacterial display

Following a similar strategy to phage display, yeast cells can be utilised as machinery cells to create *in vitro* display protein libraries (Wittrup, Tybjaerg-Hansen et al. 1997). The idea is to exploit the adhesion subunit of the yeast agglutinin protein Aga2p and design a protein library (approximately 10^7-10^9) that is co-expressed with Aga2p. When this gene is translated and transported to the yeast surface, disulfide bonds are formed with Aga1p, which presents the fusion protein from the yeast surface. The selection process is achieved by using fluorescently-labelled binding partners, and the highest affinity protein-yeast can be isolated by Fluorescence-activated cell sorting (FACS). Targets such as growth factors, cell surface receptors and antibody fragments have been studied using this method (Gai and Wittrup 2007). Bacterial systems using similar methodology have also been validated (Jose 2006). Libraries of peptides or proteins can be fused with bacterial flagella or outer surface protein expressed on the surface of Gram+ or Gram- bacterial cell surface. Similar to yeast display, the selection of bacterial displayed binding ligands can be achieved through cell sorting.

1.1.3 Ribosomal display

Ribosomal display was developed in 1994 (Mattheakis, Bhatt et al. 1994), in which the researchers proposed a cell-free transcription and ribosome-based translation system generates stable ternary complexes of mRNA, ribosome and protein totally *in vitro* (Netzer and Hartl 1997; Sawata, Suyama et al. 2004). No transformation of living cells is required, allowing generation of very large libraries of up to 10^{14} members. A C-terminal spacer sequence reduces steric hindrance between the display protein and the ribosome and allows correct protein folding. Transcription and translation strategies using rabbit reticulocyte lysate or E coli extract are compared by Hanes et al (Hanes, Jermutus et al. 1999). One selection cycle can be readily performed in less than 24 hours. Reports had shown that the introduction of mutations through ribosomal display was possible for affinity maturation (Hanes, Schaffitzel et al. 2000; Rothe, Hosse et al. 2006). Recently, ribosome display also advanced the display of alternative non-immunoglobulin scaffolds with different surface topologies to generate diversity (Binz, Amstutz et al. 2005). For detailed protocols on ribosome display, see He and Taussig (He and Taussig 2005), and Hanes et al (Hanes, Jermutus et al. 1999).

1.1.3a Overview of ribosomal display workflow

A brief cycle of ribosomal display is shown in **Figure 1.2**. The construction of a protein library for ribosomal display typically involves the insertion of a random DNA library in the form of PCR products, which codes for binding proteins into a ribosome display vector, which contains a T7 promoter and a ribosome binding site at the 5' region, and a spacer sequence at the 3' region of the vector, flanking the random DNA library. After insertion, the template containing the T7-RBS-library-spacer sequence is amplified by PCR, and thus a

library of linear templates coding for a library of protein is obtained. Millions of unique mRNA sequences can be transcribed from the library by *in vitro* transcription. When *in vitro* translation occurs, as the template lacks a stop codon and T7 termination sequence at the 3' end, ribosomes recruited for protein translation will be retained at the end of the mRNA and are unable to release the translated protein. This process generates mRNA-ribosome-protein complexes which can be used to select high affinity binding ligands against immobilised proteins. After the selection process (which typically involves multiple rounds of PBS washes), the sequences of high affinity binding ligands can be obtained by collecting the complexes and performing reverse transcription PCR. The PCR products obtained can be used directly to start a new round of ribosome display selection against the same target, thus achieving affinity maturation. Alternatively, the PCR products can be cloned into expression vectors and analysed by growing single colony on agar plates (Zahnd, Amstutz et al. 2007).

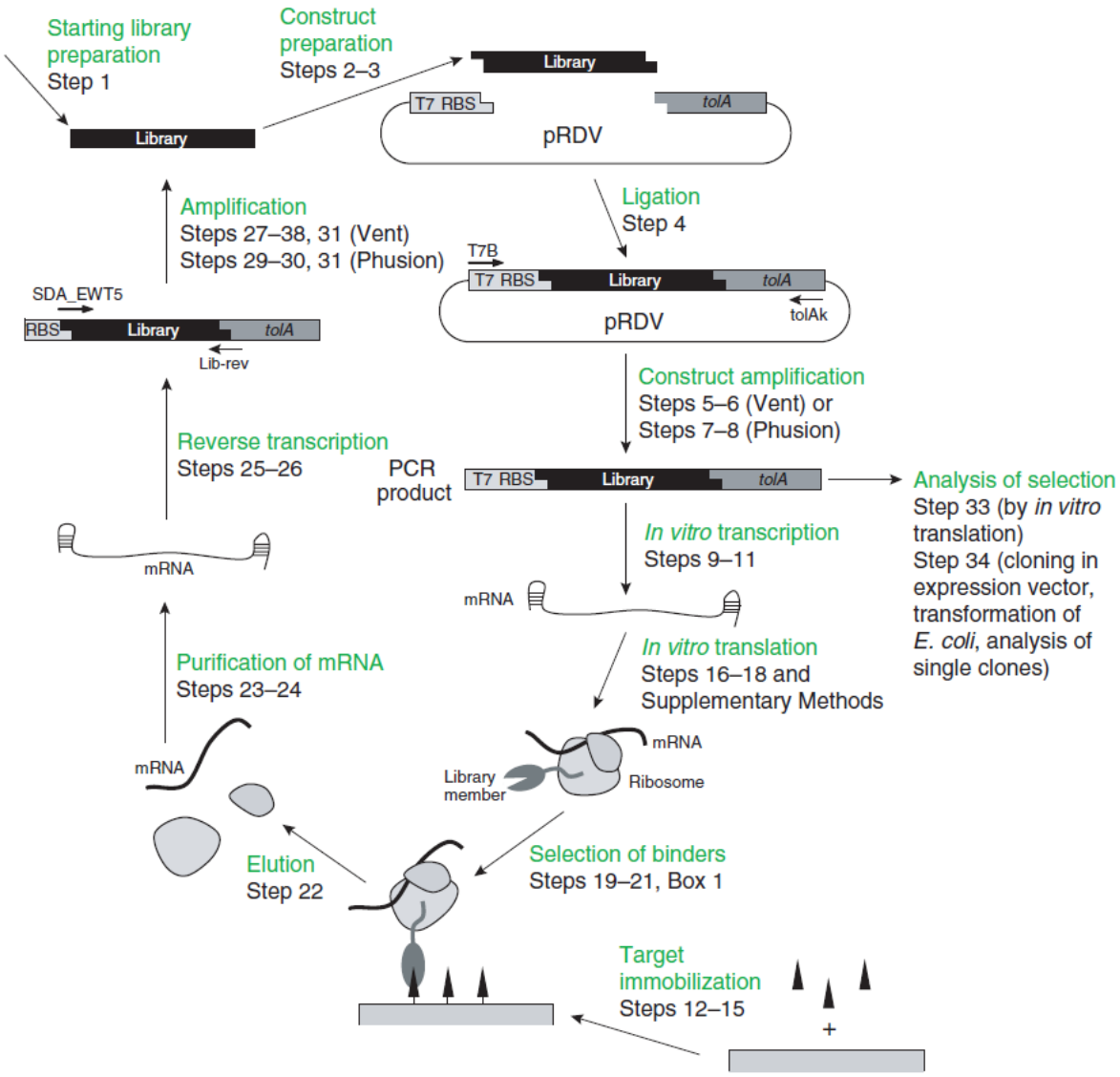


Figure 1.2 The ribosome display selection cycle. A DNA library in the form of a PCR product, coding for binding proteins, is ligated into a ribosome display vector (pRDV), thereby genetically fusing it to a spacer sequence in-frame, and providing a promoter T7 and translation initiation region (ribosome binding site, RBS) at the 5' end. The final ribosome display construct is obtained by PCR amplification of both flanking regions and the library insert from the ligated vector. *In vitro* transcription of this PCR product yields mRNA that is used for *in vitro* translation. The ribosome stalls at the end of the mRNA and does not release the encoded and properly folded protein because of the absence of a stop codon. The ternary mRNA-ribosome-protien complexes are used for affinity selection on an immobilised target. The mRNA of bound complexes is recovered, after washing from dissociated ribosomes, reverse transcribed and amplified by PCR. Thereby the selected pools of binders can be used directly for the next cycle of ribosome display or analysis of single clones after cloning into expression vectors, which are then used for E. coli transformation and small-scale in vivo expression. Adapted from (Zahnd, Amstutz et al. 2007)

1.1.4 mRNA display

An alternative mRNA display method utilises puromycin to link RNA and protein covalently, creating stable protein-mRNA complex (Roberts and Szostak 1997). It is similar to ribosome display in that both methods try to link the mRNA with the expressed protein, however mRNA display has a distinct difference that it uses puromycin as a linker and the ribosome is not retained with the complex. A brief comparison of both methods is shown in **Figure 1.3**. Puromycin acts as a stable homologue of amino-acyl tRNA. During *in vitro* translation, puromycin enters the ribosome A site and binds covalently to newly translated protein, inhibiting further translation as a result of the peptidyl transferase activity of the ribosome. Reports had shown that puromycin-mediated ribosome display can achieve up to 40% of the mRNA successfully translated (Takahashi, Austin et al. 2003). To improve the stability of the complex, the mRNA on the complexes can be reverse transcribed to cDNA to improve the stability before the rigorous washing steps. mRNA display had been reported to generate large peptide libraries of more than 10^{13} unique variants for *in vitro* selection (Takahashi, Austin et al. 2003). The application of this method has been very diverse, including the identification of RNA-binding ligands (Barrick and Roberts 2003), small molecules (Keefe and Szostak 2001; McPherson, Yang et al. 2002) and peptides (Baggio, Burgstaller et al. 2002). Short, linear random peptides with nanomolar affinity to target proteins of interest were also able to isolated by mRNA display, however 7 panning rounds were required to select peptide with such affinity (Wilson, Keefe et al. 2001).

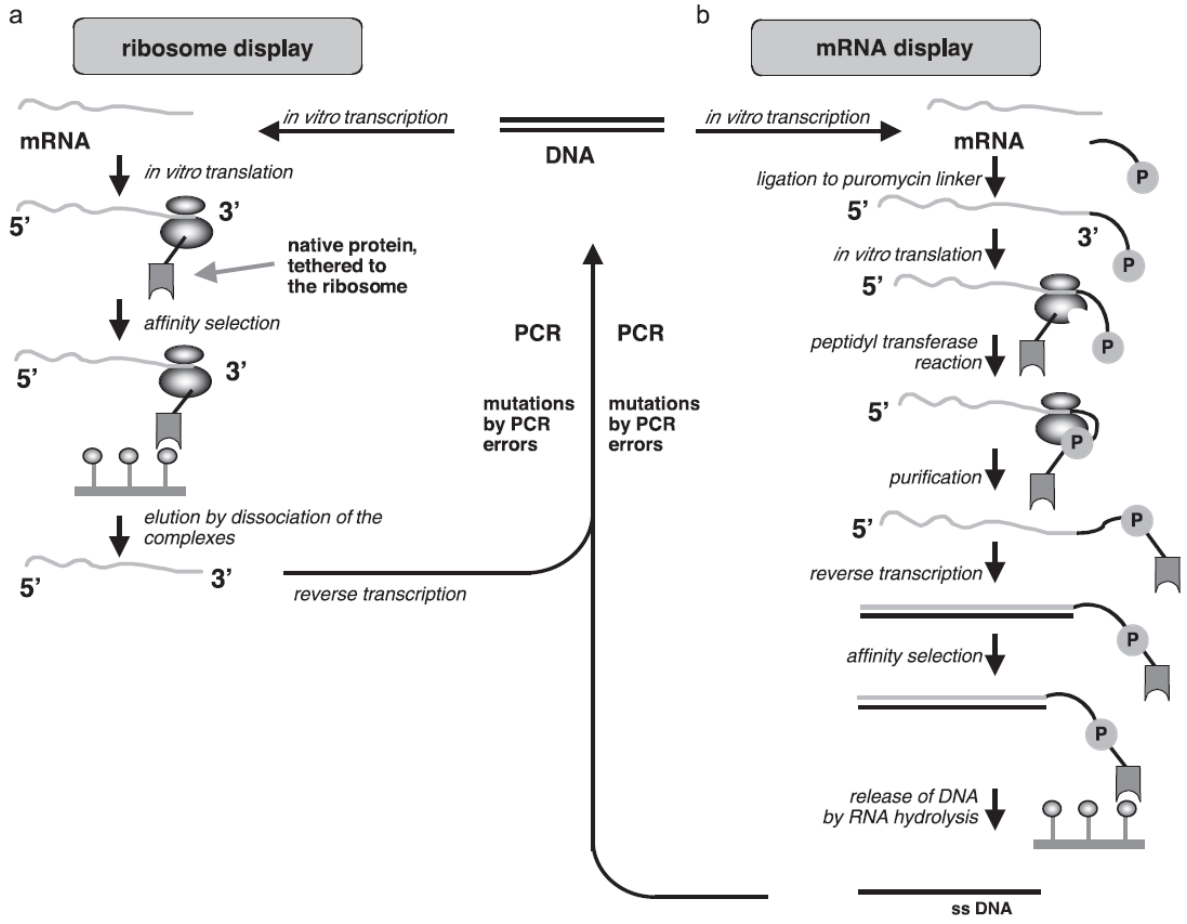


Figure 1.3 Comparisons of ribosome and mRNA display methods. (a) ribosome display (left). DNA encoding the library is transcribed *in vitro*. The resulting mRNA lacks a stop codon, giving rise to linked mRNA-ribosome-protein complexes during *in vitro* translation. These can be stabilised and directly used for selection against an immobilised target. The mRNA incorporated in bound complexes is eluted and purified. Reverse transcription PCR can introduce mutations and yields a DNA pool enriched for binders that can be used for the next cycle. (b) mRNA display (right). Covalent mRNA-protein complexes are generated by ligation of a DNA linker with a small adapter molecule, puromycin, to the *in vitro* transcribed mRNA, which lacks a stop codon. The mRNA is translated *in vitro*, and the ribosome stalls at the RNA-DNA junction. Puromycin then binds to the ribosomal A-site, and the nascent polypeptide is thereby transferred to puromycin, which acts as a homologue of aminoacyl-tRNA. The resulting covalently linked mRNA-protein complex is isolated, reverse transcribed and used for selection experiments. The DNA strand is recovered from target-bound complexes by hydrolysing the complementary mRNA at high pH, and is amplified by PCR. Adapted from (Lipovsek and Pluckthun 2004)

1.1.5 CIS and CAD display

Although ribosome and mRNA display increase the speed of the selection by bypassing the need for cellular display systems, these methods have the disadvantage of rapid RNA degradation and the need for reverse transcription PCR after selection. To overcome this, DNA-based display systems had been developed which allowed rapid library generation by direct PCR and subsequent analysis by DNA-based next-generation sequencing. CIS display utilises the bacterial replication initiator protein RepA to achieve a DNA display system. An illustration of a typical CIS selection cycle is shown in **Figure 1.4**. RepA has a unique ability to only binding to its encoding DNA at the 3' end of two non-coding regions (called CIS and ori) to the RepA coding sequence, called the cis-activity. CIS domain is involved in binding and loading of the protein whereas ori contains a RepA binding domain. The detailed binding mechanism still remains to be elucidated but it is hypothesised that the transcription by RNA polymerase is stalled at the CIS domain, enabling the RepA to be translated and attached to its binding site on the ori domain on the same template. Reports have shown that CIS display was able to recover the sequence of a specific binding ligand from a large library of variants containing 10^{10} members (Odegrip, Coomber et al. 2004). CIS display has been applied to select various targets, including small binding peptides, structural proteins, and antibody fragments (Odegrip, Coomber et al. 2004; Eldridge, Cooley et al. 2009). In peptide selection, the method has been modified to select for highly stable peptides by treating the entire libraries under harsh conditions, i.e. extreme pH, high protease environments, and selecting for peptides that were not degraded under these conditions (Eldridge, Cooley et al. 2009). CIS display was also utilised for probing cells; peptide-DNA-RepA complexes were coupled with fluorescent labels and incubated with Jurkat and CHO cells, the location of these complexes could be visualised using high resolution microscopes.

A similar DNA-based display system, the covalent antibody display (CAD), utilises a bacteriophage endonuclease P2A protein, which had a cis-DNA binding activity to form covalent bonds with the coding sequence through catalytic tyrosine activity. An example of CAD was the selection for high affinity binders to tetanus toxin with up to 300-fold enrichment rates (Reiersen, Lobersli et al. 2005).

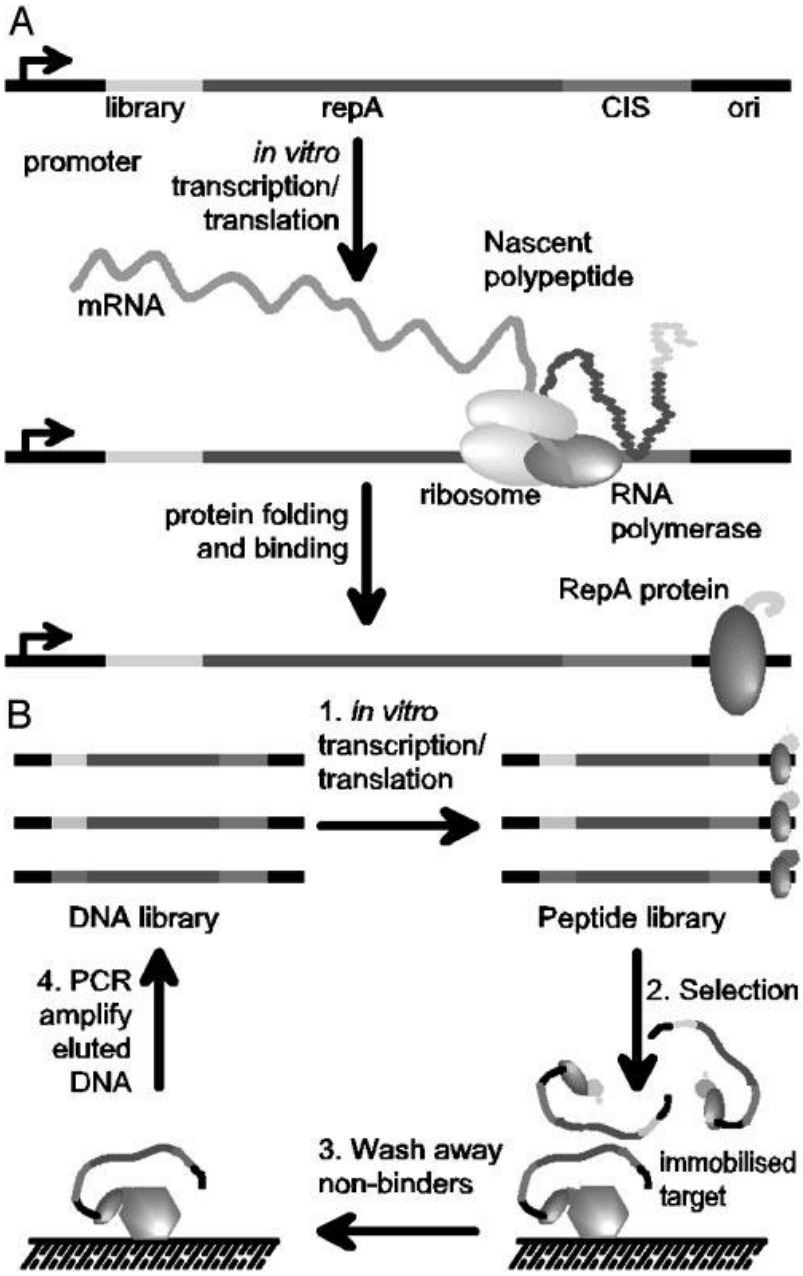


Figure 1.4 CIS display selection cycle. (a) Template DNA encoding an N-terminal library peptide is ligated to the RepA gene. *In vitro* transcription is initiated at the promoter and pauses when the RNA polymerase reaches the CIS element. Concurrent translation produces the RepA protein, which transiently interacts with the CIS element, thereby forcing its subsequent binding to the adjacent ori sequence. This process establishes a faithful linkage between a template DNA and the expressed polypeptide that it encodes. (b) CIS display selections begin with the construction of a peptide-encoding DNA library followed by *in vitro* transcription/translation to form a pool of protein-DNA complexes (step 1). The library pool is incubated with an immobilised target (step 2), nonbinding peptides are washed away (step 3), and the retained DNA that encodes the target-binding peptides is eluted and amplified by PCR (step 4), to form a DNA library ready for the next round of selection. After three to five rounds of selection, recovered DNA is cloned into an appropriate expression vector for the identification of individual target-binding peptide sequences. Adapted from (Odegrip, Coomber et al. 2004)

1.1.6 Directed evolution of proteins in emulsion – *in vitro* compartmentalisation

It is possible to create artificial cell compartments which mimic the cell cytoplasm for protein library generation. Ideally, the mechanism of probing protein interactions or probing potential ligands will directly link any identified protein or peptide to its encoding DNA, for ease of sequencing and ultimately identifying the protein sequence. To create a simpler reaction environment but at the same time preserve the feature of isolated and unique environments that only a cellular reaction can provide, the idea of *in vitro* compartmentalisation has been proposed. Griffith et al., managed to create unique and isolated directed evolution environments using water-in-oil emulsions (Griffiths and Tawfik 2006), comprising a mixture of mineral oil, surfactant (Span 80, Tween 80) and trace amount of detergent (Triton X-100) forming the "oil phase", and conventional aqueous reaction components, i.e. coupled *in vitro* transcription and translation reaction to create the "water" phase. It was reported that in a 1 ml reaction volume, more than 10^{10} water-in-oil emulsions can be created with high-speed stirring using a magnetic stir bar, where each emulsion has a mean diameter in the range of 2-3 μm and mean volume of 5 fL. At this volume, a single molecule achieves a concentration of approximately 0.5 nM, thus enabling a single DNA molecule to be transcribed and translated (Tawfik and Griffiths 1998). With proper dilution of DNA molecules, it is possible to create individual water-in-oil emulsions in which only one DNA molecule is present in an emulsion, and the protein expressed is trapped in a single confined physical space, i.e. creating 10^{10} unique directed evolution reactions. To fish out the DNA and protein samples after breaking the emulsions, the use of magnetic beads was proposed (Dressman, Yan et al. 2003; Diehl, Li et al. 2006), the method was named by the group as BEAMing (Bead, Emulsion, Amplification, Magnetic), and had been rapidly developed as one of the Next Generation Sequencing approaches by Roche 454 Sequencing, (Rothberg and Leamon 2008). The magnetic beads are coated with streptavidin (Dynabeads, Invitrogen),

and attached with biotinylated primers. During emulsion formation, the templates and primer-coated beads are critically diluted so that only one template and one bead can be found in an emulsion. After the PCR reaction, this results in a single magnetic particle with thousands of copies of DNA identical in sequence to the original ones. The population of beads (approximately 10^{10}) then becomes the representation of the starting library. The bead bound double stranded DNA templates can be detected via binding of fluorescent antibodies. In addition, there are numerous ways in which this bead bound library can be linked to its phenotype. Hideo's group sought to explore the use of emulsions to investigate protein-protein, protein-DNA interactions. The Japanese group named their approach as GLOBE (genetic library on beads), and managed to construct a genetic library of *Paracoccus denitrificans* and selected gene beads that were bound to the His-tagged transcription factor PhaR by FACS (Kojima, Takei et al. 2005).

Furthermore, they constructed a gene template which has T7 promoter systems flanking a glutathione-S-transferase (GST) linked with a Histidine tag, which the middle four histidine residues were randomised. Through the emulsion PCR, each bead will contain copies of a template with specific histidine mutations. The beads were harvested by breaking the emulsion, followed by linking with biotinylated anti-GST antibodies. And the bead-DNA-antibody mixtures formed another emulsion with the presence of coupled *in vitro* transcription and translation components. The expressed protein contains GST which will immediately be captured by the anti-GST antibody on the beads. After breaking the emulsion for the second time, the bead-DNA-antibody-protein mixtures were stained with anti-His6-FITC antibody, and the brightest beads were collected by FACS sorting and the DNA was analysed. It was discovered that after two rounds of screening, the randomised sequences were substantially converging to peptide-encoding sequences recognised by the anti-His6 antibody, at an enrichment rate of up to 917 folds per round (Gan, Yamanaka et al. 2008).

This study showed a practical alternative to phage display, which is simpler, *in vitro* and easy to handle. In addition, over the subsequent years, Hideo's group also published papers utilising a similar approach to discover the sequence for high affinity binding Fab fragments (Sumida, Doi et al. 2009), and identified the sequence of a high affinity angiotensin II-inhibiting peptide that showed significant higher affinity to angiotensin II, which regulates blood pressure (Gan, Furuzawa et al. 2010). It was also reported that instead of the magnetic bead approach, critical dilution of the gene library into one molecule per well in water-in-oil emulsion can also achieve comparable screening results (Rungpragayphan, Yamane et al. 2007) (Koga, Yamane et al. 2007). However flow cytometric based approaches may associate with loss of sensitivity and specificity secondary due to the requirement to express threshold levels of fluorescence before selection. This requires sufficiently large numbers of fluorescence binding partners to be detectable during the flow cytometry. Furthermore, if binding events are rare, then these may be lost during flow-guided cell sorting for the subsequent rounds of bead generation and selection of target-specific peptides. Lastly the existing bead methodology is limited to the identification of protein ligands of protein targets as *in vitro* transcription/translation requires a template larger than a peptide as discussed in detail below.

1.2 Comparison of current methods

A brief illustration of the distinguishing features of the described methods is shown in **Figure 1.5**). Depending on the application, there may be situations which are more suitable for one method than the other. These methods are compared in terms of speed, library size, expressed protein size, and cost of production.

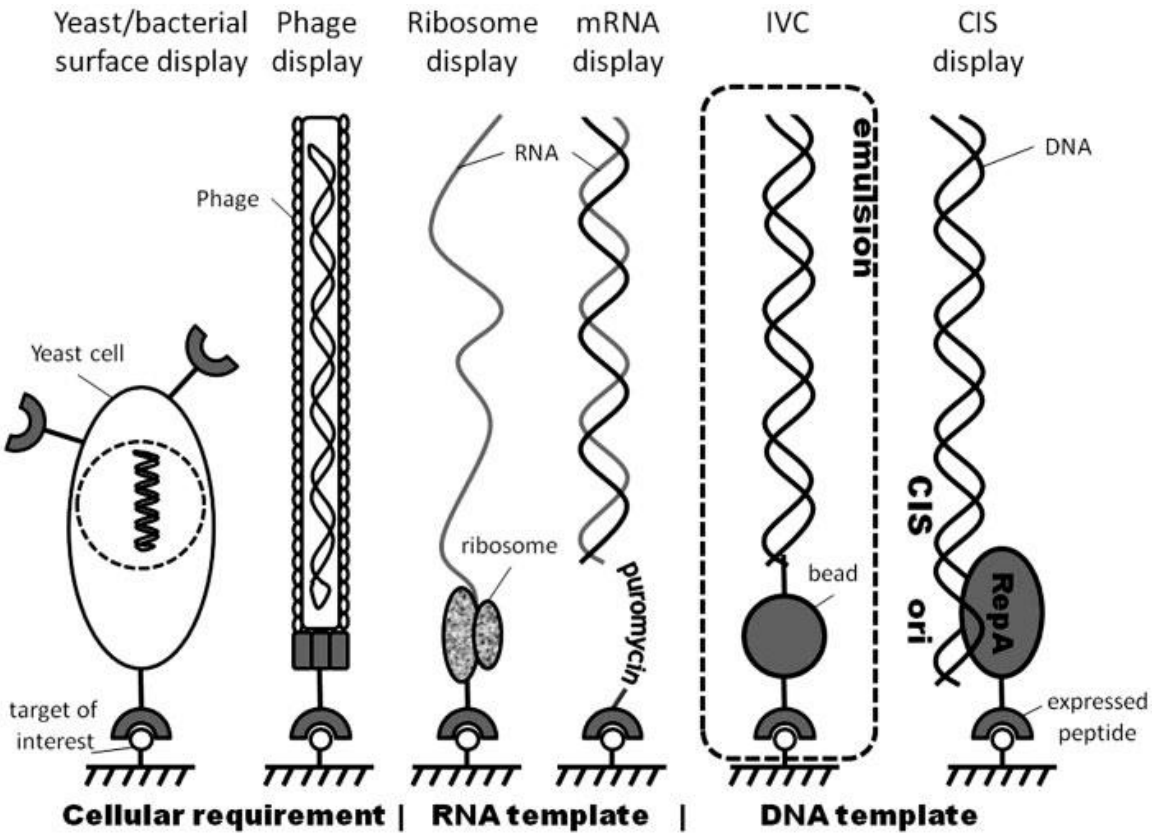


Figure 1.5 A simple illustration of different *in vitro* display methods. Phage and yeast display are categorised as cellular method, whereas the rest of the methods are acellular. Adapted from (Ullman, Frigotto et al. 2011)

The strengths and weaknesses of different methods are summarised in Table 1.1. DNA based methods such as CIS, CAD display and IVC are superior in terms of speed when compared with phage and mRNA based display methods. Phage and bacterial display requires the transformation and growth of the cells, making them the slowest and most laborious methods. mRNA-based methods have a quicker turn-around time because of the acellular experimental

conditions. However, having to work with unstable mRNA-protein complexes and the requirement of reverse transcription to cDNA slows the entire selection process and the possibility of losing sequence information by RNA degradation. DNA-based selection methods such as CIS, CAD and IVC proved to be the fastest selection methods among other methods currently available, as DNA-based methods have the convenience of library amplification by direct PCR, and DNA is much more resistant to hydrolysis and degradation than RNA molecules. In terms of library size, acellular methods are able to produce more variants in a single experiment since the transformation to live cells is not required, the proteins do not have to exit through the secretory pathway, and wider variety of antimicrobial peptides are able to be produced. Although it is possible to produce libraries contain large proteins, phage display has intrinsic limitation to display larger protein, due to secretion and folding problems (Fernandez-Gacio, Uguen et al. 2003; Hoogenboom 2005). Other methods overcome the folding issues by the use of eukaryotic cell extracts such as wheat cells and rabbit reticulocytes for the *in vitro* transcription and translation, and in these settings, the size of the protein is less limited due to cell-free environments. Of the *in vitro* systems, it is only the emulsion approach that potentially offers large multimeric libraries – these may be important for enhancing detection of lower affinity interactions that may still be relevant for in vivo biology. Also multimerisation opens the possibility of detecting interactions of the library with cells rather than with immobilised protein targets. A further key advantage of the *in vitro* systems is that it is potentially possible to use unnatural amino acids and post-translational modifications (e.g. stapling/cyclic peptides) which are more difficult using in vivo approaches. It is worth noting that all the methods described above are relatively cheap to produce a library with large degree of variants, but the cost of analysing the sequence information can be exponentially more expensive if techniques such as next-generation sequencing are chosen.

method	speed	library size	protein size
phage/yeast	slow	10^{10}	small
ribosome	medium	10^{12}	small, large
mRNA	medium	10^{12}	small, large
CIS/CAD	fast	10^{13}	small, large
IVC	fast	10^{14}	small, large

Table 1.1 Comparison of current *in vitro* display methods.

1.3 Intrinsic problems encountered in display methods

In vitro selection methods involving living cells, i.e. phage display, have several intrinsic disadvantages. Firstly, they are both time and labour intensive, as working with living cells that have to be propagated between each round of selection. Secondly, it is difficult to control the natural selection processes within living cells and it is harder for the expression of proteins that are less favourable by the expressing cells, thus limiting the diversity of the protein library. Thirdly, it is difficult to express proteins that operate at unfavourable conditions such as extreme pH or use unnatural amino acids. Fourthly, it has been found that overexpression of more than 51% of endogeneous E. coli ORFs can cause severe growth defects, while 77% of these ORFs cause growth inhibition (Kitagawa, Ara et al. 2005). Finally, the *in vivo* methods described have the disadvantage that in addition to the proteins under study there are other proteins that are on both the phage and/or the yeast surface which can interfere with binding and increase non-specific interactions. Also, fusion proteins can alter the conformation of both interacting partners and influence binding. The recombinant proteins may be toxic to the yeast or the phage or influence their replication. Post-translational modifications may also be important for binding between the natural ligands and these may not be present in the displayed proteins using yeast or phage based systems.

Although cell-free systems such as ribosome, mRNA, CIS display systems had been proposed to overcome the problems encountered in phage display, the above mentioned systems shared a common disadvantage. The environment in which the molecular evolution (DNA to protein) occurred is not isolated, which could potentially hinder the diversity of the variant protein library due to natural selection. For example, in a standardised environment, proteins which are more favourable with the pH or hydrophobicity of the standardised environment are more likely to be translated whereas proteins which are less favoured by the environment might never be able to translate. This problem would be more severe in a large library, e.g. variants $>10^{10}$.

1.4 Therapeutic peptides re-emerging as a pharmaceutical interest

The vulnerability of protease digestion and high cost in production are the main reasons that make therapeutic peptides unattractive as drug candidates. However, in recent years different routes of delivering therapeutic compounds to patients have been identified (Vlieghe, Lisowski et al. 2010. As a result, the pharmaceutical industries have shown a revival of interest in peptides as potential drug candidates. New peptide synthesis methods have been developed to increase the production yield and new chemical modifications to reduce protease activity have been designed to augment the peptide efficacy. There are a substantial number of peptide drugs being marketed now {Vlieghe, 2010 #138).

There are three common sources where peptides have been experimented to have therapeutic uses: firstly, natural occurring peptides produced by animals and plants (from fragments of bulky proteins or hormones); Secondly, high affinity ligands selected from recombinant libraries; and thirdly, synthetic compounds selected from chemical libraries (Sato, Viswanathan et al. 2006). Different synthesis methods are suitable for peptides with different lengths and sizes. Current common peptide synthesis methods include chemical synthesis,

recombinant library, *in vitro* transcription and translation systems, transgenic animals and plants or enzymatic synthesis. During the production of peptides, designing and incorporating unnatural amino acids allows the formation of chemical bonds and alters the structural properties (Neumann, Wang et al. 2010), thus increases the stability of the synthesised peptides under physiological conditions. In recent years, the difficulty in the production of synthetic therapeutic peptides has been greatly reduced with developments of solid-phase peptide synthesis (SPPS) (Chandrudu, Simerska et al. 2013). The primary SPPS workflows are sequential synthesis, convergent synthesis and chemical ligation, and the yield of the synthesis can range from 1.5 mg to 5 kg depending on different systems.

1.5 Limitations of peptides as therapeutic drug candidates

Peptides greatly vary in chemical properties; factors such as peptide delivery, absorption, cellular trafficking must be considered if peptides are to be used in therapeutic setting. For example, peptide properties such as the aqueous solubility, lipophobicity, formation of hydrogen bonds, and physiological stability under proteolytic or enzymatic actions have direct impact on the manufacture and distribution of therapeutic peptides to patients. To summarise, the main limitations of using peptides as drug candidates include: poor delivery route, intravenous delivery is often necessary; short efficacy due to the rapid proteolytic degradation by proteases in the gut and blood plasma; rapid removal from the blood circulation by the hepatic and renal clearance; hydrophilic nature which results in poor cellular membrane trafficking; lack of stable conformational structure, which results in a lack of selectivity and poor specificity, which might lead to targeting of unwanted targets; possible risk of auto-immunogenic reactions; high synthesis and manufacturing costs (the production cost of a 5000 Da molecular mass peptide exceeds the production cost of a 500 Da molecular mass small molecule by more than 10-fold (Bray 2003).

1.6 Peptides are still a pharmaceutical interest

The pharmaceutical industry suffers a huge reduction in the average number of licenced drugs per year and the increasing in the costs of research and development. The field is in need of alternative research directions to increase drug productivity. As a result, peptide as therapeutic drug options is potentially a viable direction for addressing unresolved clinical problems, particularly where small molecules have failed or show poor specificity.

Peptides or small proteins are not preferred as new drug candidates, since most of the current ones require intravenous injection. Nonetheless erythropoietin and insulin are very competitive in the market despite the fact that they cannot be orally administered (Loffet 2002). It is clear that many biological systems in humans are based on peptides as ligands (e.g. anti-microbial peptides) and so it is likely that this class of molecules has bioactive potential. Research had found that prolonged treatment with alternate chemical composition or different ways of administration could potentially give new directions to the therapeutic peptides (Antosova, Mackova et al. 2009). Currently, most of the licenced therapeutic peptide drugs are hormone homologues or derivatives that mimic the function of physiological hormones. Receptor agonists or antagonists which have significant role in cancer physiology, immune system, antibiotics, and metabolic products are notably being studied for the potential of peptides as therapeutic options. This suggests that peptides might evolve into a unique niche market. In addition, other therapeutic peptides, for example peptides with wide coverage of antimicrobial activity against bacteria, fungi and viruses, have also become an exciting future opportunity, particularly in counteracting the loss of efficiency of traditional antibiotics (Jenssen, Hamill et al. 2006; Rossi, Rangasamy et al. 2008; Rotem and Mor 2009)

1.7 Synthetic peptides as vaccines and drugs against infections agents and cancer

Peptides and their homologous compounds (protein and antibody fragments) are applied in a wide variety of clinical pathologies, including immunity diseases, infectious diseases, cardiovascular, diabetes, cancers and other diseases (Loffet 2002; Stevenson 2009).

Instead of the conventional vaccine strategies (attenuated viruses/pathogen, epitopes selected from recombinant DNA library), an alternative option is to utilise synthetic peptides that shares the sequence homology to the disease-causing proteins or the surface protein of a pathogen. Synthetic peptides offer many advantages such as flexibility, production speed, compound purity and safety over attenuated viruses or pathogens, since they do not require the use of bacterial plasmid to express.

1.7.1 Synthetic peptides as cancer, allergy and viral vaccines

Synthetic peptides are also being studied in the research and development of vaccines to trigger T cell-based responses for cancer prevention. Vaccines against the human papillomavirus (HPV) are great examples of prophylactic vaccines e.g. cervical cancer. Animal models had proven to be a success, by prolonging the CD8+ T cell response (Widmann, Romero et al. 1992). In addition, synthetic peptides which mimicked the HPV E6 and E7 oncogenic proteins and p53 protein were used to study in vulvar intraepithelial neoplasia and ovarian cancer patients with promising results (Melief and van der Burg 2008; Kenter, Welters et al. 2009; Leffers, Lambeck et al. 2009).

Synthetic peptides have also been developed as allergy vaccines. Immunotherapy which utilises allergens is conducted by injecting the patient with increment amounts of the allergen to induce host immune tolerance. For example, the synthetic peptides were developed and tested *in vitro* to determine any residual binding to IgE obtained from patients who were allergic to bee venom proteins and then vaccinated in patients. The study showed that a

reduction in the antigen-specific T cell response and a reduced IgE serum concentration (von Garnier, Astori et al. 2000).

In another study, a combination design of α-helical coiled-coil domain was utilised to induce an immune response against α-helical viral proteins, or to produce a specific drug such as Fuzeon T-20, which blocks HIV from fusing to the target cell (Wild, Shugars et al. 1994; Miceli, Myszka et al. 1996; Lu and Hodges 2002).

1.8 The aim - designing a new approach to the *in vitro* display

The potential of peptides as therapeutic drugs is promising, but the technologies which screen for these peptides remained inefficient and unproductive. The aim of this project is to focus on the advantages provided by the water-in-oil directed evolution method to screen for peptide ligands of cell surface receptors. We also attempted to create constrained peptide aptamers. Peptide aptamers are short stretch of constrained peptides that are tailored specifically to bind and interact with molecular proteins, the aptamers usually consist of a variable peptide loop attached at both ends to a protein scaffold. This double structural constraint would significantly enhance the binding affinity of the aptamer to micro/nanomolar binding level (Binkowski, Miller et al. 2005). and the screening of ligand refolded MHC class I binding to T-cell receptors. Human Major Histocompatibility Complex (MHC) is a stringent mechanism that antigen presenting cells use to present correct epitopes for T-cell receptors. The size of the correct epitopes (which are produced intracellularly) can range from 6 to 10 amino acids in length for the presentation of MHC class I presentation to $CD8^+$ T-cells, or 10-20 amino acids (produced by processing extracellular proteins) for MHC class II presentation to $CD4^+$ T-cells (Guermonprez, Valladeau et al. 2002).

In order to be recognised by T-cell receptors, peptide epitopes must be correctly presented on an MHC molecule or an MHC-like molecule. However, the affinity of the T cell receptor

(TCR) for peptide loaded MHC is usually so low (dissociation constant ~50 μM; 10,000-fold weaker than a typical antibody–antigen interaction) that it was clear from early measurement, that a labelled monomeric peptide-MHC reagent would not survive even a single washing step (Matsui, Boniface et al. 1991).

We sought to develop an *in vitro* display approach in order to identify peptide ligands of both immobilised and cell based targets. Given the limitations of existing systems, we had to develop a novel strategy that facilitated the identification of peptide ligands presented in a multimeric form (chapter 3). We use HIV-1 gp120 as a proof of concept target (chapter 4). We wished to further extend the technology to use it as an approach to the identification of the specificity of T cells (chapter 5). In the discussion (chapter 6), I discuss the relevance of the approach to the wider field and its potential use in identification of novel drugs or drug leads.

To begin with, the DNA template must be created which contains necessary coding sequence for *in vitro* transcription and translation. There are a large number of potential systems ranging from E. coli, wheat, yeast, and mammalian cells. For the reasons of cost, stability, and rapid coupling of transcription/translation, RTS 100 High Yield E coli Coupled *in vitro* Transcription/translation kit was chosen. RTS 100 utilises the T7 promoter system. It has previously been reported that a synthetic protein trimer consisting of epitope peptide – spacer – β-2-microglobulin – spacer – heavy chain can refold spontaneous into stable MHC class I molecule, which can stimulate antigen-specific T-cells and B-cells (White, Crawford et al. 1999; Yu, Netuschil et al. 2002). In addition, even partial components of the trimer such as epitope linked β-2-microglobulin can form a potent target for cytotoxic T-cells when refolded with HLA heavy chain (Uger and Barber 1998). The project aims to create a peptide library with randomised 9 amino acids linked with β-2-microglobulin and a selection tag

streptavidin-binding-protein SBP. Each of the DNA molecules can be converted to a unique β-2m-peptide conjugate, thus the threshold number of peptide combinations can have up to 20^9, or 10^{18}. By using the method of emulsion PCR, each DNA acts as a template and can be critically diluted to one template per emulsion formation. After PCR cycles, each bead within an emulsion will have thousands of copies of the template it was incubated with, creating homogenous DNA copies on a single bead. The bead-DNA is washed thoroughly and used to form another emulsion with RTS100 coupled with *in vitro* transcription/translation components. During the incubation, each emulsion will translate each unique peptide that encoded on each bead and immediately linked with the streptavidin coated beads, creating a protein-DNA-bead molecule.

It is possible to use a number of approaches to test the practicality of the system. Tests such as Western blotting, Enzyme-linked Immuno-sorbent assay (ELISA), FACS staining and sorting experiments are undertaken to ensure the quality of the expressed protein library, and the refolding of MHC and the feasibility of the system in cellular interaction.

By identifying what epitope a T-cell receptor recognises will allow the identification of the disease relevant part of a known protein. Ultimately this may lead to the identification of potential new therapeutic, diagnostic or prevention targets. It may be of interest to determine the specificity of T cell receptors within diseased tissues. Probing a cell population with a peptide/protein library is not possible with any of the other genotype:phenotype technologies, and so this would represent a significant advantage of bead display over all other approaches.

We are hoping develop a novel *in vitro* bead display system to identify peptides that bind to a target of interest, whether that is a plate-bound molecule or a cell membrane bound target.

Chapter 2 - Materials and Methods

2.1 Synthesis of DNA template containing randomised peptides

The DNA templates used for emulsion PCR and the following emulsion *in vitro* transcription/translation are synthesised by Eurofins MWG Germany, under the Complex Gene Synthesis service.

2.2 Protein Expression - HLA-A2/B8 heavy chains, β2m light chain

Day 0 transformation

The agar plates are prepared by mixing 500 ml ddH2O, 5g tryptone peptone (Sigma-Aldrich), 2.5g yeast extract (Sigma-Aldrich), 5g NaCl (Sigma-Aldrich), 7.5g bacto-agar (Sigma-Aldrich), and autoclave to fully dissolve the mixture. While hot, add 0.5ml of ampicillin/kanamycin stock antibiotic (100 mg/ml) to obtain a final concentration of 100 ug/ml. Pour approximately 20 ml of the agar mixture to each Petri dish, allow to air dry. Thaw frozen competent *E. coli* (one-shot competent BL21, Invitrogen) and plasmids (HLA-A2/B8, β2m, Louise Jones, Ogg lab) slowly on ice, add 2 μl of plasmids (10 ng/μl) to each vial of competent BL21 cells, leave on ice for 30 min. Heat shock the competent cells by incubating at 42^{o}C for 90 sec, remove immediately and place back on ice. Add 250 μl pre-warmed SOC medium (invitrogen) to the vial and incubate at 37^{o}C for 1 hour at 225 rpm shaking speed. Plate by spreading 100 μl of the competent cells evenly on the pre-made agar plates, invert the plates and incubate overnight at 37^{o}C

Day 1 setting up starter culture

Inoculate a colony from the plates the day before into 250 ml of low salt LB Broth (for 1 litre, 10 g tryptone peptone, 5 g yeast extract, 5 g NaCl, 1 litre ddH2O, autoclave) supplemented with 100 μg/ml ampicillin/kanamycin depending on the antibiotics used on the plates. Grow up the bacterial culture overnight at 37^{o}C with 225 rpm shaking.

Day 2 expanding culture and harvesting proteins

For each starting culture, make up 3 x 1 litres of low salt LB broth supplemented with 100 ug/ml ampicillin/kanamycin and 0.4% sterile D-glucose (Sigma-Aldrich). Inoculate 40 ml overnight culture into 1 litre supplemented low salt LB Broth. Grow with shaking at 37^{o}C until culture reaches an OD600 of 0.4-0.6 nm (approximately 3 hrs). Add 1ml 0.5 M IPTG (Sigma-Aldrich, sterile filtered) to 1 litre of bacterial culture. Grow culture with shaking for a further 5 hrs (OD 600 > 2.0). Centrifuge bacteria culture in a sorvall centrifuge (each bottle contains 1 litre culture) at 4000 rpm, 4^{o}C, 30 min. Resuspend bacteria pellet in 50 ml ice cold PBS. Sonicate the sample for approximately 10 min with 1 min pulse and 1 min rest, until the solution is cream smooth. Spin at 13000 rpm for 10 min, followed by 3 washes in triton wash (0.5% Triton X-100, 50 mM Tris pH 8, 100 mM NaCl, 0.1% NaN_3) with 13000 rpm, 10 min spin, 4^{o}C. Wash once with resuspension buffer (50 mM Tris pH 8, 100 mM NaCl), then dissolve in urea buffer (10 M urea, 100 mM NaH_2PO_4 , 10 mM Tris, 100 uM EDTA, 100 uM DTT) and rotor mix overnight. Spin at 13000 rpm for 10 min, keep the supernatant and aliquot in 1ml, and check protein size with polyacrylamide gel and Nanodrop for correct size and concentration. Store at -80^{o}C freezer.

2.3 Emulsion PCR linking emulsion *in vitro* transcription/translation

Creating bead-primers

Dynabead MyOne/M270 streptavidin (Invitrogen)/MACS microbeads (Mitenyl MACS) are washed three times with binding and washing buffer (10mM Tris-HCl, 1mM EDTA, 2M NaCl) by separating beads and supernatant with magnet, re-suspended in binding and washing buffer. Appropriate quantity of biotinylated primers (5'-biotin-forward/reverse, dual-labelled biotin-forward/reverse, polyethelene glycol (18 consecutive PEGs) spacer-forward/reverse, Integrated DNA Technologies Co.) are added to the re-suspended beads and incubate on a rotor at room temperature. The amount of primers added is determined by the

binding capacity suggested by the manufacturer (the maximum binding capacity of 1 mg, or 6-7 x $10^7$of Dynabead to single-stranded oligonucleotides is 200 pmol, the exact concentration of MACS microbeads is unknown).

After binding, the bound beads are washed three times with binding and washing buffer to remove any unbound primer oligos. The bounded beads are re-suspended in water and stored at 4^oC

The sequences of the modified primers are as follows:

Forward primer: 5'– GATCTCGATCCCGCGAAATT -3'

Reverse primer: unmodified, 5'- TCCGGATATAGTTCCTCCTT -3'

Emulsion PCR (emPCR)

The emulsion oil for each PCR reaction is prepared in a 10 ml universal tube as follows: Modified surfactant (Degussa Soft) and mineral oil (Sigma-Aldrich). To equilibrate the emulsion oil, a magnetic stirrer (VWR UK) is added to the tube, which is placed on a magnetic spin at ≥1000 rpm in the cold room for 10 min.

Each aqueous PCR reaction (150 µl) is prepared as follows, the detailed concentrations are omitted due to patent-sensitive issue: primer-coupled beads, complement unmodified primers, unmodified primer, PCR grade dNTPs, 50 mM $MgCl_2$, random DNA template, 10X PCR buffer, and Platinum Taq DNA polymerase. The water-in-oil emulsion is prepared by slow addition of the aqueous PCR mixture into the spinning emulsion oil, and left it to spin for an additional 10 min. The emulsions are then aliquoted and subjected to the following PCR cycles: 94^oC 2 min, 50 cycles of 94^oC 30s, 50^oC 30s, 72^oC 30s, then 72^oC 7 min, followed by 4^oC incubation.

After PCR cycles, the same reactions are pooled together in a new 1.5 ml eppendorf tube. The emulsion is broken by adding 1 ml breaking buffer (Tris-HCl, Triton X100, NaCl, EDTA)

in the emulsions, vortex for 5s, and centrifuge at 13,000 x g for 5 min at room temperature. Remove the upper solvent phase (appears as white aggregates). Repeat the washing for two extra rounds.

To remove any leftover ether, vacuum centrifuge the tubes for 5 min at room temperature. (Alternatively, left the tubes opened in the fume hood for 10 min). Re-suspend the beads in 100 µl ddH_2O.

Emulsion *in vitro* transcription/translation (emIVTT)

Prepare the emulsion oils for a single reaction as follows: 944.5 µl mineral oil, 50 µl Span-80, 5 µl Tween-80, 0.5 µl Triton X-100.

Prepare each *in vitro* transcription/translation reaction (RTS 100 E. coli HY kit, Roche) as follows: E. coli lysate, reaction mix, amino acids (no methionine), methionine, reconstitution buffer, H_2O up to 40 µl. Add the re-suspended beads to *in vitro* transcription/translation mix to form a 50 µl reaction, mix thoroughly by pipetting. Add the 50 µl *in vitro* transcription/translation mixture to the spinning emulsions as described in emulsion PCR. Incubate the reaction at 30^{o}C for 4 hours on a heat block. Break the emulsions as described. Re-suspend the beads in 50 µl PBS. Examine the integrity of the translated proteins by Western Blot.

2.4 Topo-cloning – extracting sequence information of the epitopes that bind to T-cells

Wash A2-specific T-cell clones with RPMI 1640 medium, re-suspend 10^4 cells in 400 µl RPMI 1640 medium for each reaction. Incubate 50 µl of the protein-DNA-bead resuspension with 10^4 T-cell clones for 30 min at 37^{o}C. Transfer the reaction to a 24-well Millipore Millicell filter culture plate, spin the plate at 1500 rpm, 90 sec. Unbound beads will be

washed through the filter, re-suspend the remaining of the cells with ddH_2O. Transfer the reaction on ice. Depending on the rate of reaction, if a large number of beads are internalized by the cells due to T-cell receptor-MHC complex interaction, cell lysis treatment will be required to harvest the beads. If the beads are attached on the cell surface, the DNA can be amplified directly. Run a standard PCR amplification (94^{o}C 30 sec, 50^{o}C 30 sec, 72^{o}C 30 sec, 35 cycles) with primers flanking the randomized protein region. Purify the PCR samples with PCR purification kit (Qiagen). Run the purified DNA sample with an extension PCR with 5' primer adding additional 4 bases (CACC) on the 5' end of the DNA template. This will be needed for directional TOPO cloning. Purify the PCR samples with PCR purification kit (Qiagen). Perform the Directional TOPO cloning reaction (pENTR Directional TOPO Cloning Kits, Invitrogen) as described by the manufacturer's instructions. Plate the topo cloned reactions on kanamycin supplemented agar plates, incubate at 37^{o}C overnight. Next day, pick 20 to 50 colonies and expand the bacteria by incubating in 5 ml LB culture supplemented with kanamycin, incubate at 37^{o}C with shaking until the OD600 is 0.5 nm or above. Extract the plasmids by running plasmid extraction (Qiagen Miniprep kit) according to the manufacturer's instructions. Sequences the plasmid to understand the exact protein sequences which interact with A2-specific T-cell clones

2.5 Western Blot

20 µl aliquot of protein-DNA samples (from emIVTT) was incubated at 70^{o}C for 10min in 4X LDS sample buffer (NP0008, Invitrogen) and run on a 10% SDS-PAGE Bis-Tris gel (Invitrogen) with NuPAGE MES SDS buffer and novex mini cell sure lock tank (Invitrogen). Molecular weight markers were included on all gels (NEB). Proteins were transferred onto nitrocellulose membranes (Amersham Hybond-CTM Super) using TE-77 ECL semi-dry transfer unit (Amersham) (1h, 50mA/gel). The membranes were blocked at RT for 1h or

overnight at 4^{o}C in blocking buffer (PBS supplemented with 3% BSA and 0.1% Tween-20). Membranes were incubated with the HRP-conjugated antibodies for 1h at RT, i.e. primary antibody mouse monoclonal BBM1 (anti-β2m antibody, HIU in-house) washed with PBS-T (0.2% Tween) 3 times, 10 min each. Then the membrane was stained with secondary antibody goat anti-mouse conjugated with HRP (Dako); the streptavidin-binding protein was detected by streptavidin polymer conjugated HRP (Sigma-Aldrich). The proteins on the membrane were visualised using ECL Western blotting detection reagent (FujiFilm) and exposure to ECL hyperfilm (FujiFilm).

2.6 gp120 synthesis

Gp120 was produced by transient transfection of 293T cells using the JRFL gp120 – pLex construct. HIV-1 JRFL gp120 protein was captured from the supernatant by Co^{2+} affinity chromatography to the C-terminal his_6 tag. The protein was further purified by size exclusion chromatography on a SD200 16/60 column and the protein purity was verified to be >98% pure by SDSPAGE. Conformation and binding function was confirmed by binding with recombinant human CD4 and a panel of HIV-1 Env-specific monoclonal antibodies in ELISA assays.

2.7 Beads selection

Gp120 at a concentration of 100 µg/ml was coated to an immuno 96 micro-well plate (Nunc, UK) at 4^{0}C overnight. After washing in PBST (0.05% Tween 20) 6 times, 50µl PBS containing approximately 10 million beads collected after IVTT was added and incubated at 4^{0}C for 1 hr. After the incubation, the supernatant was collected and the wells were washed with PBST for 6 times. The beads remained on the plate were then collected in TE buffer for subsequent PCR to pull down the DNA fragment containing the random nucleotide region.

2.8 Peptide analysis and synthesis

The sequence of 9 amino acids or 15 amino acids (cys-cys) was translated from DNA sequencing data. The sequence analysis is performed using software R to characterise the frequency of different amino acids in each position and the frequency of 3 or longer amino-acid fragments. The sequences containing a high frequency of certain amino acids or fragments were chosen for synthesis. A GS-linker (GGGGSGG) and a biotin were added at the C terminal of the peptides. The purity was established by high-pressure liquid chromatography, and the individual peptides were dissolved at 20mg/ml in dimethyl sulfoxide and diluted at 1 mg/ml with PBS.

2.9 Peptide selection via amino acid sequence analysis

The 9mer peptides were broken into 3mer fragments and the frequency of these 3mers was analysed with scripts written in R software. The peptides contain higher-frequency 3mers appeared in independent replicates were chosen to synthesis. For example, GP16 (LWCRRLNLL) was the peptide containing the NLL repeated for 6 times in 4 replicates and LWC repeated 2 times in 2 replicates.

2.10 ELISA (antibody blocking)

ELISA plate (Greiner, 655061) was coated with 1 μg / ml purified gp120 in PBS and incubated at 4^{o}C overnight. The plate was washed with ELISA wash buffer (PBS, 0.05% Tween20, VWR) followed by 100 μl blocking buffer (PBS, 0.05% Tween20, 2% BSA powder, Sigma) for 1 hr at room temperature. For antibody blocking experiment, V3/VRC01/F105 antibodies (NIBSC, CFAR3219/3291/3115) were used in 1 μg / ml for 1 hr at room temperature. After blocking, the plates were washed and peptides were added at

different concentrations and incubate for 1 hr at room temperature. After washes, the detection of the binding was done by the addition of streptavidin conjugated peroxidase (Sigma, at 1:2500 dilution in PBS for 1 hr at room temperature and developed by TMB substrate (Thermo, 8008743723). After colour change was observed, stop buffer (0.16 M sulphuric acid, Sigma) was added to stop the reaction, and the absorbance was read at OD_{450} with ELISA plate reader.

2.11 ELISA (peptide blocking)

JRFL gp120 protein was coated on ELISA plate at 1 μg / ml 4^{o}C overnight and blocking with BSA as described above. Peptides were prepared in 4 dilutions and added to the plate and incubated for 1 hr at room temperature. The plates were washed and recombinant CCR5 GST tag (H00001234-P01, ABNOVA) was added to the plate at 0.1 μg / ml and incubated for 1 hr at room temperature. The plates were washed and the detection antibody anti-GST HRP conjugated was added at 1:2000, 1 hr, room temperature. The developing process was identical as described above. Paired t-test was used to compare between different concentrations of peptides (GP1-22) and the negative control peptide (hemagglutinin-binding peptide, 9 amino acids) in the ELISA absorbance reading, 95% confidence level.

2.12 Surface plasmon resonance (BIAcore)

Ligand immobilisation

Immobilisation of purified gp120 protein (1 mg/ml) and purified HLA-A2 heavy chain (reference control, 1 mg/ml) onto the surface of individual flow-cells (Fc1 and Fc2 respectively) of CM5 sensor chip (GE Healthcare) was performed under conditions of 20 μl/min flow rate and 25°C, with HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% v/v Surfactant P20; GE Healthcare) as the running buffer. After initiating the system,

100 µl 0.05M solution of EDC [1-ethyl-3-(3-dimethylaminopropyl) carbodiimide] / NHS (N-Hydroxysuccinimide) at 1:1 ratio was injected to activate the dextran matrix of the sensor chip surface. After immobilisation, 150 µl of 1M ethanolamine hydrochloride was injected to deactivate the unreacted carboxylmethyl groups on the dextran matrix. Finally, 30 µl of Glycine pH2.5 was injected to remove any non-covalently bound proteins.

Measurement of the binding affinity of the interaction between predicted peptides and gp120

The affinity analyses of the interaction between 9-mer predicted peptides and purified gp120 were performed using the BIAcore T200 (GE Healthcare), under conditions of 10 µl/min flow rate and 25°C and using HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% v/v Surfactant P20; GE Healthcare) as the running buffer. Prior to experiment, the 9-mer peptides were diluted to 1 mg/ml with HBS-EP buffer. A 2-fold serial dilution of totally 8 concentrations was used to pass through the immobilised peptides. The data were analysed using Microcal Origin version 7.5 (Microcal Software Limited).

2.13 HIV-1 inhibition Assay

Intracellular p24 antigen expression in $CD4^+$ cells were used to quantify HIV-1 infection. Viral inhibition was assessed by the HIV-1 infection rate after the peptide incubation. Cryopreserved PBMCs were thawed and stimulated with PHA (5 µg/mL) in RPMI 1640 medium (Sigma, UK) supplemented with 10% fetal calf serum (R10) for 3 days. Biotinylated peptides were incubated with streptavidin for 10 min in room temperature. The peptide-streptavidin complexes (10µM) and HIV-1_{BaL} (National Institute for Biological Standards and Control, United Kingdom), HIV-1 IIIB and HIV-1 clade C (ES-X1936-C) were incubated at 37^0C for 30min. Licensed HIV-1 fusion inhibitor T20 (5µM) and HIV-1 without peptides were used as positive and negative controls. After the incubation, PBMCs were infected with HIV-1 isolates (with or without peptides incubation) at a multiplicity of infection (MOI) of

0.002. HIV-1-infected PBMC (2×10^5) were cultured in triplicate in R10 with interleukin 2 (20 IU/mL) in 96-well round-bottomed plates. Cells were harvested at day 3, stained with Aqua Live/Dead Fixable stain (Invitrogen), fixed with 1% paraformaldehyde/20 μg/mL lysolecithin at RT, permeabilized with cold 50% methanol followed by 0.1% Nonidet P-40, finally stained with p24 antibody (KC-57-FITC; Beckman Coulter) and antibodies to CD3, CD4, and CD8 (conjugated to APC-Cy7, PE and APC, respectively; BD Biosciences) and analyzed by flow cytometry. HIV-1 infection rate was expressed as percentage of HIV-1-infected CD4+ T cells (identified by gating on CD3+ CD8- cells, to allow for HIV-induced CD4 downregulation) and determined as follows: [(p24+, $CD4^+$ T cells) + (p24+, CD4- T cells)] / [(p24+ , $CD4^+$ T cells) + (p24+, CD4- T cells) + (P24-, CD4+ T cells)] × 100. The protocols were described previously by Dorrell et al(Yang, Wu et al. 2012).

2.14 HLA-peptide Library

The streptavidin bead displayed DNA library was generated by emulsion PCR using biotinylated primers and templates containing a T7 promoter, 9 random codons (NNK x 9), glycine-serine linker and sequences coding β-2-microglobulin (β2M) and streptavidin binding peptide (SBP).

Emulsion PCR was undertaken as previously described in the work of Diehl et al (Diehl, Li et al. 2006). Streptavidin beads were incubated with primers (forward primer: 5'- biotin – GATCTCGATCCCGCGAAATT; reverse primer: unmodified, 5'-TCCGGATATAGTTCCTCCTT -3'). The emulsion oil for each PCR reaction was prepared with ABIL EM90 and mineral oil. To equilibrate the emulsion oil, a magnetic stirrer was added to the tube. Each aqueous PCR reaction (150 ml) was prepared as follows: 6 ml of primer-coupled beads, 3 ml of complement unmodified primers (400 mM), 3 ml of unmodified primer (2.5 mM), 3 ml of 10mM PCR grade dNTPs, 4.5 ml of 50 mM MgCl2, 10

ml of 20 pM DNA template, 15 ml 10X PCR buffer, and 9 ml of Taq DNA polymerase (5 U/ml). The water-in-oil emulsion is prepared by slow addition of the aqueous PCR mixture into the spinning emulsion oil, and left to spin for an additional 3 min. The emulsions were then aliquoted into 100 ml each, and subjected to PCR. The emulsion was broken by adding 1 ml breaking buffer (10 mM Tris-HCl pH7.5, 1% Triton-X100, 1% SDS, 100 mM NaCl, 1 mM EDTA).

After breaking the emulsion PCR reaction, a new emulsion was formed with components required for *in vitro* transcription translation. Emulsion oils were prepared with each *in vitro* transcription/translation reaction (IVTT, RTS 100 E. coli HY kit, Roche) comprising: 12 ml E. coli lysate, 10 ml reaction mix, 12 ml amino acids (no methionine), 1 ml methionine, 5 ml reconstitution buffers, H2O up to 40 ml. The 10 ml re-suspended beads were added to 40 ml *in vitro* transcription/translation mix to form a 50 ml reaction, and mixed thoroughly by pipetting. The 50 ml *in vitro* transcription/translation mixture was added to the spinning emulsions as described in emulsion PCR. The reaction was incubated at 30^{o}C for 4 hours and the emulsions broken as described above.

Following Western blot identification of β2M and SBP on the surface of the beads, we refolded with HLA heavy chain (HLA-A*0201 or CD8 null A*0201 using domains 1 and 2 of HLA-A*0201 and domain 3 of Kb with mutated residues 226A/227L, which can specifically bind to high affinity HLA-A2 restricted T cells (Choi, Chen et al. 2003). The quality of the expressed protein or the heavy chains on the beads was assessed by Western blot with streptavidin/anti-streptavidin antibody and HC-10 antibody (specific for free heavy chains). The refolded HLA-peptide complex on beads was assessed with antibody W6-32 staining (Shields and Ribaudo 1998) (specific for fully assembled conformationally correct HLA class I molecules).

2.15 PBMC separation and setting up antigen-specific T-cell line

Donors whose HLA types match the desired antigen were bled, e.g. the antigen used is reported to be presented by HLA-B8$^+$ T-cells, then donors who have been identified HLA-B8$^+$ will be bled. Approximately 30-50 ml is required. Quickly dilute the whole blood with RPMI 1640 medium (Sigma-Aldrich) at 1:1 ratio. Add 15 ml lymphoprep in a fresh 50 ml falcon tube, turn the tube at 45^o, and slowly add the whole blood/R_0 mixture along the wall of the tube. Keep the blood on the top of the lymphoprep without shaking. Spin at 2000 rpm, 20 min, without accelerated brake. Use a syringe to remove the buffy coat (leukocyltes and platelets), collect the buffy coat in a fresh tube. Top up the mixture to 50 ml by adding more RPMI 1640 medium. Spin the 50 ml mixture at 1500 rpm, 10 min with breaks. Discard supernatant and resuspend pellet in 10 ml RPMI 1640 medium supplemented with 10% human serum. Count and dilute the cells into 2 x 10^6 cells /ml. Plate the cells in 24-well plate, 2 ml cell mixtures each. Prepare and dilute the antigen to 1 mM, add 8 μl to each well so that the final concentration is 4 μM (optimised concentration). Incubate the mixture at 37^{o}C, 5% CO_2. Change the medium and add IL-2 supplement (4 μl each well, from 10^5 stock IL-2) every 3 to 5 days. At day 10 to 14, check the specificity with ELISpot.

2.16 Enrichment in the sequences containing HLA-A2 anchors

A library containing all possible variants of a nine amino acids would comprise 10^{12} sequences. As we used a library containing 10^7 beads, we designed a perl scripts to identify HLA-A*0201-binding motifs and shorter (5-mer) sequences as such variants would be predicted to be sampled within a 10^7 library. Of note, however, the methodology can be up-scaled to larger numbers of beads. HLA-A*0201-restricted T cell clones were generated to the peptides GILGFVFTL (influenza A matrix), NLVPMVATV (CMV pp65) and

GLCTLVAML (EBV BMLF1) and shown to be specific. The T cells were incubated directly with the HLA-A*0201 beads or after anti-CD8 antibody treatment. The HLA-A*0201 beads were generated using either wild-type HLA-A*0201 or the CD8 null version. Unbound beads were separated from those associated with the T cells using centrifugation through a 3mm filter. The DNA on the remaining beads was isolated and sequencing using high-throughput methodology. The unique sequences accounted for 98% ± 0.2% (mean ± STD) of the total and showed a good diversity of the peptide library. After the selection, there is no obvious decrease in the number of total peptide sequences and the unique sequences account for 95% ± 1.4% of the total peptides (P=0.0002, Student's t test, pre vs post), indicating there is enrichment after the selection. The percentage of unique sequences decreased from 96% ± 0.5% to 94% ± 1.1% (P<0.001) when using anti-CD8+ indicating that CD8 blockade can improve the selection.

The sequences containing HLA-A2 anchors were selected based on a pattern search using: XLXXXXXXV derived from CMV peptide, XLXXXXXXL for EBV peptide and XIXXXXXXL for Flu peptide. The highest frequency and the number of anchor-containing sequences with the frequency ≥2 were normalised to the number of total sequences in each sample. There is an increase in the number of the anchor-containing sequences in post selection samples suggesting that bead selection has occurred. The CD8-null HLA-A*0201 show a significant enhancement of relevant selection of anchor residues. The prior addition of anti-CD8 antibody further enhanced selection of peptide sequences with correct anchor positions. The frequency of peptides showing I2V6L9, L2V6L9, and L2V6V9 were also enriched in the relevant T cell selected populations. These data show that primary anchor residues at positions 2 and 9 can be selected for HLA-A*0201 but in addition the secondary anchor in position 6 is also important.

2.17 Enrichment of the peptide fragments after T cell selection

Perl scripts were developed to identify sequential 5mer peptides within each of the peptides GILGFVFTL, NLVPMVATV and GLCTLVAML. Both forward and reverse sequences were used as the peptides coupled to b2m could associate with the HLA heavy chain in either direction. The frequencies were further normalised to the total sequence number for each sample. None of the 5mers were reduced in both the forward and reverse directions, but several directly matching 5mers were enriched in both forward and reverse directions in the post-selection samples.

2.17.1 Perl script – translation script (the script is assisted by Computational Biology Research Group in WIMM)

The perl script includes the following commands during the search: 1. exclude contaminants, 2. find the ATG start codon, count 15 amino acids downstream, 3. count the number of repeating sequences, show the sequences in frequency order. The contaminants (incomplete sequence fragments that present across all samples, irrespective to the experimental conditions) were removed from the sequence pool and the sequences were ranked in frequency order. Detail scripts can be found in the **Appendix**.

2.17.2 Perl script – identifying HLA-A*0201 anchor residues specific to influenza M peptide antigen, CMV pp65 peptide

In addition to the translation script described in 2.17.1, HLA-A*0201 epitope study requires the query of peptides with restricted anchoring residues in particular position. The additional script would isolate the peptide sequences with anchor residues XIXXXXXXL (influenza) and XLXXXXXXV (CMV p65) and XLXXXXXXL (EBV BMLF-1) and the sequences will

be ranked in frequency order. Read the peptide from N-terminal to C-terminal and from C-terminal to N-terminal to ensure that the sequences are evaluated in both 5' to 3' and 3' to 5'. Detail scripts can be found in the **Appendix**.

2.17.3 Perl script – check the randomness of the library

The variation of the sequences in the library is critical for a quality selection, i.e. all amino acid combinations are present in an unbiased manner. To check the randomness of the library, a perl script was developed to query a random set of peptide fragments in the library. If each sample consists a random peptide library, the search for the fragments should return with similar frequencies for each fragment. Detail scripts can be found in the **Appendix**.

Chapter 3 - Design and optimisation of OrBIT technology

3.1 Introduction

3.1.1 The optimisation of emulsion PCR linking emulsion *in vitro* transcription translation

Emulsion PCR has unstandardized protocols

Since Griffith {Miller, 2006 #141} reported the possibility of conducting PCR in emulsion, there were numerous reports demonstrating different ways to achieve individual experimental needs. However, these methods often differ greatly in various aspects summarised in these main categories: a) the type of beads, b) emulsion composition, c) emulsion formation, d) PCR cycle conditions, e) primer modification, f) emulsion-breaking buffer. There are more than three available options for each of the categories described above, and each protocol seemed to have preference to one type of products or methods than the others. Hence it is important for us to pre-determine our experimental goals and tailored different protocols to serve the purpose. This chapter will describe the fundamental ideas behind this project and describe a step-by-step approach to find suitable conditions for each category.

3.1.2 Peptide library generation: 1. Creating DNA-bead library by emulsion PCR

Figure 3.1 shows a basic schematic of the emulsion PCR with magnetic beads. It is not a requirement to implement magnetic beads to the process. The purpose of making PCR with emulsion is to amplify unique DNA templates coding for unique peptides. With the use of streptavidin coated magnetic beads and biotinylated primers, one could conjugate either forward or reverse primers onto the beads and when the primer-bead conjugates are critically diluted with the emulsion oils and the PCR reagents to 1 bead in 1 emulsion, it is possible to create millions of isolated water-in-oil microenvironments within which the PCR amplified DNA templates are homogeneous, i.e. one double-stranded DNA-bead complex contain one

unique peptide sequence. After breaking the emulsions, DNA-bead complexes can be easily harvested and washed by magnets, the beads are re-suspended in water for further experiment. Each Dynabead contains approximately 2.4 x 10^5 binding sites offered by the coupled streptavidin, therefore when successful PCR is performed and homogeneous DNA templates are attached on the bead, there will be abundant DNA templates for protein synthesis by *in vitro* transcription and translation.

At this stage the quality of beads can be checked by DNA gel electrophoresis, i.e. validate the integrity of DNA templates attaching on the beads.

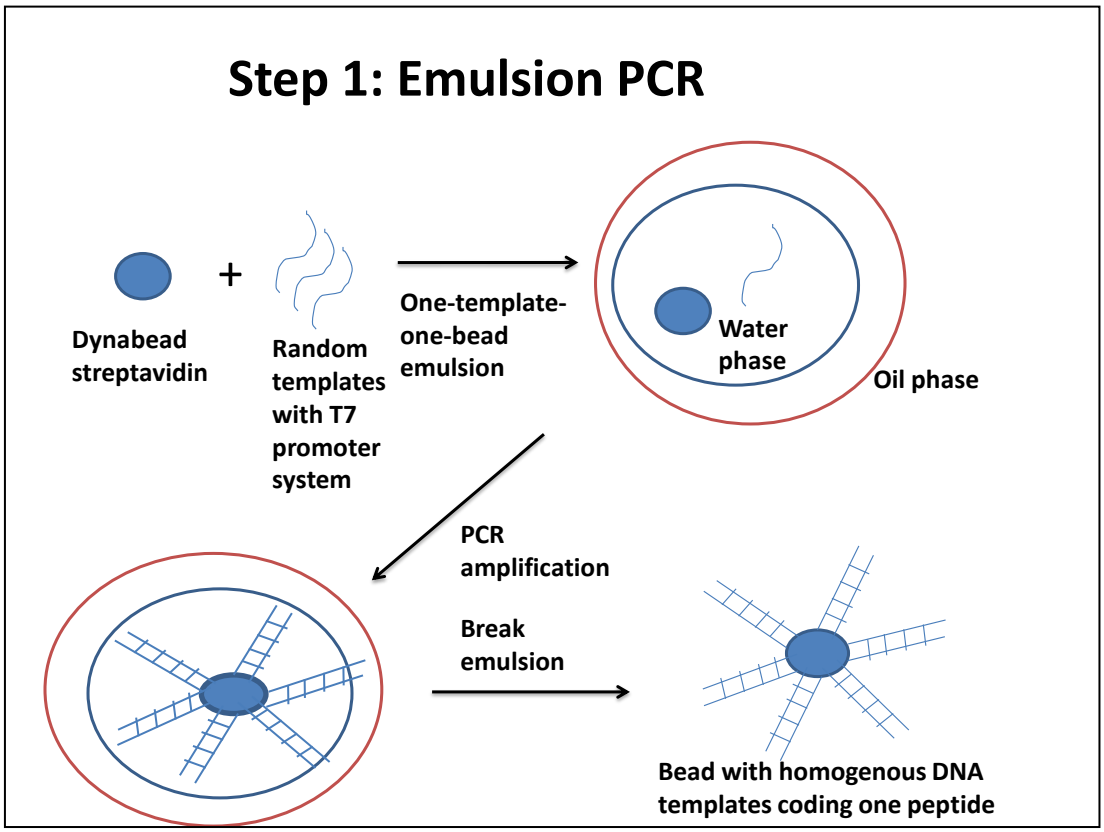


Figure 3.1 Emulsion PCR with beads (emPCR). Streptavidin coated magnetic beads (Dynabead) are pre-coated with biotinylated primers and incubated with one DNA template per bead per emulsion formation. The emulsion mixtures are subjected to PCR cycles and the DNA templates will be amplified and attach on the beads. The emulsions are broken and beads with homogenous copies of double-stranded DNA can be harvested.

3.1.3 Peptide library generation: 2. Synthesis of peptide library by emulsion In vitro transcription and translation

After harvesting DNA-bead complexes from emulsion PCR, the peptide library can be synthesised by incubating the beads with *in vitro* transcription and translation components. In order to maintain the homogeneity of the peptide on the bead, another round of water-in-oil emulsions is formed. Since protein synthesis and PCR utilise different reagents and temperatures, different surfactants (oil phase) are used. **Figure 3.2** shows a schematic view of the emulsion *in vitro* transcription and translation, the DNA-bead complexes are incubated with *in vitro* transcription and translation components (RTS100 system, *E. coli* purified lysates) and water-in-oil emulsions are formed by mixing DNA-bead complex with emulsion oils. The entire mixtures were incubated for 4 hours at 30-37^{o}C depending on the expressed proteins. If the proteins were translated accurately, it should consist of a streptavidin binding protein (SBP) on the C-terminal, which would readily bind to the available streptavidin site on the DNA-bead complex, forming a protein-DNA-bead complex. Since the protein synthesis was conducted in a confined microenvironment with homogeneous DNA template, all the expressed proteins linking on the bead should also be homogenous. Therefore at the end of emPCR linking emIVTT reactions, we would obtain a randomised bead library of unique peptides with traceable coding sequences on the same bead.

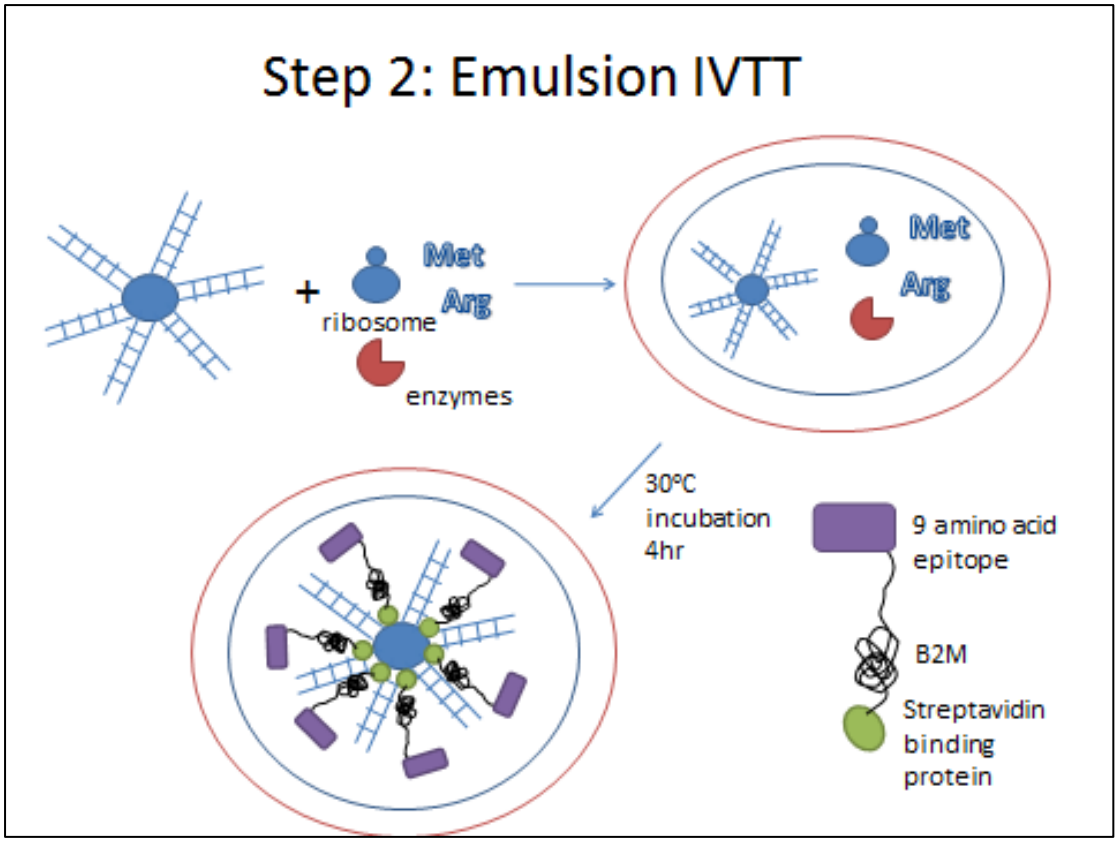


Figure 3.2 emulsion *in vitro* transcription and translation (emIVTT). DNA-bead complexes from emPCR mixed with transcription and translation components, and mixed with emulsion oils to create one bead per emulsion. The water-in-oil emulsion mixtures were incubated for 4 hours at 30^{o}C, allowing proteins to be synthesised from the DNA template (bottom right). Once the protein was translated, it would attach on the available streptavidin binding sites on the bead, creating a homogeneous protein-DNA-bead complex.

3.1.4 Peptide library generation: 3. Use the bead library directly: selection of high affinity binding peptides from the bead library

There are two ways to utilise this bead library for selection, the first and the more conventional method is to use the bead library directly for the selection of high affinity peptides binding to the protein of interest. As shown in **Figure 3.3**, the library of beads was incubated with the immobilised protein of interest on a plate for selection. Beads which contain peptides with high affinity to the target protein would be able to bind and be retained on the surface during washing processes. The binding beads would be eluted and the peptide sequence information can be obtained. The identification and characterisation of high affinity peptides binding to HIV coat protein gp120 will be described in detail in Chapter 4. The second approach is to use beads to probe cells as described below.

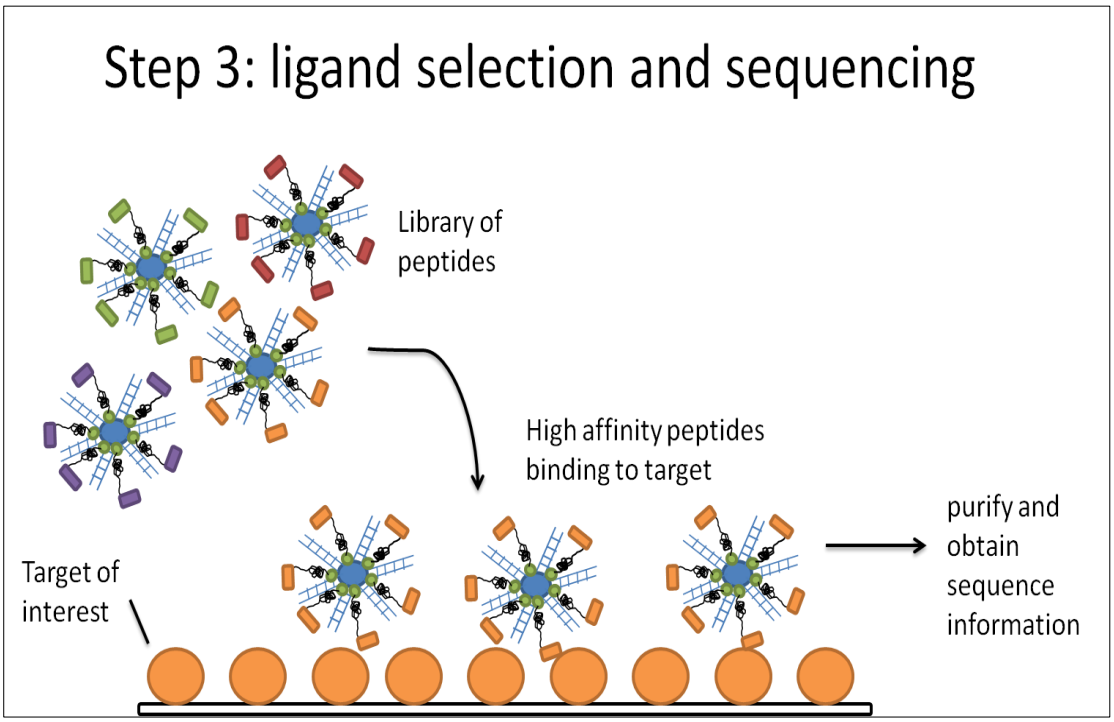


Figure 3.3 target immobilisation and selection of high affinity beads. Protein of interest was immobilised on high binding plate and incubated with bead library briefly. Beads with high affinity peptides would retain on the plate while the remaining ones were washed away. After series of washes, the bounded beads were harvested and sent for sequence analysis.

3.1.5 Peptide library generation: 3. Use the bead library directly: generation of refolded MHC class I molecules on the bead and the selection of antigens recognised by antigen-specific T-cell clones

The template was designed specifically that the randomised peptide library was synthesised with β-2-microglobulin conjugations. When the peptide-DNA-bead complexes were incubated with a recombinant HLA heavy chain under the correct refolding conditions, peptide which was recognised by the heavy chain would be able to refold and form an antigen-presenting MHC class I molecule. As shown in **Figure 3.4**, since the peptides attaching to a single bead are homogenous, it is possible to form a homogenous, multi-meric MHC class I antigen presenting bead. Essentially, a library of epitopes recognised by the specific HLA heavy chain was generated. In our study, we chose to refold the bead library with recombinant HLA-B8, HLA-A0201 heavy chains. By interacting this library with antigen-specific T-cell clones, e.g. T-cell clones derived from flu matrix antigen, we were able to select beads expressing high affinity antigens which could be recognised by the T-cell receptors on the T-cell clones. The sequences of these binding peptides could be amplified and sequenced.

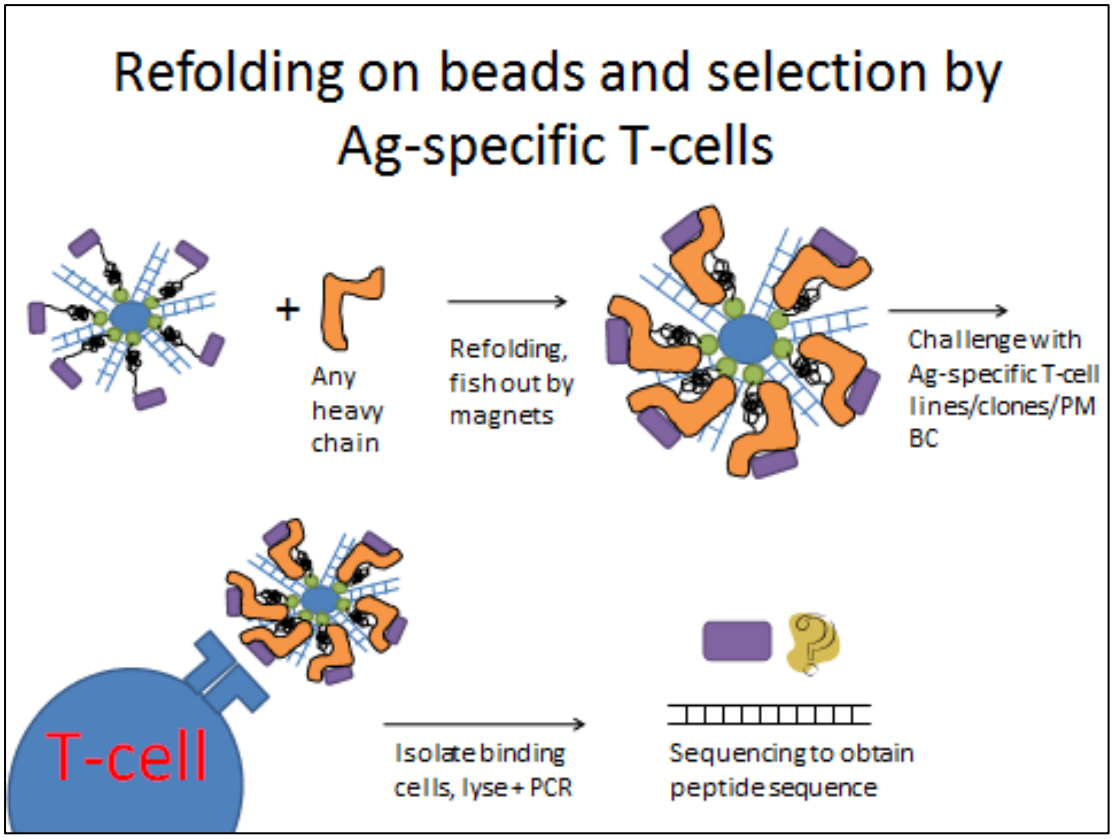


Figure 3.4 formation of MHC class I molecules on the bead and selection by antigen-specific T-cell clones. Peptide-DNA-bead complexes were harvested after emIVTT. The bead complexes were refolded with recombinant HLA heavy chains, forming MHC-DNA-bead complexes. The complexes were then interacted with antigen-specific T-cell clones, and beads which bound to T-cell clones would be isolated and the peptide sequences would be amplified and sequenced to obtain sequence information.

In both selection methods, peptide sequence information was first obtained via PCR amplification of the DNA domain containing randomised peptides. The templates were purified, inserted into plasmid vectors and plated as bacterial colonies. Since each colony contains one unique plasmid, each colony would be representing one peptide sequence. Hundreds of colonies were selected, purified and sequenced to obtain the sequence information of these binding peptides. The working example will be described in Chapter 4.

Alternatively, the more efficient but more expensive way to obtain the sequence information was to use Next Generation Sequencing (Described in Chapter 5).

3.1.6 Construction of randomised peptide library with T7 promoter systems

Figure 3.5 shows the basic design for the randomised peptide template construction. The synthetic protein template is flanked by T7 promoter and T7 terminator, which is the requirement for subsequent *in vitro* transcription and translation experiment. The positive control BZLF1/B8 is a peptide fragment of EBV viral protein BZLF1 (exact sequence RAKFKQLL) which plays a key role as a transcription factor in the switch from latent infection to lytic cycle, and its stringent specificity to HLA-B8 has been utilised to make high affinity MHC tetrameric molecules to study antigen-specific $CD8^+$ T-cells (Uger and Barber 1998) (Callan, Tan et al. 1998). There were four different types of randomised amino acid domains. Firstly, the randomised 9 amino acids encoded 9 consecutive randomised amino acids after T7 promoter, to enhance antigen presentation by MHC class I molecule. Secondly, we also designed randomised amino acids with cysteine residues at intervals which will form disulphide bonds spontaneously, creating a loop-like secondary structure around the randomised amino acids. This design was inspired by the structure of β-defensin, a host defence peptide with 18 to 45 amino acids, including 6 cysteine residues. Defensin is resistance to protease actions and the loop structure allows easier contact and presentation to the targets. We designed templates with single and double cysteine bonds. The sequence of the flexible linker is a long repeat of glycine and serine residues, which were reported to provide flexibility of the epitope which projects outward of the stable β-2-microglobulin molecule (Uger and Barber 1998), thus increases the chance for interaction. The length of the flexible linker linking epitope to β-2-microglobulin is 15 amino acids, whereas the length of which links β-2-microglobulin to streptavidin binding protein is 7 amino acids. The extended length of the flexible linker would allow the refolding of MHC class I molecules with the translated β-2-microglobulin. DNA constructs with only T7 promoter system, epitope and SBP tag had also been synthesised and tested for *in vitro* transcription and translation,

however our result showed that short peptide fragment (smaller than 15 kDa) was extremely difficult to be translated successfully (data not shown). β-2-microglobulin also acts as a marker to test the correct translation by Western blotting. The randomised sequence were constructed by Eurofins MWG with special instruction that the third bases of each randomised codon are restricted to thymine and guanine bases, i.e. this design is to eliminate the maximum number of random stop codons (TAG, TGA, TAA) being generated by random combination. It is also possible to avoid adenine and guanine in the third codon, however this will be accompanied by the loss of majority of amino acid combinations. The absence of adenine will eliminate 2/3 of the stop codons while preserving the variability of amino acid combinations. The decision of not using cytosine in the third base is to enhance the stability of the epitope by limiting the possibility of unwanted G-C base pairs.

The advantage of using linear template is the short and minimum nucleotides necessary for protein to be synthesised *in vitro*. The entire template is less than 1 kilobase pairs, allowing rapid PCR amplification. Fast PCR cycle is a critical part for successful emulsion PCR, as longer templates require longer annealing time and significantly increase the overall PCR process and increase the vulnerability of breaking the water-in-oil emulsion mixtures.

Template design

5' 3'

T7 promoter	BZLF1	Flexible linker	b2m	linker	SBP	linker	T7 terminator

T7 promoter	Random 9 a.a.	Flexible linker	SBP	linker	T7 terminator

T7 promoter	Random 9 a.a.	Flexible linker	b2m	linker	SBP	linker	T7 terminator

T7 promoter	4 Cys + 19 a.a.	Flexible linker	b2m	linker	SBP	linker	T7 terminator

T7 promoter	2 Cys + 11 a.a.	Flexible linker	b2m	linker	SBP	linker	T7 terminator

Figure 3.5 Designs of template. The BZLF1 (EBV viral peptide) template is the template for positive control. All templates except the second one contain β-2-microglobulin (b2m). Randomised amino acids are in 3 variations, the random 9 consecutive amino acids, the 19 random amino acids with 4 fixed cysteines, and the 11 random amino acids with 2 fixed cysteines. Key: T7 promoter/terminator = T7 translation system utilised by *E coli*. a.a. = amino acids, flexible linker = 4 x GGGS amino acid linkers, linker = 1 x GGGS amino acid linker, SBP = streptavidin-binding-peptide, 38 amino acids.

3.2 Results

3.2.1 Emulsion PCR – making water-in-oil emulsions

The first and the most important step in emulsion PCR has to be the formation of water-in-oil emulsion mixture with one bead per emulsion. There are an abundance of streptavidin-coupled magnetic beads available commercially, such as Dynabeads (Invitrogen) and MACS beads (Miltenyi MACS). These beads differ mainly in the diameter of the beads and the amount of binding sites available. Previous research had shown that as long as the water-in-oil formation is optimised, i.e. the ratio between the number of beads and one water-in-oil emulsion should be 1 to 1 or less, one can perform emulsion PCR with any type of beads, with or without magnets. Our method exploits the convenience of magnetic separation, therefore we have experimented with three major streptavidin-coupled magnetic beads: Dynabead MyOne C1 (diameter = 1μm), Dynabead M270 (diameter = 2.7μm) and MACS bead (diameter = nm range, specific diameter unrevealed by manufacturer).

Figure 3.6-8 showed the use of 3 types of beads to form water-in-oil emulsions under two different conditions, the mixing time and the speed of spinning used for mixing. Due to discrepancies in the published literature, it is important to determine our own optimised protocols for water-in-oil emulsion formation. In general, low mixing time and speed resulted in the formation of larger emulsion droplets, as shown in 1 min, 100 rpm microscopic images in Dynabead C1 and Dynabead M270. Lower mixing time and speed were also inversely proportional to the size of the droplets. Due to the mechanical limitation of the mixer (miniature magnetic stir bars), it is not possible to produce a mixture with 100% homogenous emulsion size. However, we found that Dynabead C1 between 150 to 300 rpm mixing speed usually produced the most standardised emulsions, with one single bead in one emulsion (observed as black spot within emulsion). However longer mixing times will substantially reduce the number of water-in-oil emulsions observed under the microscope, as prolonged

mixing will break the emulsions already formed. If mixing speed was increased 5-fold to 1000 rpm, the number of micro-size emulsions would form immediately, however no or little bead will be able to incorporate in the emulsion. Therefore it is important to balance the amount of time on the mixer and the speed of the mixing.

Both Dynabead M270 and Mitenyl MACS bead produced smaller-sized emulsions than the Dynabead C1. There were a significant number of M270 beads observed outside the emulsions (observed as black spots outside the circular droplets), regardless of conditions. The discrepancy between C1 and M270 beads could be due to the fact that M270 beads were 2.7-fold larger than the C1 beads, beads with larger physical size could be harder to incorporate in the emulsion under the same condition achieved by C1 bead. Mixing M270 beads in 1000 rpm for 1 min had proven to be one of the more favourable conditions for M270 emulsion formation.

Due to the limited information from the Miltenyi MACS beads manufacturer, we were not able to obtain the exact size and quantity per ml information of the MACS beads. However we were informed that the diameter of the MACS bead was in the nanometre range. Under the same mixing conditions, emulsions with much smaller size were obtained. In multiple attempts, 200 rpm mixing speed for 5 min produced the most standardised sized emulsions.

By repeated comparison, we found that mixing Dynabead C1 at 200 rpm for 1 min yielded the most number of successful emulsions, which were defined as one bead per emulsion. This mixing condition was used as standardised protocol.

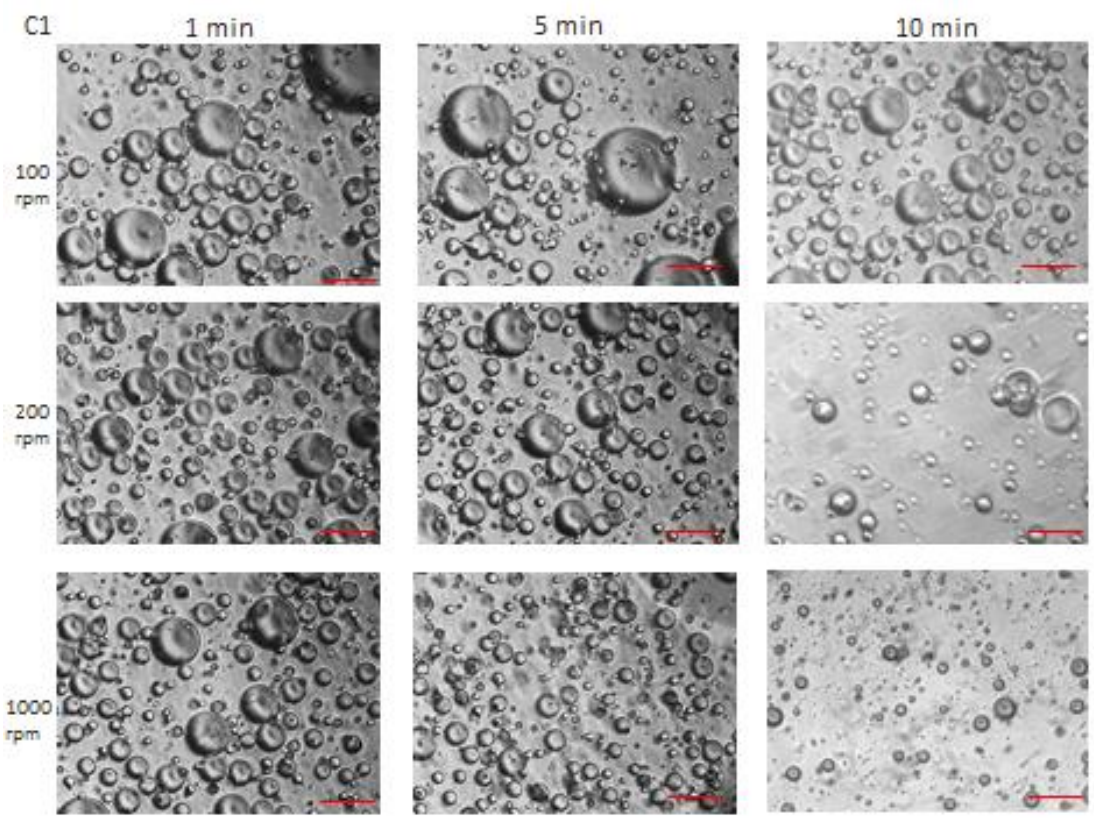


Figure 3.6a the formation of water-in-oil emulsions with Dynabead C1 under different conditions, pictures taken with 10X inverted microscope. Water-in-oil emulsions were presented as circular droplets, and beads could be observed as black dots within the droplets. The red scale bar represents 20 µm.

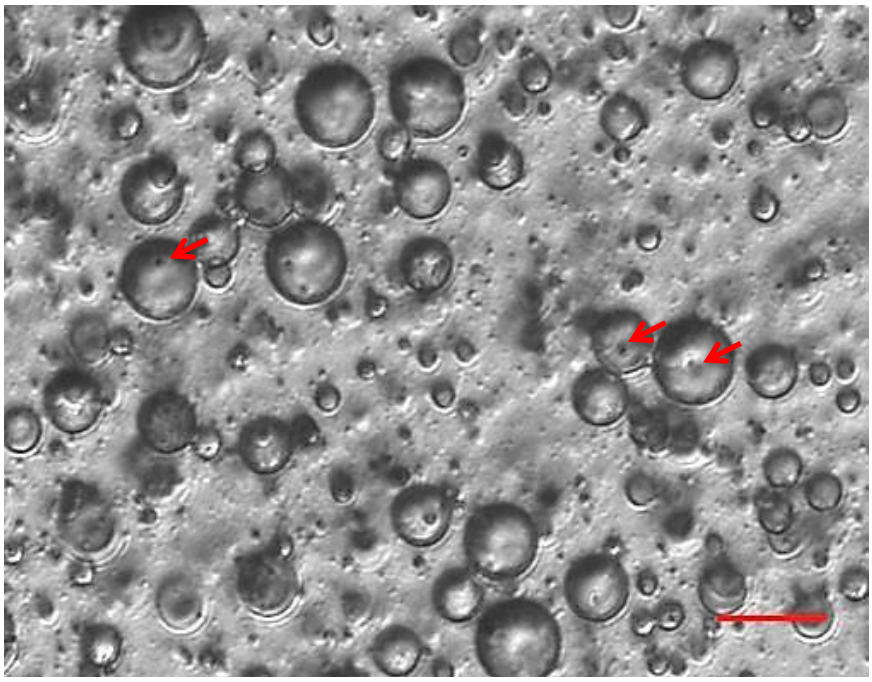


Figure 3.6b The 20X microscopic image of a water-in-oil emulsion mixture with Dynabead C1. Successful emulsion is determined by one bead per emulsion. The scale bar represents 20 µm.

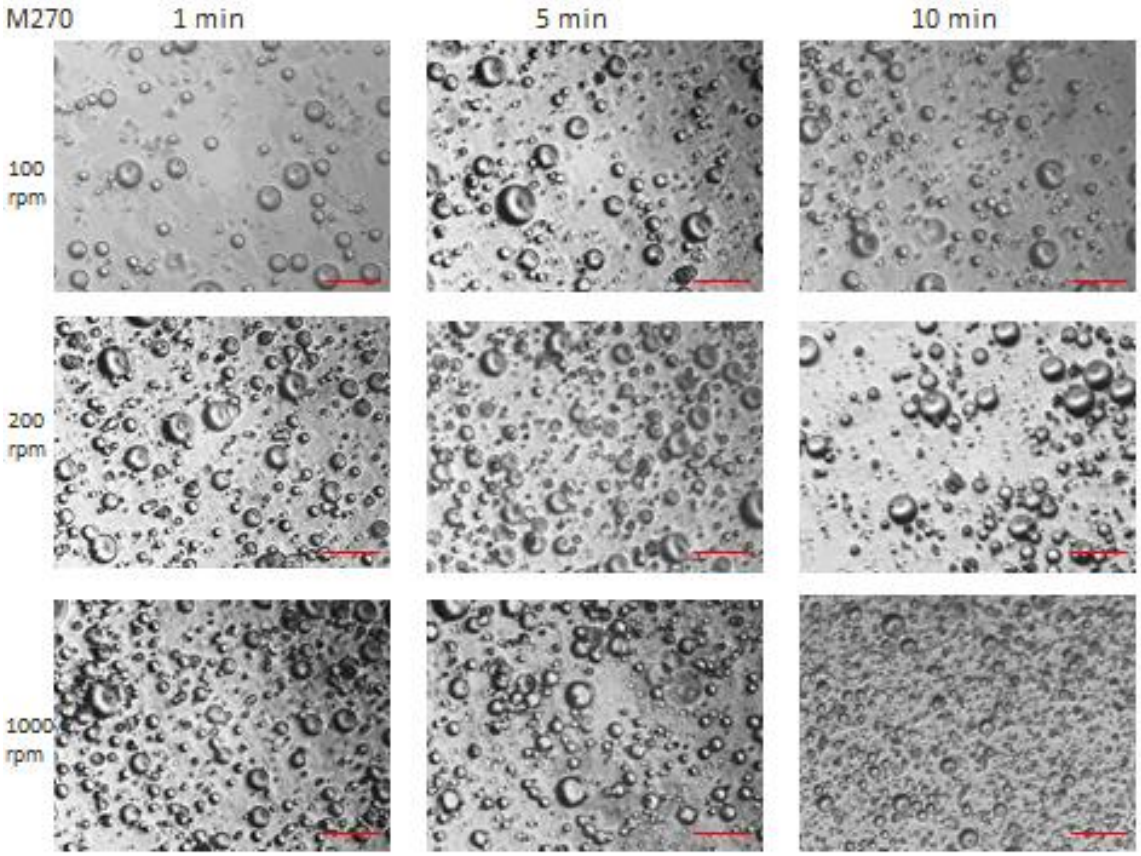


Figure 3.7 the formation of water-in-oil emulsions with Dynabead M270 under different conditions, pictures taken with 10X inverted microscope. Water-in-oil emulsions were presented as circular droplets, and beads could be observed as black dots within the droplets. The red scale bar represents 20 µm.

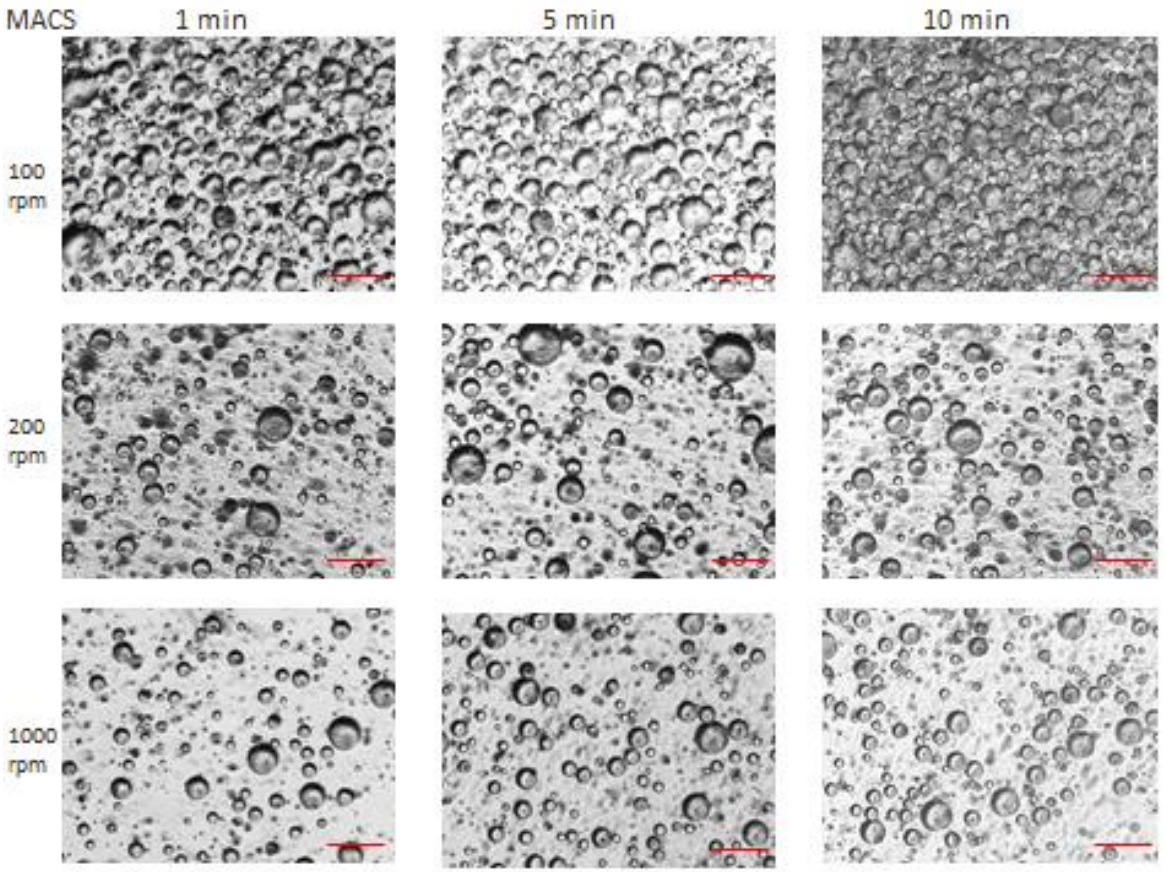


Figure 3.8 the formation of water-in-oil emulsions with Mitenyl MACS bead under different conditions, pictures taken with 10X inverted microscope. Water-in-oil emulsions were presented as circular droplets, and beads could be observed as black dots within the droplets. The red scale bar represents 20 µm.

3.2.2 Emulsion PCR – dual labelled biotinylated primer

There are multiple factors which would affect the success of emulsion PCR. Among those factors, the use of different biotin-modified primers will directly impact on the amplification of templates and the successful annealing on the beads, by influencing the retention of amplified templates on the beads after breaking the emulsions (**Figure 3.9**). Initially the primers were designed with simple single biotin addition on the 5' end of the primer. It was assumed that the strong streptavidin-biotin bond can sustain the harsh conditions in PCR cycles, as a report showed that a short incubation in water above 70^{o}C will reversibly break the interaction without denaturing streptavidin, allowing re-use of the streptavidin solid support (Holmberg, Blomstergren et al. 2005). However, several attempts were made and the results confirmed that most of the amplified DNA was irreversibly lost from the beads after breaking the emulsions, and the remaining templates on the beads were insufficient to carry out emulsion *in vitro* transcription and translation. As shown in **Figure 3.9a**, most of the amplified DNA templates (bands at approximately 800 base pair) were found in the flowthrough of the bead washes, and only very low quantity of templates were found in the re-suspended beads. On the contrary, performing emPCR with dual-biotin, PEG labelled primer-beads amplified templates remaining attached on the beads after several washes.

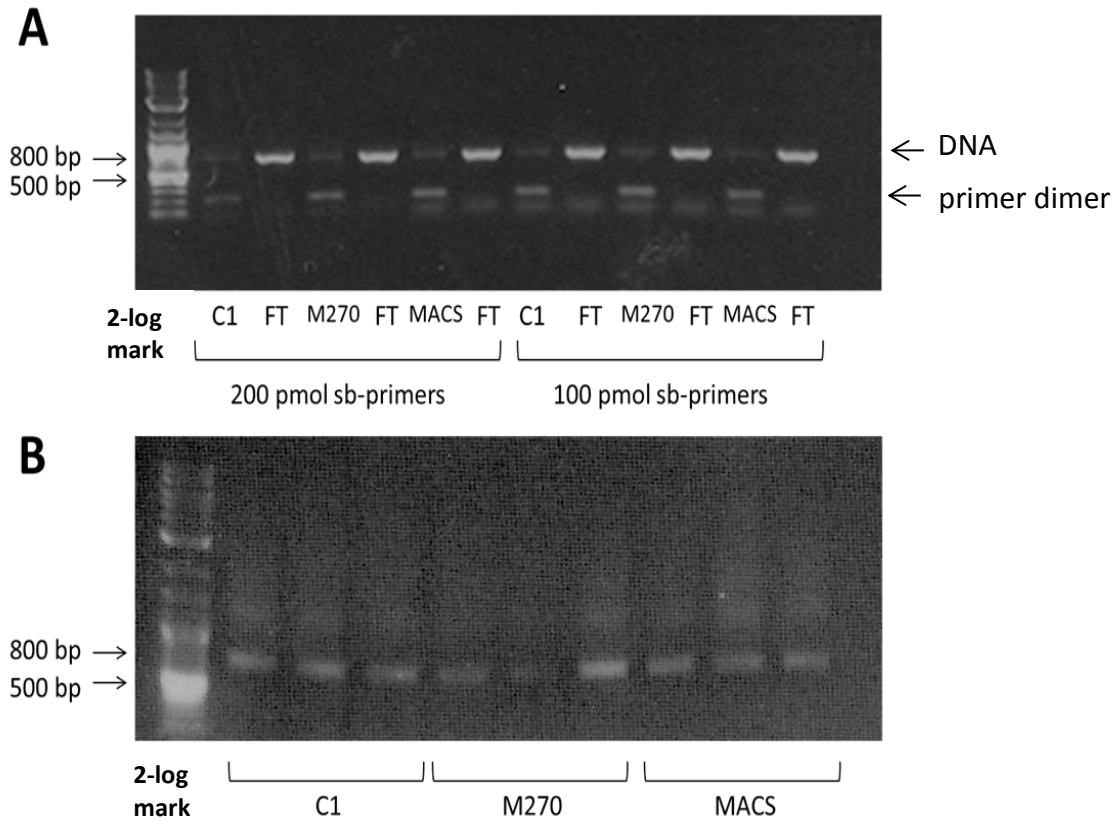


Figure 3.9 single labelled biotin primers v.s. dual labelled biotin primers in emPCR. Fig. 3.9A post-emPCR products using single-biotin labelled primer-beads. Samples were shown in pair, C1 showed the re-suspended beads and FT is the corresponding flowthrough of the beads during bead washes. C1, Dynabead C1 streptavidin. M270, Dynabead M270 streptavidin. MACS, MACS streptavidin.

Fig. 3.9B post-emPCR products using dual-biotin, PEG labelled primer-beads. All samples were re-suspended beads after 3 washes. In C1/M270/MACS column, the wells represented beads linked with 1000 pmol dual-biotin labelled primers, 500 pmol, and 250 pmol. The expected band size of the PCR product would be 780 base pairs.

3.2.3 Emulsion PCR – Dynabead C1, M270, MACS for emPCR

If the emulsions were made and broken under ideal conditions, it was possible to use any of the beads described for emPCR. As shown in **Figure 3.10**, different beads were used for emPCR. The emulsions were broken after PCR cycles and the DNA-beads were harvested, washed and purified. A small amount of the beads were used to examine if the DNA had attached on the beads correctly. It was done by amplifying the DNA domains flanking the random amino acid sites and a portion of β-2-microglobulin encoding sequences. As shown by the DNA agarose gel, all three beads gave the correct size DNA bands. As a result, we had observed that Dynabead C1, M270 and MACS were capable of performing emPCR. In multiple attempts, we observed that Dynabead C1 generally produced more products than the other beads, given the same PCR cycle conditions.

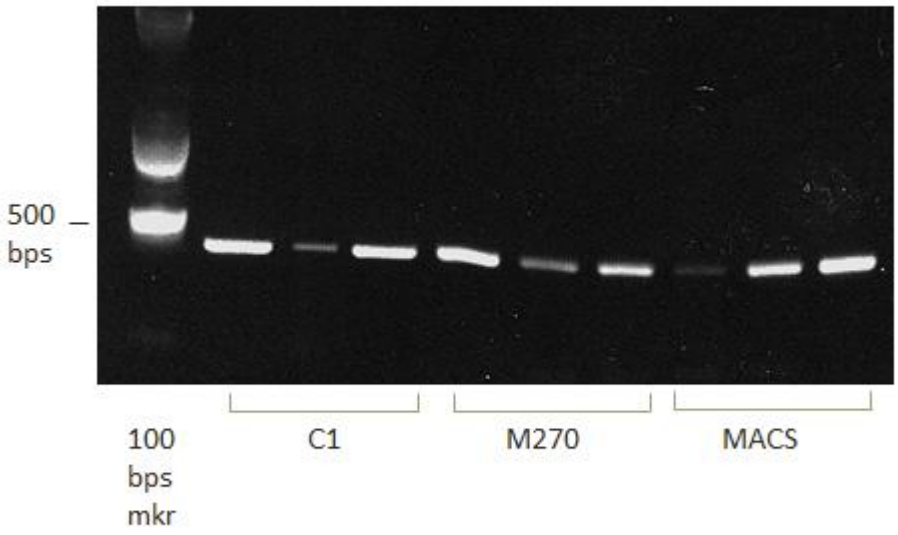


Figure 3.10 post-emulsion PCR gel examinations, performed with Dynabead C1, M270 and MACS beads. Each type of bead was used to form emulsions individually, in triplicates. The emulsions were broken after PCR, and the DNA region containing random amino acids (approximately 450 bps, template = 9 random amino acids) were amplified by standard PCR and checked with DNA gel.

3.2.4 Emulsion PCR – all templates designed were capable of attaching on the beads

We wanted to test whether different templates which encoded different number of randomised amino acids would affect the generation of successful emPCR products. **Figure 3.11** showed the result of emPCR with Dynabead C1 and three different templates described in **Figure 3.5**. As shown in the DNA gel, it was possible to use templates with variable number of randomised amino acids as emPCR templates under the same conditions. The difference in the length of random domain could be clearly observed at the higher position with higher molecular weight, e.g. the DNA band with 4 cysteines and 19 random amino acids was noticeably higher than that of the others. An EBV BZLF1 template which has fixed amino acids and the identical number of base pairs as the 9 random amino acid templates was used as the positive control in this experiment. Note that primers which amplified smaller regions were used to better see the small difference in molecular weights between primers.

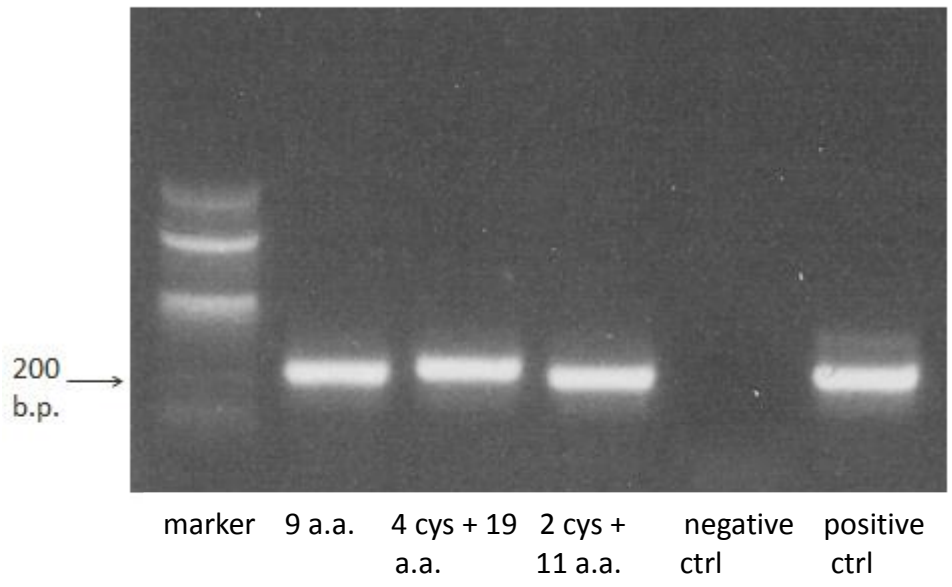


Figure 3.11 emPCR with three types of random templates. The DNA bands represented the PCR products of beads after emPCR. The expected sizes for the three templates were 210, 252 and 218 bps respectively. The positive control was EBV BZLF1 template.

3.2.5 Emulsion PCR – Restriction digestion showed C1 bead had most DNA templates attaching on the beads.

To examine the relative amount of DNA templates attaching on different types of beads, restriction digestion of the emPCR products were performed. **Figure 3.12** showed the restriction digestion of the DNA-beads after emPCR. The DNA templates used had a restriction cutting site for XbaI closer to the 5' end around the T7 promoter region, therefore we could examine the integrity of the template by analysing the restriction digest products. As shown on the DNA gel, for Dynabead C1 and M270, most of the beads were immobilised near the starting wells (as shown by the bands near the wells), some DNA templates on the beads were cleaved and migrated downwards to the correct size (700 base pairs), whereas the other DNA product remained attached on the beads. Dynabead C1 in general had more correct DNA templates attached compared to M270, however both were relatively low. We reasoned that it was due to the position of the restriction site, which was very close to the 5' end where the beads were attached, therefore there might be possible physical steric hindrance preventing the actions by XbaI restriction enzymes. No DNA bands were observed in MCAS beads, this could be due to the fact that only a small quantity of the samples (5 µl from 50 µl) were taken for restriction digestion hence the quantity of MACS beads was not sufficient for restriction digestion. Also note that MACS beads were minute magnetic particles, which could not be retained by the agarose gel, as shown by the lack of band near the starting well. Nonetheless, PCR amplification in **Figure 3.10** suggested that there were evidence of DNA templates on MACS beads.

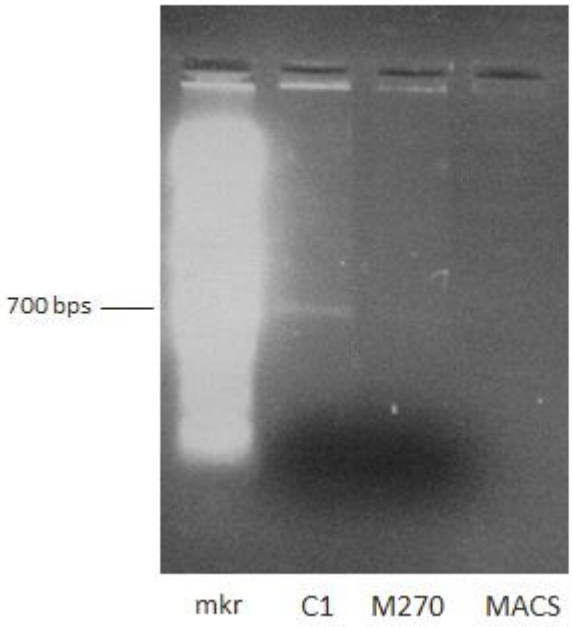


Figure 3.12 XbaI restriction digestion of Dynabead C1, M270 and MACS beads. The restriction cleaving site is approximately 100 bps from 5' terminus, and the final products were (100 + 700 bps). The 100 bps products were on the beads and immobilised near the well due to small pore size. The 2-long DNA marker (mrk) was over-exposed therefore showed as bright smear.

3.2.6 Emulsion PCR – prolonged annealing time in PCR cycle is required for emPCR amplification.

Since the emulsion PCR was amplifying DNA templates under unusual conditions, the time required for each PCR step would also vary from the standard PCR cycles. It was reported that extended annealing time was required for the successful amplification of the template (Diehl, Li et al. 2006). However other emPCR research suggested that standard time was sufficient when the number of PCR cycles was used (Kojima, Takei et al. 2005). We examined both conditions and found that annealing time was critical for the emPCR amplification. **Figure 3.13** showed two emPCR reactions with different PCR annealing time. Dynabead C1, 9 random amino acid templates, 55 PCR cycles were used as standard conditions. As shown on the gel, samples with 1 min annealing time were not able to produce any visible bands, i.e. no templates were amplified and attached on the beads. Although the exact reasons for this difference remained to be elucidated, we reasoned that it was because the primers in the aqueous phase of the emulsion were not in abundance compared to the standard environment, hence longer time were required to pair up the primers on the templates.

We also found that if modified polymerase with fast PCR polymerase enzyme was used in place of standard taq polymerase, it was possible to produce successful emPCR products without the use of 9 min annealing time (only 12 s was needed for annealing, along with 5 s for denaturing and 5 s for extending, data not shown).

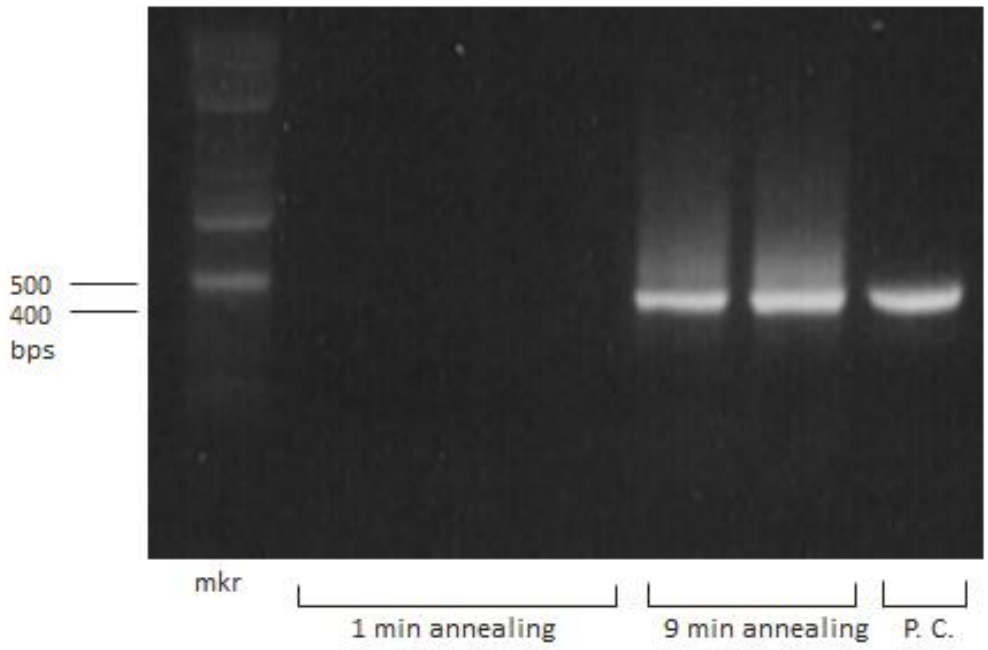


Figure 3.13 emPCR under different PCR annealing conditions. 9 random amino acid templates were used for all samples. After emPCR, the emulsions were broken and the beads were harvested and the target region was amplified and examined (approx. 450 bps).

3.2.7 Emulsion PCR under various conditions

There are several factors which would affect the success of emulsion PCR, a summary of the conditions we had tested are listed in **Table 3.1**. Among those factors, the use of different biotin-modified primers has the direct impact to the amplification of templates, by influencing the retention of amplified templates on the beads after breaking the emulsions. Secondly, insufficient annealing time in PCR cycles also directly affect the success of emPCR, which could be prevented by the use of modified polymerase enzyme. Thirdly, different mixing speed during the emulsion formation created different sized emulsions, which would affect the homogeneity of the templates on the beads.

It is also worth mentioning that we had tried various different reagents for the generation of oil phase in the emulsion, as well as different reagents for breaking the emulsion. We have found that modified surfactant from Degussa was best suited to perform emPCR, and Span 80/Tween80 mixture was better for emIVTT.

We had successfully summarised a standardised pipeline to produce bead library using emPCR, the detailed protocol was described in Material and Methods.

Table 3.1	
Conditions	Effects
Types of beads	
Dynabead C1, M270, MACS	different types of beads does not influence emPCR, as long as streptavidin coated
PCR cycle conditions	
1s, 5s, 10s, 30s denaturation at 94C	altering denaturatation time does not influence templates falling
30s, 1 min, 9 min annealing at 50C	annealing time less than 1 min results in no amplification; alternatively FastTaq polymerase may be used
30, 40, 50, 60 cycles	cycles less than 50 result in very low products
Primers	
forward/reverse primer on beads	Primer direction does not influence with emPCR or emIVTT results
single biotin	nearly none of the ampliified template is able to retain on beads
dual biotin	necessary for emPCR
PEG spacer	Reduce steric effect (not tested)
Emulsion formation	
Span 80 Tween 80	only low amount of templates retains on beads
Modified surfactant	better yield than Span/Tween 80
100, 200, 1000 rpm spinning	200 rpm spinning gives more standardised size emulsions
Emulsion breaking	
breaking buffer	breaks emulsion well but bad for IVTT (Triton-X was used)
H2O-saturated diethyl ether	difficult to handle, samples appeared as white aggregates after wash
hexane	difficult to handle, samples appeared as white aggregates after wash
isobutanol	better separation of samples, visible beads after wash

Table 3.1 factors affecting emulsion PCR.

3.2.8 *In vitro* transcription and translation

As each emulsion PCR reaction produced variable quantity of DNA templates on the beads, it is important to have the knowledge of exact template concentration required for a successful *in vitro* transcription and translation reaction. As shown in **Figure 3.14**, the Western blot staining of a standard *in vitro* transcription and translation reaction without emulsions. For each 50 µl reaction 10 µl of DNA template with different concentrations were used. The Western blot examination of the product indicated that only templates more than 200 ng would be sufficent to produce detectable β-2-microglobulin signals on the blot. Also the level of expressed proteins was not increased if the template concentration was above 400 ng. This information is critical for calculation of the total amount of biotin primers needed to attach to the beads.

We have found that it was difficult to express proteins with *in vitro* transcription and translation if the protein to be expressed has molecular weight smaller than 15 kDa.

Figure 3.14

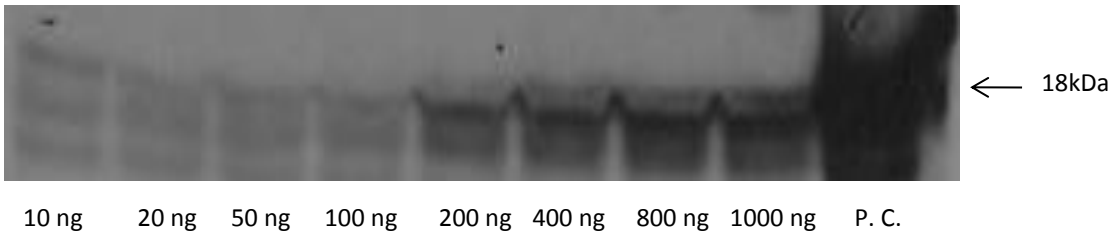


Figure 3.14 Western blot results showing the effect of different dilution of templates on the efficiency of IVTT. The result was the standard *in vitro* transcription and translation with different quantity of DNA templates containing randomised peptide sequences. The blot was stained with anti-human β-2-microglobulin antibody. Positive control, 1 ul synthesised β-2-microglobulin at 1 mg/ml

3.2.9 emIVTT - Translating proteins from DNA templates linking with beads

Following a successful emPCR reaction, the re-suspended bead-DNA complexes will be used to form another emulsion to allow the translation of different protein sequences attaching on each bead. Similar to emPCR, the formation of emulsions in emIVTT requires the dilution of beads such that only one or less than one bead could be present in each emulsion. As the transcription and translation of the template occurred in emulsions, the immediately expressed protein which carried SBP will spontaneously link with the available streptavidin on the bead inside the emulsion. **Figure 3.15** showed the Western blot result of protein-bead-DNA complexes staining with anti-human β-2-microglobulin and streptavidin conjugated HRP. The result showed that all of the three types of beads were able to allow protein expression and subsequent attachment on the beads. Surprisingly, the quantity of primers on the beads before emPCR has no significant effect on the level of protein expression in emIVTT. We reasoned that once the beads were coated with sufficient quantity of the DNA template (as shown in **Fig 3.14**), the proteins would be continuously expressed and occupy the available streptavidin binding space on each bead. Once the expression and attachment of β2m and SBP on the beads were confirmed, this batch of beads will be quantified for refolding with HLA-heavy chains in refolding buffer. We were unable to express proteins at detectable quantity that contained only randomised peptide and a streptavidin-binding-protein (approximately 7 kDa) and protein with only randomised peptide and β-2 microglobulin (approximately 12 kDa). We reasoned that it was because of large protein such as β-2-microglobulin provided overall structural stability to the entire protein, which was more stable under the vigorous washing procedures when the emulsions were broken.

When we confirmed the successful expression of proteins after emIVTT, we were able to use this batch of beads as a bead library to probe against any protein of interest.

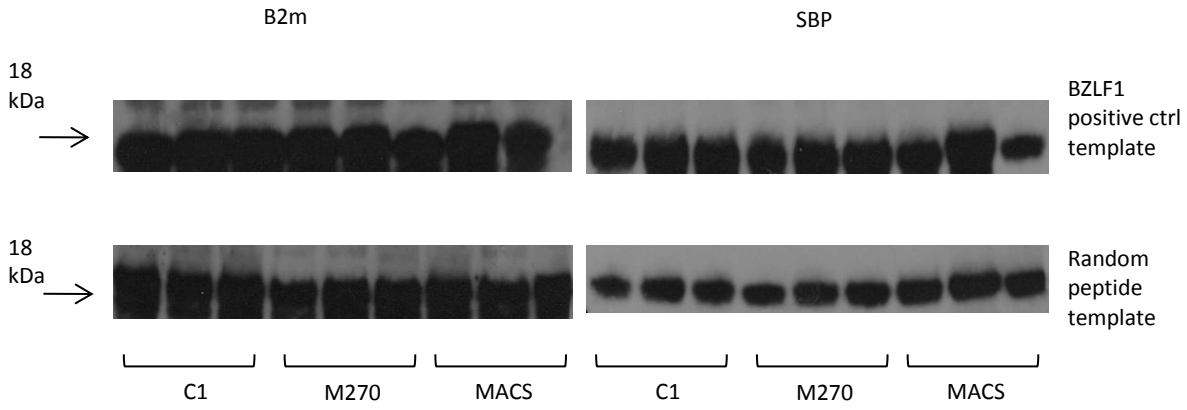


Figure 3.15 emIVTT on different types of beads. The bands were Western blot results showing protein-DNA-beads attaching with either EBV BZLF1 positive control template (known peptide sequence) on upper row, or templates containing randomised peptide sequences (lower row). The left column showed the results of human β-2-microglobulin (β2m) staining and the right ones was streptavidin-binding-protein staining. Each type of beads was labelled with three different concentrations dual-biotin labelled primers attaching on the beads, e.g. for C1 column, 1000 pmol, 500 pmol, 250 pmol primers were attached on 1mg of beads (6-7 x 10^7).

3.2.10 Generation of BZLF1 epitope-specific HLA-B8 refolded MHC tetramers

In order to isolate T-cells with specificity to EBV BZLF1 epitope, tetramers were generated with EBV BZLF1 peptide and HLA-B8 heavy chain and subsequently linked with streptavidin-conjugated PE to generate the BZLF1 antigen-presenting tetramer. As shown in **Figure 3.16**, patient's PBMC were stimulated with BZLF1 epitope RAKFKQLL, maintained for 14 days and stained with anti-CD3, anti-CD8 and the tetramer. This was used to sort BZLF1 antigen-specific T-cell lines (grown from PBMC). Number of attempts had been made to generate BZLF1-specific T-cell clones by pooling the sorted cells in a culture well re-suspended with PBMC feeders. However we were unable to successfully grow the lines into clones so far. Therefore we were seeking well-developed influenza epitope GLCGFVFTL-restricted HLA-A2 T-cell clones (courtesy of Dr Tao Dong) for the rapid test of our system.

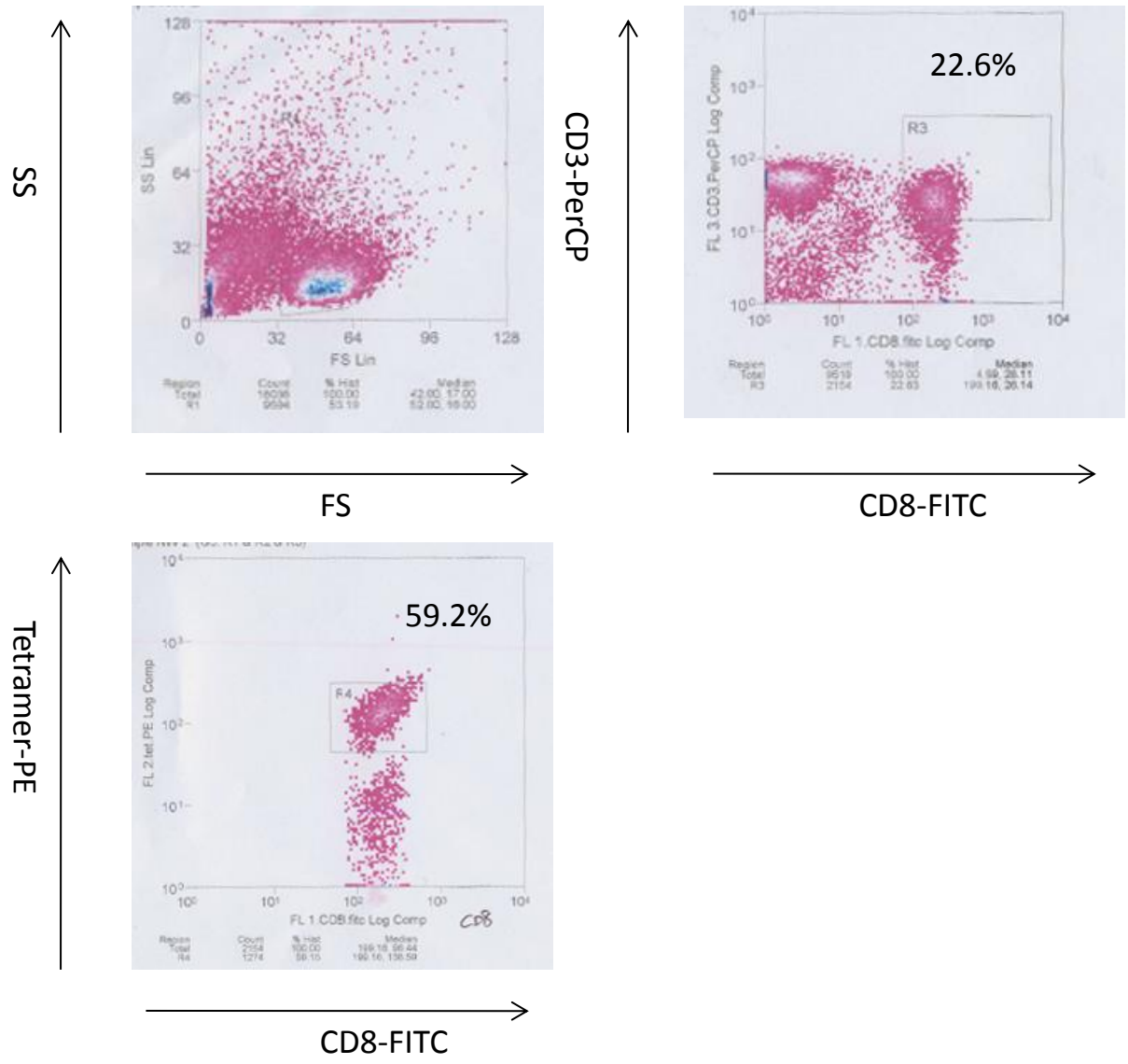


Fig. 3.16 Tetramer staining of antigen-specific B8 cell lines (Day 14) by refolded B8 tetramers.

T-cell lines (BZLF1 epitope RAKFKQLL restricted) were grown from PBMC for 14 days and sorted by staining with anti-CD3 conjugated PerCP, anti-CD8 conjugated FITC and BZLF1-HLA-B8 tetramer conjugated with PE.

3.2.11 the refolding of HLA-B8 MHC molecules on the beads after emIVTT - FACS

In order to probe the bead-linking random peptide library with T-cells, proteins attaching on the beads needed to be folded into antigen-presenting MHC molecules. Although it was supported in the literature that epitope- β2m -HLA-heavy chain could spontaneously refold into MHC molecules (Uger and Barber 1998) (White, Crawford et al. 1999), it was uncertain that the same held true when the proteins were attached on a bead. An ELISA was carried out to examine whether mixing the correct HLA-heavy chain with protein-bead-DNA would trigger the MHC complex refolding. **Figure 3.17** showed the absorbance reading for beads refolded with different types of HLA-heavy chains. W6/32 antibody only recognises the correctly folded HLA-A/B/C molecules so it was used to examine the conformation of MHC molecules. As shown in the graph, at zero dilution, classical refolds (positive control) had OD_{450} reading at approximately 2.7; whereas beads refolded with B8/B7/A2 heavy chains had approximately 3.0, 1.2 and 2.0 respectively, and the negative control had the lowest of all at approximately 1.1. Upon each dilution, the absorbance reading for each samples reduced accordingly, confirming that the W6/32 antibody was binding to the MHC-bead-DNA complexes. Interestingly, while the absorbance readings for samples refolded with HLA-A2 remained low, the readings for HLA-B7 refolded samples remained high, despite the fact that it was still lower than HLA-B8 samples. This could be due to the fact that HLA-B8 and HLA-B7 shared a large degree of sequence homology, thus HLA-B7 molecules could partially react with HLA-B8 restricted EBV BZLF1 epitopes.

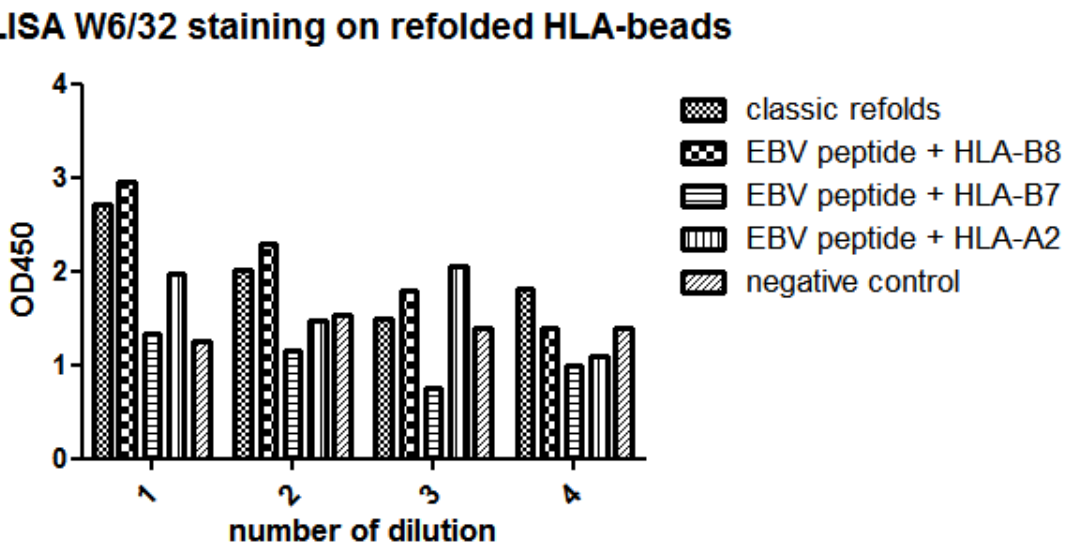


Figure 3.17 ELISA results showing W6/32 antibody staining on refolded HLA-M270 beads. HLA-B8/B7/A2 heavy chains and B2m proteins were synthesised by protein expression. Protein-bead-DNA complexes (template BZLF1 coding sequence) were refolded with three types of HLA heavy chains, washed, and coated on ELISA plate, which was stained with mouse W6/32 antibody and anti-mouse HRP and measured the absorbance. Classic refolds were the refolding of B2m, BZLF1 epitope, HLA-B8 heavy chain, which are acted as positive control for successful refolding. In negative control reaction, only B2m and HLA heavy chains were added. Each dilution had 5-fold difference in bead concentration in descending order.

3.2.12 The refolding of HLA-B8 MHC molecules on the beads after emIVTT – FACS

Since ELISA could not give a quantitative measure of the percentage of beads with correctly refolded MHC molecules using randomised peptide, the beads with refolded HLA-B8 MHC molecules were stained with W6/32 antibody conjugated with PE fluorophore and the samples were tested with FACS. In **Figure 3.18**, the dual-biotin primer-linked Dynabeads were used to perform emulsion PCR to amplify the BZLF1 peptide-β-2-microglobulin-SBP DNA templates, which were linked with the beads by streptavidin-biotin interaction. The DNA-bead complexes were then taken to translate proteins by emulsion *in vitro* transcription/translation, and the subsequent DNA-bead-protein complexes were refolded with free HLA-B8 heavy chains. The final product of antigen-presenting MHC-DNA-bead complexes were stained and analysed by W6/32 conjugated PE antibody. As shown in **Figure**

3.18, the beads were stained by W6/32 antibody even if no proteins were present, the background signal was approximately 7-8%. Despite the background, we were able to see that more than 50% of the beads were stained positive on using HLA refolded beads. If we considered the background staining as 10%, we could reason that more that 40% of the beads carried correctly refolded HLA-B8 MHC molecules. If 40% of the beads were successfully linked with refolded MHC molecules, the number of beads with correctly refolded MHC molecules would be $7x10^6$ (starting number of beads) x 40% = 2.8 x 10^6 beads carrying antigen-presenting MHC molecules. This calculation would be different if the randomised peptides were used to refold. Our result showed that the W6/32 staining of random peptide refolding could yield from 0.1% to 3.2% correctly refolded MHC-bead complexes.

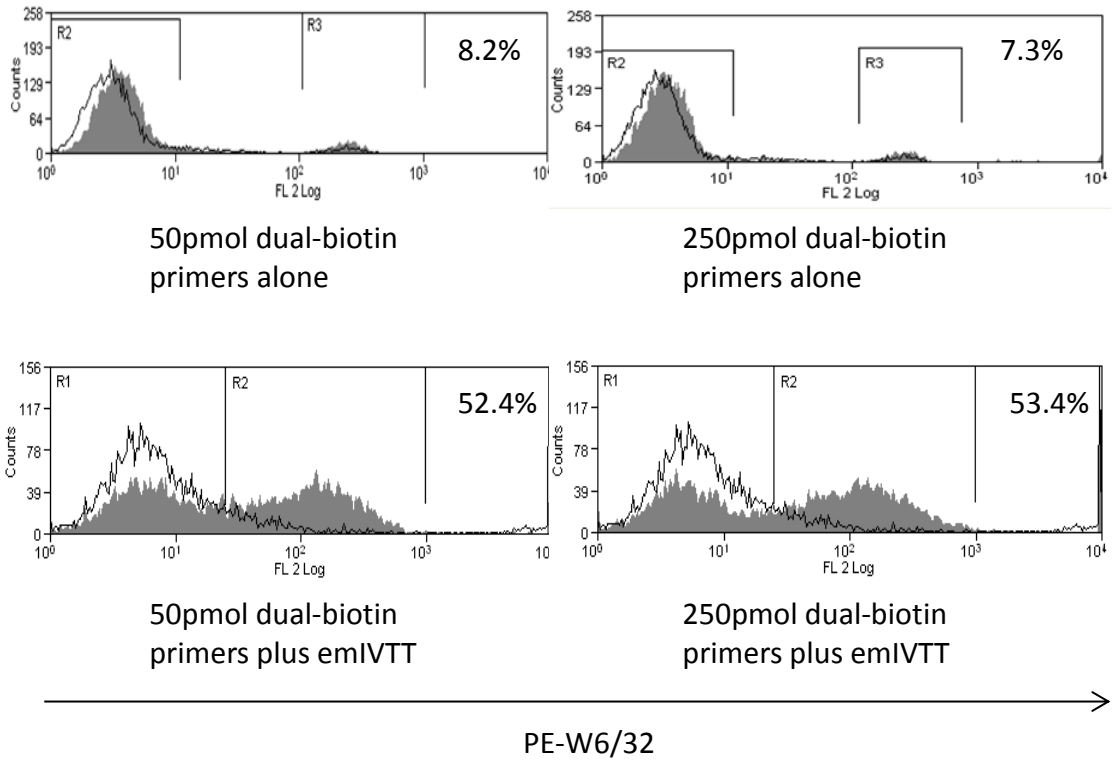


Figure 3.18 FACS staining of beads with refolded HLA-B8 MHC molecules. The refolded beads containing HLA-B8 MHC molecules refolded with EBV BZLF1 epitope (EBV) or randomised peptides were stained with W6/32 antibody conjugated with PE fluorophore. 50/250 pmol represents the amount of dual-biotin labelled primers coated on the beads before emPCR. Black line represented the unstained beads as negative control, whereas the grey shade represented the bead samples stained with W6/32-PE antibody.

3.2.13 Beads linking refolded MHC class I complexes created antigen-presenting multimers

In order to confirm the ability of beads linking with multiple refolded MHC molecules and maintaining the correct conformation for antigen-presentation, fresh beads were linked with pre-synthesised HLA-B8 MHC heterotrimer refolded with BZLF1 epitope and stained with W6/32 antibody. As shown in **Figure 3.19**, only 1 ug of refolded HLA-B8 (stock concentration 1mg/ml) will be enough to saturate more than 90% of the beads, suggesting that the beads were capable of acting as a high affinity antigen-presenting platform.

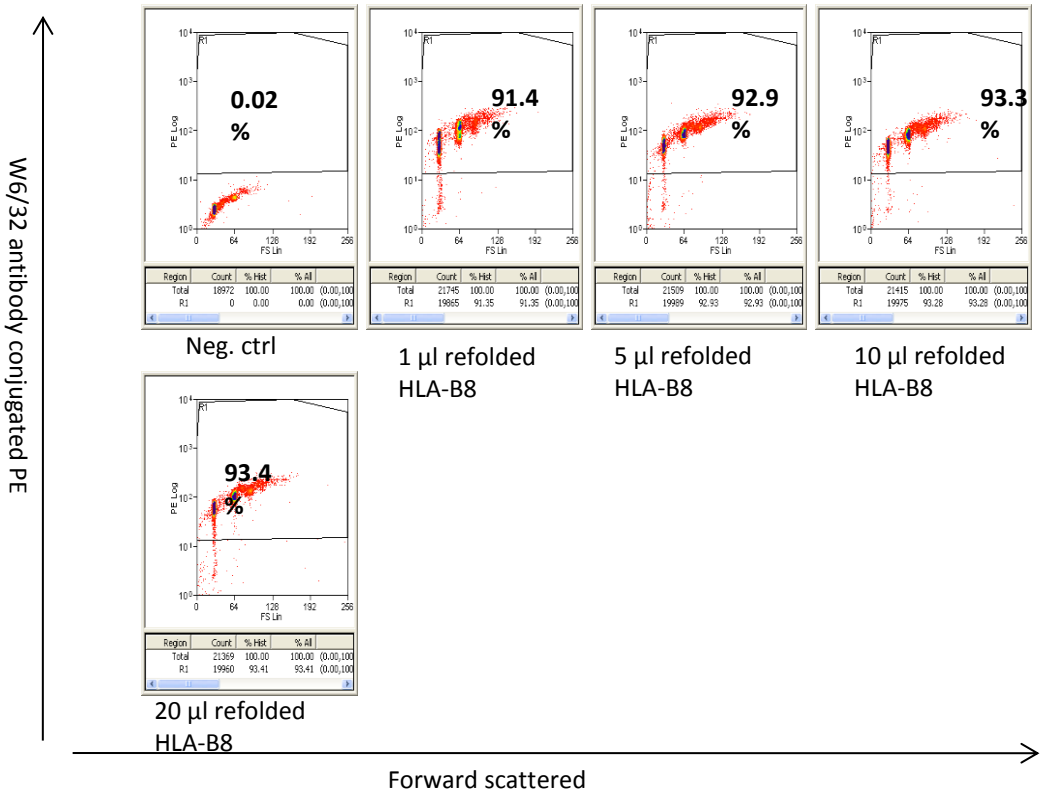


Figure 3.20 FACS staining of beads coated with pre-synthesised refolded HLA-B8 MHC molecules. 7 x 10^6 beads were used to link with pre-synthesised refolded HLA-B8 heterotrimer (HLA-B8+B2m+BZLF1 epitope, concentration 1mg/ml), stained with W6/32 antibody conjugated to PE and analysed by FACS.

3.3 Discussion

3.3.1 Emulsion bead library as an option for next generation peptide selection

We have established a streamlined protocol for the production of a bead library containing randomised peptides or refolded MHC class I molecules for the selection of high affinity peptide or antigen-specific epitopes. **Figure 3.20** illustrates the complete production and quality control in each interval process. The entire process depends primarily on the quality and homogeneity of templates attached on the beads after emulsion PCR. We found that if one batch of beads passed both the microscopy and DNA gel tests, the subsequent emulsion IVTT would definitely be successful with high protein yield. For emulsion IVTT, we found that incubating the emulsions at 32^{o}C produced more proteins (data not shown) than using 30^{o}C suggested by many protocols, and prolonged incubation time (> 4 hours) produced proteins with aggregates. In addition, we also optimised the refolding of MHC class I molecules from the peptide library by miniaturising the refolding reaction to 1 ml.

To quantify each process, we yielded approximately 2 x 10^6 beads (Dynabead C1) in a 50 µl emulsion PCR reaction, with each bead contains approximately 2.4 x 10^5 biotin binding sites for biotinylated primers and streptavidin-binding protein. If we make 10 preparations and pool them together after the emulsion IVTT process, we would in theory produce a peptide library with 10^7 unique peptides (assuming half of the beads were lost during washing processes). Automated emulsion systems will greatly enhance the library to 10^{12}-10^{14}. The advantage of utilising emulsions and magnetic beads provided easy separation and handling during the synthesis process, the cell-free environment means that the use of unnatural amino acids and the production of very acidic or basic proteins which would usually be lethal to the host organisms are possible. In addition, we have optimised the protocols with quick turnover time. The time required for the generation of peptide library from emulsion PCR and emulsion IVTT combined was 2 days, one additional day for the refolding of the library. This

cell-free protocol by-passes the laborious and time-consuming methods which rely on cell growth to synthesise the library. Furthermore, as discussed in Chapter 5, the beads can be used to probe cell populations which are not readily achievable using phage display.

3.3.2 Potential problems encountered by emulsion PCR linking emulsion IVTT methods

Several improvements can be made to increase the fidelity of our systems. Firstly, more beads per PCR/translate/screening could be used to expand the scale of the library. Currently for each reaction we used approximately 10^6 beads, however not all of the beads would be attached with DNA template and not all of the templates would be translated, initially we opted for 10% of the starting beads successfully linked with translated proteins. If we could increase the number of starting beads, we would have a gross higher number of quality bead-protein-DNA complexes. Secondly, it might be beneficial to elute any biotinylated proteins in the *in vitro* transcription/translation kit before the kit was used for emulsion IVTT, techniques such as MACS streptavidin column could be used. However whether the exclusion of these proteins would affect the translation of beads need to be tested. Thirdly, we could order random templates in which the third codons did not contain Alanine and Glycine, hence eliminate the presence of stop codons (TAA, TAG, TGA) in the randomised peptide sequence. However by doing so a large number of amino acids would also unable to be translated. One possible way is to eliminate either A or G in the third codon but not both, thus we reduced the likelihood of stop codons but also preserve the randomness of codon choices.

We were able to demonstrate that this technique could be feasible for the identification and characterisation of antigen-cell interactions. The ability to form high-affinity antigen-presenting MHC molecules with known peptide sequence significantly broadens the scope of

investigations that can be undertaken to study antigen presentation to T-cells, and the possibility of other lymphocytes e.g. NK cells.

The approach therefore has the potential for use in synthetic biology in which whole libraries of simplified proteins constructed *in situ* on a fixed platform and engineered to bind with high affinity with any proteins or cells of interest. Such libraries of simplified, synthetic, and high-affinity proteins would have vast potential both as research tools and as a route to identify novel therapeutics.

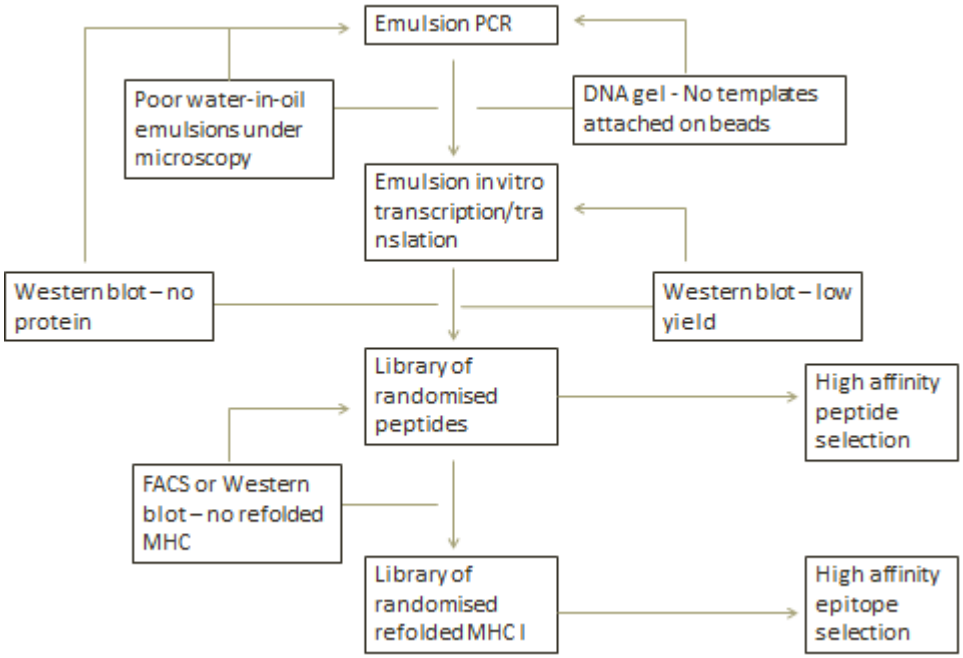


Figure 3.20 overall processes for the generation of bead library and quality control points.

Chapter 4 - Selection of high affinity peptides with emulsion technology – targeting of HIV-1 coat protein gp120

4.1 Introduction

HIV-1 proteins show enormous sequence diversity which is believed to contribute to evasion of the humoral immune response. Until relatively recently, the robust isolation of cross-clade neutralizing antibodies has been rare, with only a handful of well documented antibodies described in the literature.

During the stage of chronic HIV-1 infection, neutralising antibodies target, to a variable degree in different patients, HIV-1 viral envelope protein Env (Willey and Aasa-Chapman 2008; Gray, Madiga et al. 2011). Env is a part of the HIV viral membrane, and it comprises three homologous trimeric gp120/gp41 proteins loosely linked together. Env is embedded in the viral membrane via the gp41 transmembrane subunit and present in a scattered formation on the surface of the virion (Shibata, Igarashi et al. 1999). The HIV-1 virus is thought to evade antibody neutralisation by the post-translational modification of a mixture of surface glycosylation on the Env, inter-strain variability, and conformational masking (McCoy and Weiss 2013).

Since the beginning of HIV-1 studies, only a few neutralising monoclonal antibodies (mAbs) from infected patients has been characterised, the critical regions targeted by these neutralising mAbs targeted are described in **Figure 4.1** (Sheppard and Sattentau 2005). In brief, neutralising mAb b12 was a potent anti-HIV mAb with broadly neutralising ability and high potency. Several studies showed that it targeted the CD4-binding site of Env and could neutralise up to 40% of HIV-1 strains tested (Burton, Pyati et al. 1994). As Env binds CD4 co-receptor on host T-cells, a conformational change takes place that exposes the induced epitopes (CD4i epitopes) which had been shown as the targets of neutralising antibodies (Moulard, Phogat et al. 2002). It had also been found that the neutralising ability of CD4i-targeting mAbs such as 17b could be augmented by the presence of sub-inhibitory

concentrations of sCD4. In addition, a study also found that if the mAb is present as FAB rather than the entire IgG, the potency and broadly neutralising ability of the mAb would be greatly increased. However, these results were completely opposite if the mAb tested was b12, which targets the CD4-binding site (Labrijn, Poignard et al. 2003). These findings implicated that the access route of neutralising mAb to CD4i epitopes on HIV-1 virus could be limited due to conformational changes. As a result, although CD4i binding activity is induced in the sera of immunised humans, the mAbs found in the sera do not have broad neutralisation activity (Vaine, Wang et al. 2010).

Furthermore, mAb 2G12, a neutralising mAb that targets specific glycan domain on Env was identified (Trkola, Purtscher et al. 1996). 2G12 has a domain-swapped conformational structure (Calarese, Scanlan et al. 2003). The N-linked glycans in Env in conserved domains C1, C2, C3, C4 and variable domains V4 were found to be bound by 2G12 (Chaillon, Braibant et al. 2011). A recent study also suggested that modifications in V1, V2 and V3 domains could alter the binding activity of 2G12. The V3 loop has also been shown to be a target for neutralising mAbs (Durda, Leece et al. 1988), in particular the 447-52D mAb (Stanfield, Gorny et al. 2004; Rosen, Chill et al. 2005) and B4E8 mAb (Bell, Pantophlet et al. 2008). Gp41 was also a viable target for neutralising mAbs purified from patients infected with HIV-1. Among all neutralising mAbs, mAb 2F5 (Barbato, Bianchi et al. 2003) and mAb 4E10 were most researched (Stiegler, Kunert et al. 2001), both mAbs bind within the membrane proximal external region (MPER) as well as possessing lipid binding ability (Alam, Morelli et al. 2009). 4E10 shows a broad spectrum of neutralisation across different HIV-1 strains (Mehandru, Wrin et al. 2004), despite the fact that the neutralisation activity is only weak to moderate potency (Binley, Wrin et al. 2004).

The studies of neutralising mAbs had led to great understanding of HIV-1 in terms of structures and sites for neutralisation targeting, but the highly polymorphic nature of HIV-1

viruses remained a significant concern as to whether we could ever generate mAbs with complete coverage of all HIV-1 strains and achieve neutralisation potency. Therefore scientists have raised doubt to whether induced host immune response could have a long-lasting and protective effect. In addition, although studies had shown that when mAbs were introduced to non-human primates there would protect from infection (Mascola, Lewis et al. 1999; Shibata, Igarashi et al. 1999; Parren, Marx et al. 2001; Veazey, Shattock et al. 2003; Hessell, Rakasz et al. 2009; Watkins, Siddappa et al. 2011), viral mutants developed in animal models and led to viral escape (Andrus, Prince et al. 1998; Poignard, Sabbe et al. 1999). Similar findings were also reported in patients who had intermittent antiretroviral treatment with neutralising mAb transfer. (Trkola, Kuster et al. 2005).

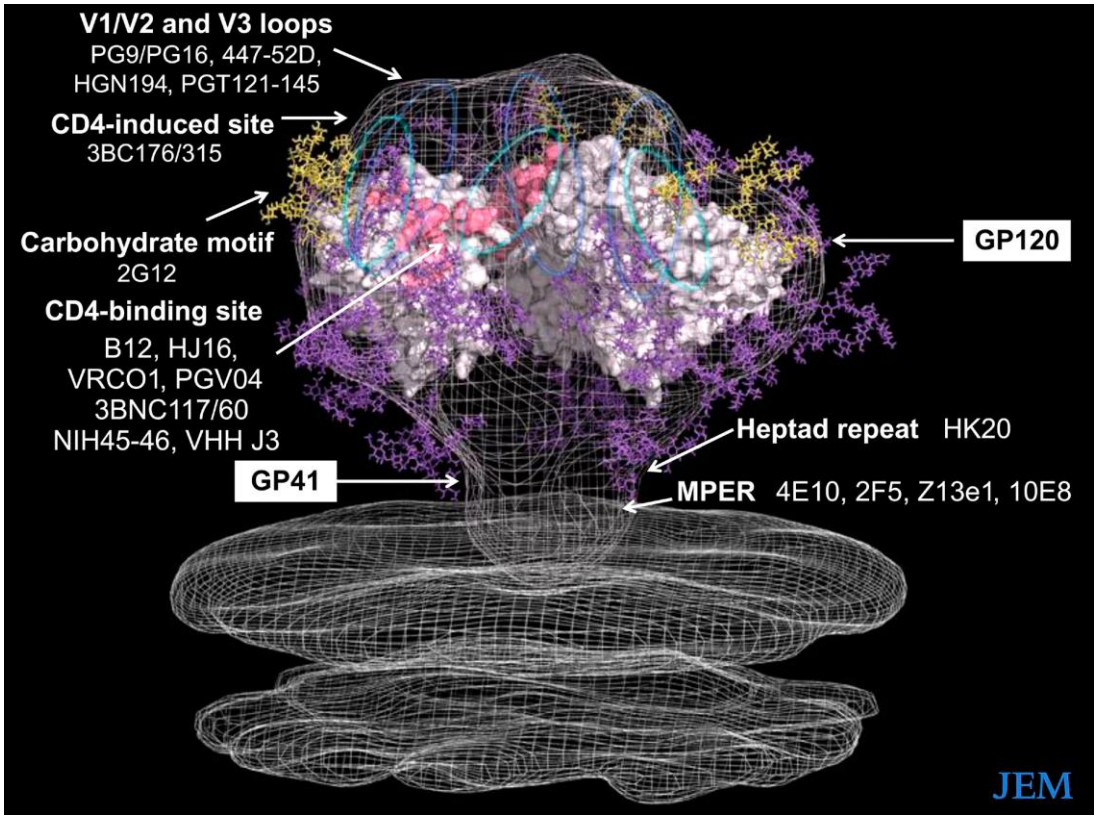


Figure 4.1 Neutralising mAbs targeting HIV-1 Env. Adapted from (McCoy and Weiss 2013)

However with advances in antibody isolation technology, it is becoming apparent that infected individuals not only develop neutralizing antibodies, but these can be isolated and characterized (Walker, Phogat et al. 2009; Wu, Yang et al. 2010; Walker, Huber et al. 2011). With high levels of somatic mutation, it seems likely that parrying rounds of mutation and selection occur between Env and the evolving humoral immune response. There are some common features of the emerging Env target sites, including the CD4-binding site, quaternary V2/V3 loops, gp41 Membrane Proximal External Region and Env carbohydrates (McElrath and Haynes 2010). Monoclonal antibodies based on gp120 protein antigens had shown to have neutralisation ability (Flynn, Forthal et al. 2005; Pitisuttithum, Gilbert et al. 2006; Gilbert, Wang et al. 2010). Indeed Env acquisition of an extensive evolving glycan shield is believed to restrict access to conserved protein backbone amino acids (Wei, Decker et al. 2003). Interestingly, many of the isolated antibodies show poly-reactivity including self-reactivity which may be important for neutralization capacity (Mouquet, Scheid et al. 2010). Understanding of the characteristics of neutralizing antibodies is important for the development of vaccination strategies, but is also of importance for the development of HIV-1-binding or HIV-1-fusion inhibitors. For example, Enfuvirtide is a licensed peptide HIV-1 fusion inhibitor that through binding to gp41 inhibits HIV-1 fusion with the target cell membrane. Given the glycan shield protection, we hypothesized that peptides may offer advantages over the relatively large Env-specific antibodies, by targeting small accessible niches on the protein backbone. Peptides may offer further advantages such as the relative stability of stapled, cysteine-linked peptides that in some cases, are reported to survive the gastric environment and be potentially suitable for oral administration, e.g. ciclosporin (Craik, Daly et al. 1999). Furthermore peptides may provide structural clues to aid the subsequent development of small drug compounds or other peptide mimetics.

We screened ORBIT bead display peptide libraries on HIV-1 gp120 and identified peptides after a single round of selection that bound gp120 with micromolar affinities. The highest binding peptides competed for gp120 binding by antibodies specific for the V3 loop. Peptides showed cross-clade neutralising capacity. ORBIT bead display is a powerful technology to present peptides on a scaffold of β-2-microglobulin, which can be used to select against disease relevant targets and therefore is a new approach to leads for drug discovery.

4.2 Results

4.2.1 Validating the binding of predicted peptides to bovine serum albumin

In order to confirm that the peptide-bead library was capable of selecting target-binding peptides and their encoding sequences, we tested the ability of the library to select high affinity bovine serum albumin-binding peptides. After the library selection, six peptides were selected and synthesised. **Figure 4.2a** showed the result of the ELISA binding experiment of these peptides against immobilised BSAs. The peptides were prepared in 3 serial dilutions at very low concentration (1-100 nM) and two negative control peptides which were confirmed to have no binding with BSA were also used. As shown in the graph, all six peptides (BSA 1-6) had higher absorbance compared to the control peptides in any given concentration. The binding activity of these peptides to BSA was also dose-dependent. As a result, we were able to confirm that the peptide-bead library we generated was able to select BSA-binding peptides, and reproduced them based on the sequences encoded on the beads.

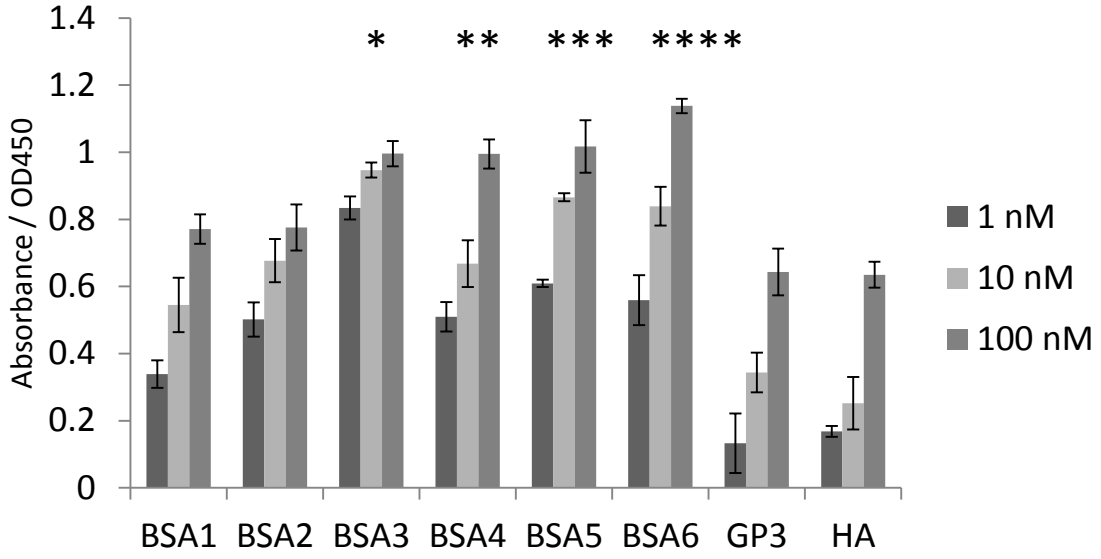


Figure 4.2a ELISA experiment to test the binding of synthesised peptides to BSA. 1 µg/ml purified BSA was immobilised on ELISA plate and 1-100 nM of peptides were added to test the binding activity. Negative control peptides were GP3 (gp120-binding peptide) and HA (hemagglutinin-binding peptide). *p-*values* < 0.05

peptide	sequence	Hydrophobicity
BSA1	L A V R K R M P H	hydrophilic
BSA2	I T F D V Y K V A	hydrophobic
BSA3	I A N G R H F Q H	hydrophobic
BSA4	I R N Q D R C G L	hydrophilic
BSA5	K V H S I M A W K	hydrophilic
BSA6	V P L F K K Q F E	hydrophilic
GP3	S Y S F C P R R Q	hydrophilic
HA	R T L G V M V A F	hydrophobic

Table 4.2b peptide sequences of BSA1-6 and controls

4.2.2 Validating the binding of predicted peptides to gp120 - ELISA

By validating the ability of the library to select BSA-binding peptides, we proceeded to utilise the technique for the selection of HIV-1 coat protein gp120.

Twenty two predicted peptides were synthesised and tested in different batches of binding to JRFL gp120 protein. Peptide GP1-17 was selected from the 9 random amino acid libraries, whereas GP18 to 22 were selected from the 11 random amino acid library containing one cysteine bridge (**Figure 4.3b**). **Figure 4.3a** showed the result of ELISA tests of the twenty two peptides. All peptides were prepared in 4 serial dilutions and the absorbance was read at OD450. An irrelevant peptide (flu virus HA binding peptide) was used as a reference for the peptides which has no binding to gp120. As shown in **Figure 4.3a**, GP2, GP3, GP11 and GP16 had significantly higher binding to gp120 than the control hemagglutinin-binding peptide HA in a dose-dependent manner. At 1000 nM concentration, GP2 and GP11 had approximately twice the absorbance reading compared to the control. At 1000 nM and 125 nM, GP16 had absorbance readings, which were 2.5 and 1.5 times higher than the control. Interestingly, although the peptides we synthesised were few, we were not able to observe any positive binding with the peptides that had cysteine bridge. It is possible cysteine bridges cross link neighbouring peptides and therefore lose the conformation of individual peptides. While being a potentially useful screening tool, ELISA is susceptible to detection of non-specific and low affinity binding. We therefore proceeded to determine specific binding affinities using surface plasmon resonance.

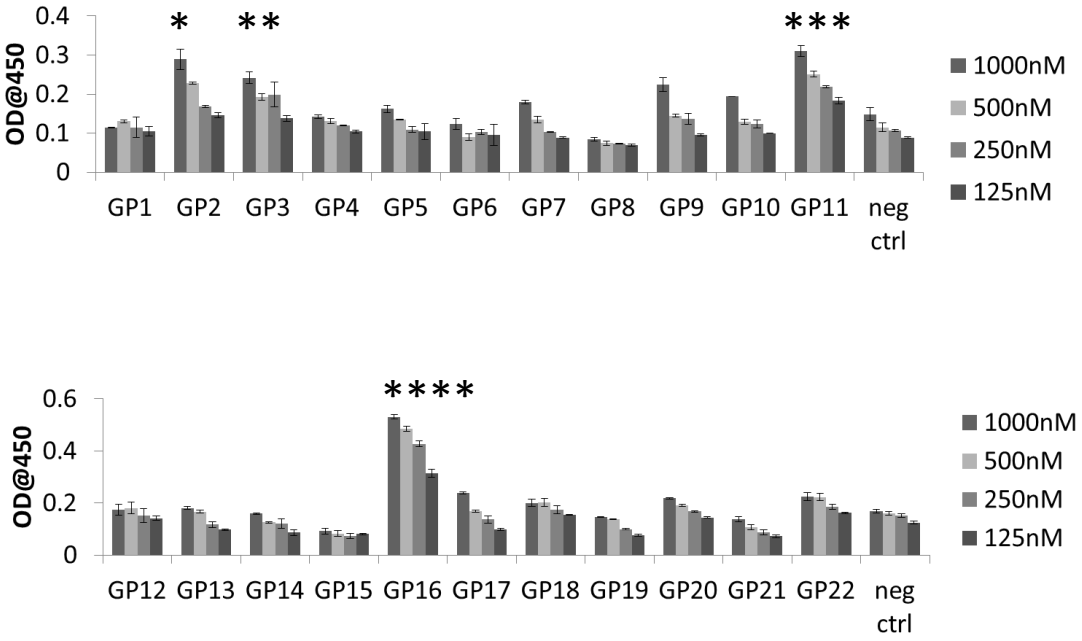


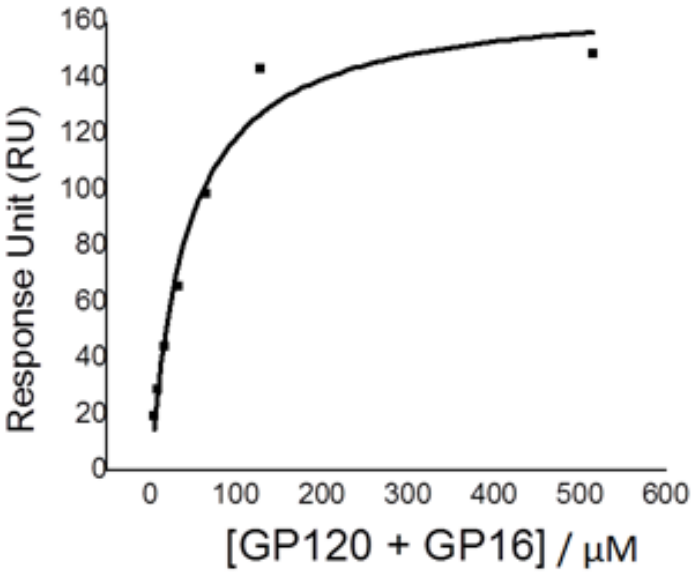
Figure 4.3a Two batches of ELISA for the binding tests at OD450 between predicted peptides and GP120 protein. Some predicted peptides (GP2, GP3, GP11, GP16) showed higher absorbance, i.e. peptides bind to GP120 in a dose-dependent manner, compared to the negative control (HA, hemagglutinin-binding peptide). While most peptides bind to GP120 briefly, GP16 showed the highest binding affinity to GP120 among other peptides. (T-test against negative controls at 1000 nM, GP2 (p = 0.003), GP3 (p = 0.043), GP11 (p = 0.014), GP16 (p = 0.001); n = 3)

peptide	sequence	hydrophobicity
GP1	MSTSCCLPS	hydrophobic
GP2	RRRNVPLIL	hydrophilic
GP3	SYSFCPRRQ	hydrophilic
GP4	AVNIVGYSN	Hydrophobic
GP5	LVLLSSASS	Hydrophobic
GP6	LRPLLSSMR	hydrophilic
GP7	LVLLLSSMS	Hydrophobic
GP8	EMANRRSGS	hydrophilic
GP9	WFFTSLVTC	Hydrophobic
GP10	SILSSCWQS	Hydrophobic
GP11	LLALSAYMV	Hydrophobic
GP12	RPLPILAPC	Hydrophobic
GP13	GTRVFMRSL	hydrophilic
GP14	LSVLSRGML	Hydrophobic
GP15	WHVLIWLLF	Hydrophobic
GP16	LWCRRLNLL	Hydrophobic
GP17	LWHTRRGPG	hydrophilic
GP18	SSCAGNLPREDQECC	hydrophilic
GP19	SSCQRLFVIAYPFKC	Hydrophobic
GP20	SSCCFLPASWLKSLC	Hydrophobic
GP21	SSCSVRSSCRLYYSC	Hydrophobic
GP22	SSCLLFTSMLPDKTC	hydrophobic

Table 4.3b peptide sequences of GP1-22

4.2.3 Validating the binding affinity of predicted peptides to gp120 – surface plasmon resonance

In order to obtain quantitative measurements of binding reactions between predicted peptides and gp120, surface plasmon resonance (BIAcore) was used to measure the equilibrium dissociation constant K_D of the predicted peptides and gp120. Six peptides which showed higher absorbance in the ELISA experiments were selected for BIAcore experiments, together with five predicted to have intermediate or no binding from the ELISA. Since the K_D to gp120 is relative to different analytes, purified protein HLA-A*0201 heavy chain was used as a reference control to measure the relative binding affinity of peptides to gp120. Purified gp120 and HLA-A*0201 were immobilised on chips, and the peptides were passed through both flow cells at a constant rate of 10 μl / min. **Figure 4.4** showed an illustrated example of the K_D calculation for GP16 peptide. K_D for different peptides is shown in **Figure 4.4**. In general the peptides with positive binding on ELISA had a K_D value ranging from approximately 20 to 80 μM, indicating that these peptides had the ability to bind gp120 protein, but with relative weak affinity. The BIAcore experiment acted as a second confirmation that the randomized peptide library had successfully captured peptides which bind to gp120 in a single round of selection, and that the sequence information coding the binding peptides was correctly preserved and sequenced. Although binding affinity was in the μM range, we were aware that this was also the case of Enfuvirtide T20 which is a licensed therapeutic for HIV-1, and so proceeded to determine whether the peptides had HIV-1 replication inhibition activity either as monomers or as multimeric complexes.



peptide	KD / µM
GP1	75.2
GP2	78.1
GP3	39.1
GP5	70.9
GP6	no binding
GP7	no binding
GP8	no binding
GP9	no binding
GP10	no binding
GP11	48.3
GP16	42.7

Figure 4.4 Equilibrium dissociation constant K_D calculations using surface plasmon resonance (SPR). An illustrated example of the K_D calculation for GP16 peptide. The left of the figure is the non-linear fit of the Langmuir binding isotherm for GP16 binding. The table summarises the K_D of the peptides tested.

4.2.4 HIV-1 viral inhibition – determining the optimum HIV infection

We were interested in learning if the peptides we identified were able to inhibit HIV-1 infection by blocking the gp120 domains which are responsible for HIV-1 viral entry. Therefore an HIV-1 inhibition assay was conducted with the addition of these peptides prior to HIV-1 infection.

Figure 4.5 shows the HIV-1 infection rate measured by staining HIV-1 core antigen p24, HIV-1 BAL stock virus were added at different concentration from 12 to 20 µl (MOI = 0.01) and incubated with healthy PBMC for 3 to 5 days. As shown in **Figure 4.5**, approximately 17.5% of CD4+ cells were infected with 12 µl HIV-1 BAL, and doubling the virus to 25 µl only increased the infection by 2% (Day 3). In addition, we also observed a decline in infection with increased incubation time (Day5) and viruses. Therefore we had decided to use 20 µl of stock HIV-1 virus (MOI = 0.01) for the following HIV-1 viral inhibition assay with peptides.

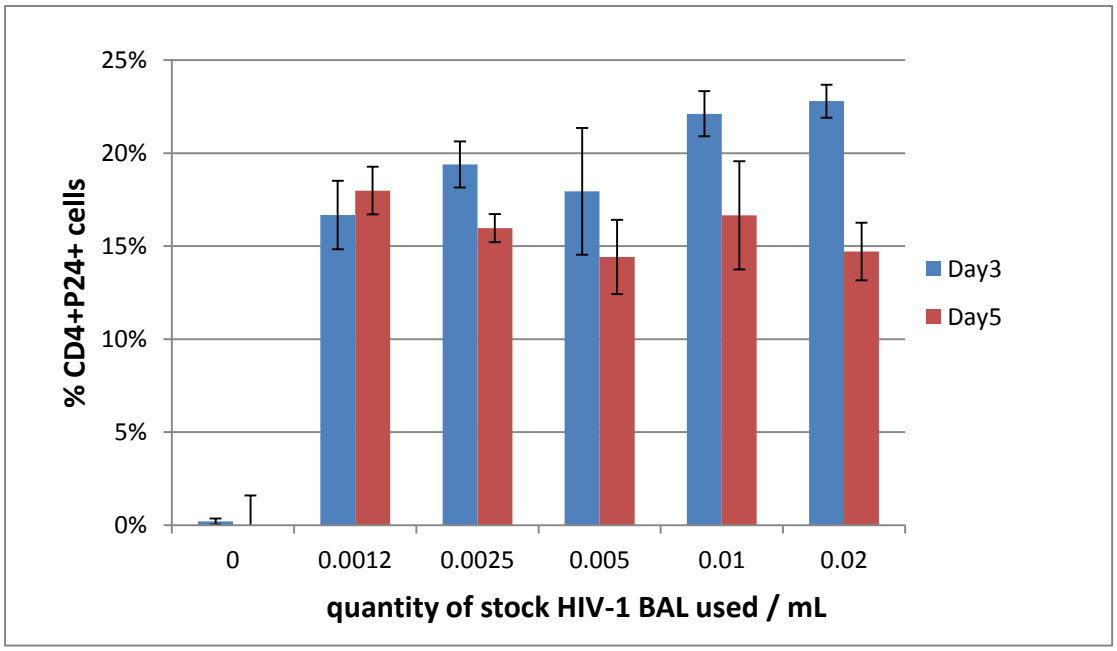


Figure 4.5 HIV-1 BaL infection to healthy PBMC. HIV-1 BaL stock MOI = 0.01. Different quantities of stock virus were incubated with PBMC for 3 or 5 days and stained, the infection of a single cell was determined as $CD3^{+}CD4^{+}CD8^{-}p24^{+}$.

4.2.5 HIV-1 viral inhibition with streptavidin-conjugated peptides – HIV-1 BaL

Several peptides which bound to gp120 were selected and incubated with HIV-1 BaL stock virus before incubating the virus with healthy PBMC. **Figure 4.6** showed the HIV-1 BaL inhibition assay with peptide administration. The infection was divided into two groups, for the duration of 3 or 5 days at 37^{o}C. As shown in **Figure 4.6**, 20 µl of HIV-1 BaL had approximately 5% infection for the untreated cells, and increased to 5.7% in average if the infection was left for 5 days. The peptides were administered at 50 µM conjugated with streptavidin. Among all administered peptides, GP3, GP5 and GP11 showed a reduction in HIV-1 infected cells at Day 3, in which the percentage of infection reduced to 3%, 2.5% and 3.5%, p = 0.044, 0.005, 0.012 respectively. However, the reduction effect disappeared if the infection was carried on for 5 days. We reasoned that it was because the peptides could inhibit the HIV-1 viral entry temporarily, but the proteases present in the serum digest the peptides hence the viruses were free to enter the cells again.

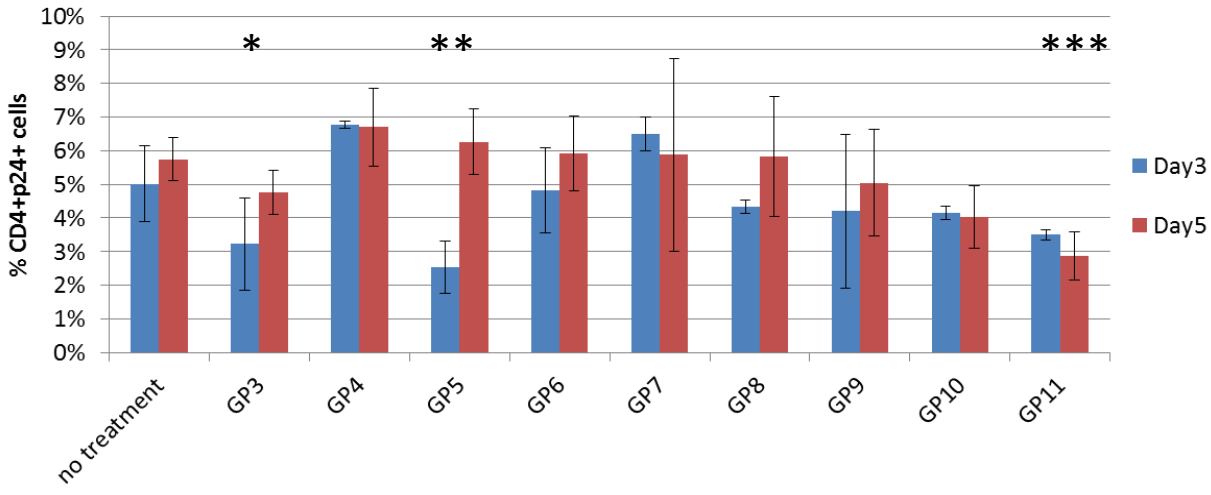


Figure 4.6 HIV-1 BaL inhibition assays. Peptides GP3-11 were administered in 50 µM conjugated with streptavidin. *p-values* <0.05

4.2.6 HIV-1 viral inhibition with streptavidin-conjugated peptides – HIV-1 clade C

Since GP3, GP5 and GP11 appeared to be able to reduce HIV-1 BaL infection, we were interested to know if the peptides showed similar reduction in different HIV-1 strains. **Figure 4.7** showed the HIV-1 inhibition assay with clade C virus. Streptavidin and BSA were added as negative controls and the Enfuvirtide (T20) was introduced as the positive control for inhibition. GP3, GP5 and GP11 and other peptides were added to investigate the HIV-1 reduction effect. As shown in the graph, the three peptides which showed reduction in HIV-1 BaL infection were also capable of reducing the HIV-1 clade C infection. While all other peptides failed to show a significant reduction compared to the negative controls, GP16 showed similar infection level to the GP5 and GP11 peptides, which were lower than controls.

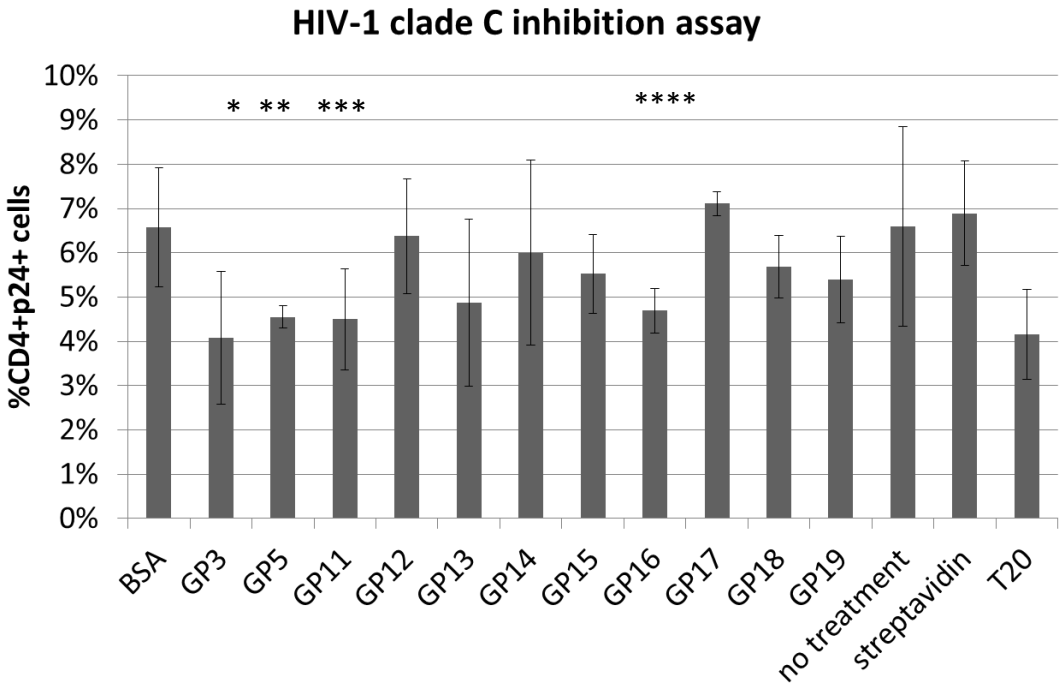


Figure 4.7 HIV-1 clade C inhibition assays. Peptides GP3-11 were administered in 50 μM conjugated with streptavidin. P24 staining was conducted 3 days post-infection.

4.2.7 HIV-1 viral inhibition with streptavidin-conjugated peptides – BaL, clade C, IIIB

Figure 4.8a shows the HIV-1 replication assay using biotinylated forms of the four peptides of highest affinity, with or without streptavidin conjugation (GP3, GP5, GP11, and GP16). In general, peptides administered directly without streptavidin conjugation had a very low inhibitory effect on HIV-1 replication, e.g. GP16 inhibited HIV-1 BaL infection by less than 1% at 50 and 100 μM, and 50% at 250 μM; GP3 had no inhibitory effect at 50 μM, and 34.5% and 54.7% inhibition at 100 μM and 250 μM respectively (**Table 4.8b**). Without streptavidin conjugation, peptides only exhibited viral inhibition at high concentration, and the effect was weak compared to the positive control Enfuvirtide (T20) (91.1% inhibition at 50 μM). Surprisingly, when the streptavidin conjugated peptides were used, the inhibition was efficient and required concentrations comparable to Enfuvirtide T20, e.g. GP16 achieved 62.6% viral inhibition at 50 μM when conjugated with streptavidin. Similar patterns could be observed with different peptides and different HIV-1 clades. GP16, GP3 and GP11 exhibited cross clade HIV-1 inhibitory activity, and the effect was augmented with streptavidin conjugation. As streptavidin conjugation did not significantly influence K_D (data not shown), we reasoned that it may provide a steric hindrance protection to enzymatic digestion and/or preserve the peptide integrity or multiplicity. Of note, GP21 did not show any binding on ELISA and did not show any HIV-1 inhibition, consistent with the ELISA being useful screening tool for subsequent biological activity.

As gp120 is a highly polymorphic protein, HIV-1 neutralising antibodies are an invaluable tool to assess binding sites within the GP120 molecule. Three commercially available HIV-1 broadly neutralising antibodies V3 (anti-V3 447-52D, CCR5/CXCR4 binding sites), VRC01 (discontinuous CD4 binding site) and F105 (CD4 binding site) were prepared to test the binding sites of GP3 and GP16 peptides as these showed the greatest inhibition of HIV-1

replication. As shown in **Figure 4.8c**, peptide GP3 showed significant reduction ($p = 0.025$) in absorbance when gp120 was pre-blocked with the V3 antibody, whereas GP16 had no reduction in gp120 blocked by any of the three antibodies, suggesting that GP16 bound to gp120 on other domains. Our data showed that peptide GP3 exhibited inhibition of replication of the HIV-1 clade C isolate and HIV-1 BaL, which are CCR5 tropic viruses; however no inhibition was observed when GP3 was co-incubated with HIV-1 IIIB, which is a CXCR4 tropic virus. This evidence suggested that peptide GP3 may be binding to gp120 CCR5 binding region. Indeed, we observed a significant reduction in recombinant CCR5 binding to gp120 that had been pre-incubated with GP3 (65.8% reduction in absorbance at 125ng peptide, $p = 0.0038$)

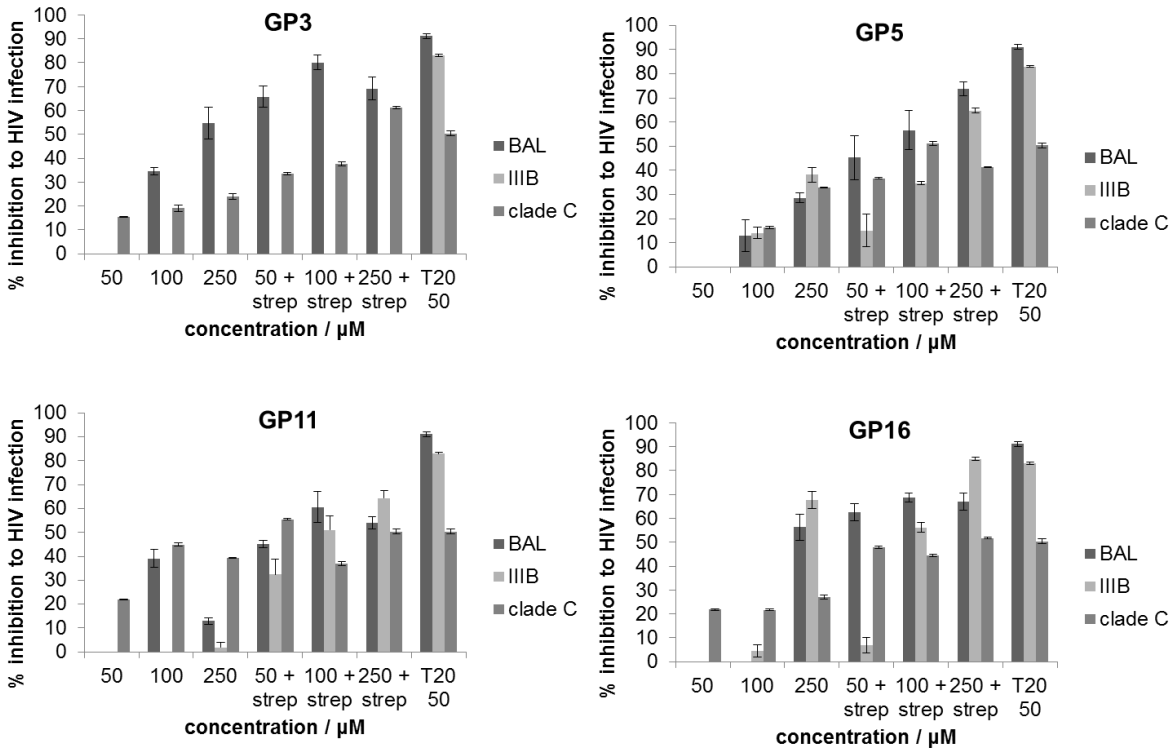
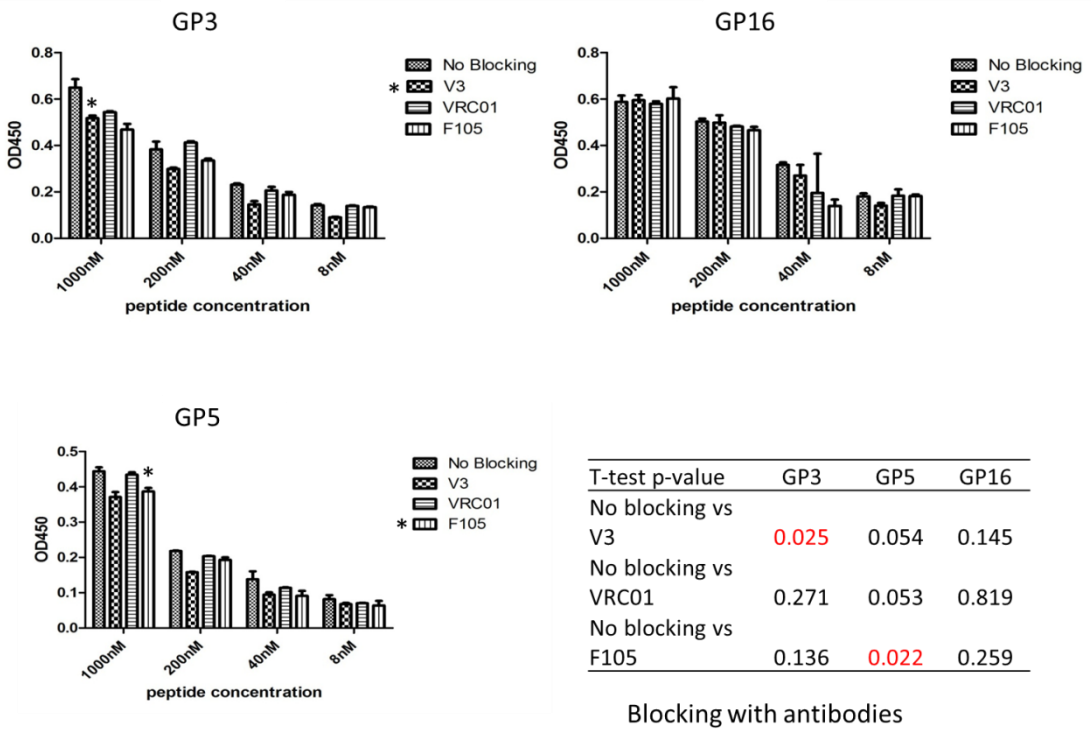


Figure 4.8a HIV-1 (HIV BAL, IIIB, clade C) inhibition assays. Peptides GP3, 5, 11, 16 were administered in 3 concentrations with or without streptavidin conjugation. T20 (Enfurvitide) was used as positive control.

	concentration/uM	HIV BAL/ 50ul % inhibition to HIV infection	HIV IIIB / 50ul % inhibition to HIV infection	HIV C / 50ul % inhibition to HIV infection
GP16	50	no	no	21.8±0.3
	100	no	4.4±2.5	21.7±0.3
	250	56.3±5.5	67.8±3.6	27.0±0.8
	50 + strep	62.6±3.6	7±3.2	47.9±0.5
	100 + strep	68.7±1.8	56.2±2.0	44.4±0.6
	250 + strep	67.0±3.6	84.9±0.8	51.7±0.3
GP3	50	no	no	15.4±0.2
	100	34.5±1.5	no	19.1±1.4
	250	54.7±6.7	no	23.9±1.1
	50 + strep	65.8±4.3	no	33.6±0.5
	100 + strep	80.1±3.0	no	37.6±0.9
	250 + strep	69.2±4.9	no	61.2±0.4
GP5	50	no	no	no
	100	12.8±6.6	14±2.4	16.3±0.4
	250	28.6±20.0	38.2±3.0	32.8±0.2
	50 + strep	45.3±9.2	15.1±6.8	36.7±0.4
	100 + strep	56.7±8.0	34.7±0.7	51.1±0.8
	250 + strep	73.8±2.8	64.8±1.0	41.3±0.1
GP11	50	no	no	22.0±0.2
	100	39.1±3.8	no	44.9±0.6
	250	12.9±12.6	1.8±2.3	39.3±0.1
	50 + strep	45.1±14.2	32.5±6.3	55.5±0.3
	100 + strep	60.6±6.5	51.1±5.9	36.9±0.8
	250 + strep	54.1±2.6	64.3±3.2	50.4±1.0
GP21	250	no	no	no
	250 + strep	no	no	no
T20	50	91.1±1.1	83.0±0.4	50.4±1.0

Table 4.8b table of 4.8a with quantitative data.



T-test p-value	GP3	GP5	GP16
No blocking vs V3	0.025	0.054	0.145
No blocking vs VRC01	0.271	0.053	0.819
No blocking vs F105	0.136	0.022	0.259

Figure 4.8c antibody blocking ELISA experiment. Immobilised gp120 were blocked with three types of antibodies V3, VRC01 and F105 before peptides were added. Absorbance for each blocking was compared against unblocked samples and the *p*-values were shown on the table.

4.3 Discussion

High sequence variation and the evolving glycosylation status of gp120 present challenges to the development of cross-clade neutralizing antibodies specific to HIV-1. However there are conserved regions that can be reached despite glycosylation shielding, such as the CD4 binding site. Known CD4 binding site specific antibodies are often highly mutated suggesting multiple rounds of selection and affinity maturation before effective targeting can be obtained. We hypothesized that peptides may access partly obscured regions of gp120 that may not be accessible by most conventional antibodies. Such peptides may have diagnostic, monitoring or therapeutic uses.

Using the bead display approach (ORBIT) with b2m as a scaffold structure to present variant peptides we screened for reactivity to gp120. The use of b2m as a scaffold allows effective translation and is known to deliver linked peptides to tertiary molecules. Such an *in vitro* transcription/translation approach offers many advantages over in vivo systems including phage display. Advantages include the possibility of using extremely large library sizes that are not limited by cloning efficiency into organisms. *In vitro* expression does not risk bias away from variants that confer growth disadvantage to particular phage or advantage to non-recombined phage. The beads are very simple, containing just peptide/protein and the encoding DNA; this means that there are fewer opportunities for non-specific interactions and that selection rounds can be reduced. Furthermore selection against a target can be undertaken in specific milieu that may not be possible using in vivo presentation, e.g. low pH that may be relevant for viral membrane fusion inhibitors. Lastly it is feasible to modify peptides whilst on the surface of the beads, for example through stapling or use of unnatural amino acids – this is challenging whilst using in vivo approaches.

Instead of using fluorescence based bead selection, we have used a binding cell system based on molecular interactions to retain the selected beads. This has many advantages, as small numbers of beads can be selected from large libraries without dependence on threshold levels of fluorescence for detection. Furthermore the beads can be readily isolated after each round of selection and transferred to sequential flow cells with different targets or conditions (eg target concentration, temperature, pH, buffer constituents) in order to sustain serial positive and/or negative selection to define multiple characteristics. Beads do not have to be re-derived between each sequential step and so even multiple selections can be performed quickly. The peptides may be functionally important, but in addition, they may provide structural data that is of value to the design of small drugs or other peptide mimetics. Furthermore, once the peptide sequence has been defined then further subsequent variants can be readily introduced such as the use of unnatural amino acids. Although we wanted to use peptides defined after a single round of selection, to assess whether the micromolar affinity peptides could have HIV-1 replication inhibition activity, it is clearly possible to use multiple rounds of selection to pick out peptides with higher affinity. It is also possible to constrain peptides through the use of cysteines at defined positions in order to facilitate the formation of secondary structures. Peptides (and peptide drug precursors) offer potential advantages over antibody approaches, as the inclusion of certain sequence and structural components can allow the absorption from the gut and the targeting of intracellular proteins as shown by ciclosporin for example. Lastly, peptides are readily amenable to multimerisation and modulation of pharmacokinetics such as through the use of pegylation with multiple peptide binding sites.

The binding of the GP3 peptide to GP120 can be partially blocked by anti-V3 antibody. This gave an indication that anti-V3 antibody might share the same binding site to GP3. Although the exact binding site for GP3 peptide remains to be elucidated, the critical epitope motif

recognised by 447-52D anti-V3 antibody is the GPxR-motif at the V3 region (Zolla-Pazner 2004) and this is partly shared by residues within the GP3 peptide (CPRR). We reasoned that the reduction in absorbance in anti-V3 antibody blocked ELISA experiment was due to competition binding of anti-V3 antibody to GP3 peptides, resulting in less GP3 peptide binding to gp120 on the plate. The gp120-binding peptides themselves, their variants or small mimetic drugs may therefore have diagnostic, monitoring or therapeutic uses. In addition, we showed reduction in binding between GP3-gp120 and recombinant CCR5, suggesting the possibility that a part of the CCR5 binding domain (possibly the V3 loop residues) on gp120 could bind to GP3 peptide, preventing the access of recombinant CCR5 protein. Recent studies also suggested the possibility that small peptides derived from CCR5 are capable of binding to HIV-1 gp120 and have neutralising effect (Schnur, Noah et al. 2011; Dogo-Isonagie, Lam et al. 2012). Furthermore, it has been known that small cyclic peptides such as β-defensin 2 and 3 have anti-HIV-1 activity (Quinones-Mateu, Lederman et al. 2003). The advantage of working with small peptides including easy modification and production could make a promising therapeutic potential in HIV-1 treatment.

In summary, ORBIT bead display of peptides using β2m as a scaffold base is a powerful approach to the identification of peptide ligands of disease relevant targets. There is broad applicability of the technology to the identification of drug leads for targets of interest.

Chapter 5 - High-throughput identification of CD8+ T cell specificity using HLA-peptide bead libraries

5.1 Introduction

$CD8^+$ T cell recognition of antigen peptides presented by class I human leucocyte antigen (HLA) is an essential step in adaptive immunity. An ability to identify the specificity of T cells would be an enormous step forward in the understanding of the pathogenesis of T cell associated disease such as tumours (Hanahan and Weinberg 2011), infections (Appay, Douek et al. 2008) and autoimmune diseases(Walter and Santamaria 2005; Goverman 2009). Studies to identify peptide antigens recognised by T cells usually require knowledge of the target proteins of interest (Lee, Ha et al. 2008; Precopio, Butterfield et al. 2008). Such studies provide important information in antigen presentation and vaccination design, but they cannot be easily adapted to probe a T cell population with a completely unknown specificity. Clearly, if the specificity of tissue infiltrating T cells could be defined then such an approach would provide significant insights into disease pathogenesis.

It has been shown that pathogenic T cells can be isolated from patient tissues and the matching TCR α and β chains from these T cells cloned and expressed (Seitz, Schneider et al. 2006). T cell hybridoma lines expressing these cloned T cell receptors (TCRs), together with plasmid-encoded combinatorial peptide libraries presented by COS-7 cells, have been used to search for antigens (Siewert, Malotka et al. 2012). A yeast-displayed peptide-HLA complex library with the selection against a given TCR has also been used to identify peptides that bind to the TCR (Adams, Narayanan et al. 2011). These elegant approaches use a large library containing random peptides and allow the discovery of previously unknown antigens. However, the expressing of recombinant refolded TCRs and establishment of T cell clones from tissue are technically complex and time-consuming.

Approaches to directly identify the specificity of T cells without the need to clone and express a recombinant TCR, have been difficult to establish because of several challenges.

Firstly, T cells are known to be cross-reactive and can therefore detect many variants of an index peptide. Furthermore T cells can express pattern recognition receptors which can introduce non-TCR specific interactions if the presenting structure is too complex. In addition, each member of the HLA-peptide library is likely to require multiplexing in order to form a stable association with a T cell because of low affinity interactions between TCR and HLA. Conventional peptide libraries produced by F-moc chemistry are prohibitively expensive to cover the number of peptides required for a comprehensive non-biased library. Lastly, interactions between CD8/CD4 and HLA class I/II may increase the non-specific association of particular HLA-peptide combinations through CD8 or CD4 instead of specific interactions through the TCR. We aimed to develop a system that would solve all of these problems.

We reasoned that it would be possible to use a bead based approach to probe T cell populations for antigen specificity, as beads are known to have the capacity to associate with T cells through interactions between multiplexed specific molecules on the beads. Such approaches are very difficult to develop using phage/yeast based systems because of multiple non-specific interactions that occur between the cells and the phage/yeast. We further reasoned that instead of using F-moc chemistry to generate the peptide library, it would be possible to synthesize a library of HLA-coated beads using the ORBIT emulsion based *in vitro* transcription/translation expression system where the production of each bead is separated from its neighbour in individual water-in-oil microcompartments (Diehl, Li et al. 2006). Individual beads would carry multiple copies of their own unique peptide/HLA combination and the DNA encoding the peptide. In this way the DNA would act both as the synthetic template for the peptide, and also as the barcode readout to identify the peptide sequence after selection. By identifying peptides bound to beads using high-throughput sequencing, it would be possible to rank peptide candidates between conditions which may

provide information about critical non-cross-reactive residues within the identified peptides. Lastly we reasoned that using CD8 blocking antibodies or CD8-null HLA variants, we would be able to circumvent the non-specific background interactions between CD8 and HLA class I.

In this study, we used emulsion PCR and emulsion *in vitro* transcription translation to express β-2-microglobulin (β2m) covalently linked to a random nine amino acid peptide library on beads alongside the DNA encoding the β2m and peptide. HLA heavy chain with or without the ability to bind to CD8 was refolded with peptide- β2m on the surface of the beads. The HLA-peptides library was then selected against HLA-A*0201-restricted T cells with specificity for Influenza matrix (58-66): GILGFVFTL, CMV pp65 (495-503): NLVPMVATV, EBV BMLF-1 (280-288): GLCTLVAML peptides. High-throughput sequencing was used to assess the library and identify the sequence of selected peptides. Peptide sequence analysis was used to rank and characterize the sequences of identified peptides.

5.2 Results

5.2.1 HLA-peptide Library

The construction of the HLA-peptide Library is shown in **Figure 5.1**. The emulsion PCR and emulsion *in vitro* transcription/translation processes were described in Chapter 3.

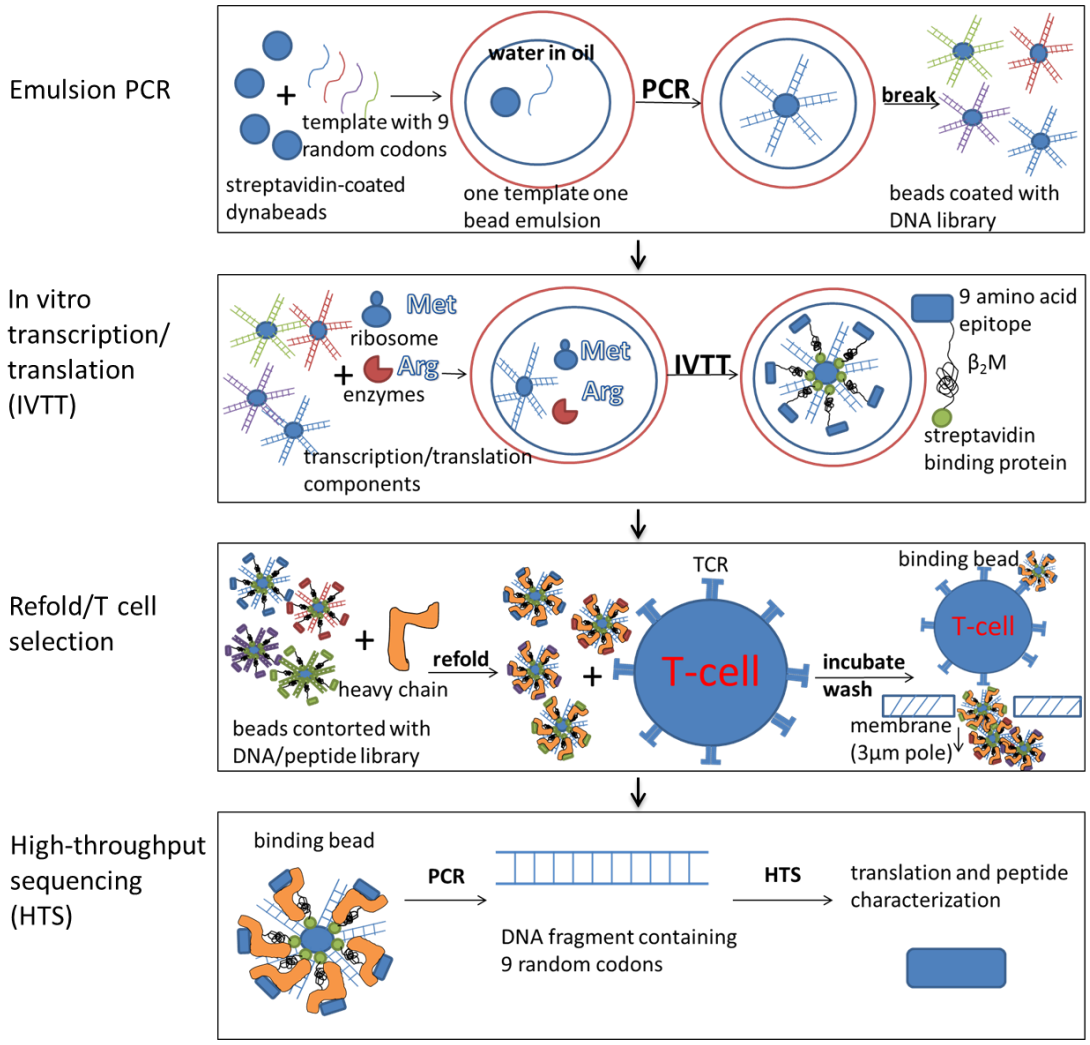
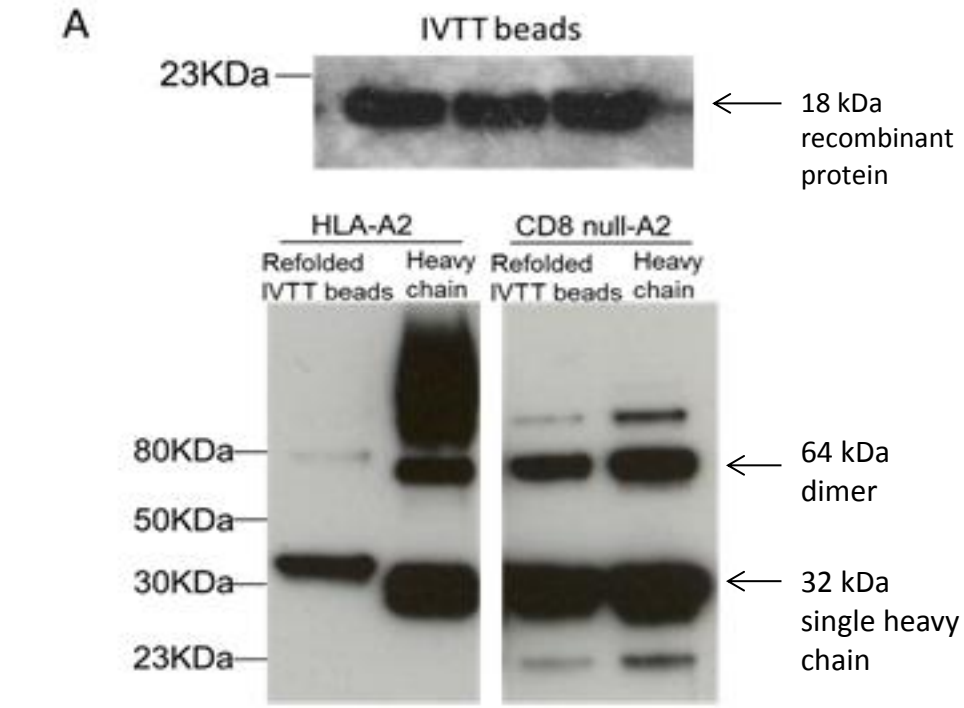


Figure 5.1 Overview of the construction of bead library and T cell epitope selection. Emulsion PCR: a water-in-oil emulsion is formed for each bead, with each emulsion containing a single template with PCR components, generating bead-DNA complexes after PCR. IVTT: Each bead-DNA complex forms a new emulsion with *in vitro* transcription/translation components for protein synthesis. Since the protein translated contains streptavidin-binding-protein, it will attach to the bead as soon as it is translated, resulting the formation of a protein-DNA-bead complex. Refold/ T cell selection: The protein-DNA-bead complexes were refolded with heavy chain to form the HLA-peptide-bead library. The library was incubated with T cell clones and high-throughput sequencing was applied to identify and characterise these peptide sequences.

Following Western blot identification of β2M and SBP on the surface of the beads, HLA heavy chains (HLA-A*0201 or CD8 null A*0201 using domains 1 and 2 of HLA-A*0201 and domain 3 of Kb with mutated residues 226A/227L courtesy of Prof Vincenzo Cerundolo, which can specifically bind to high affinity HLA-A2 restricted T cells (Choi, Palmowski et al. 2002) were added to the bead library under suitable refolding conditions. The quality of the expressed protein or the heavy chains on the beads was assessed by Western blot with streptavidin/anti-streptavidin antibody and HC-10 antibody (specific for free heavy chains) as shown in **Figure 5.2a**. The refolded HLA-peptide complex on beads was assessed with antibody W6-32 staining (Shields and Ribaudo 1998) (specific for fully assembled conformationally correct HLA class I molecules, **Figure 5.2b**).



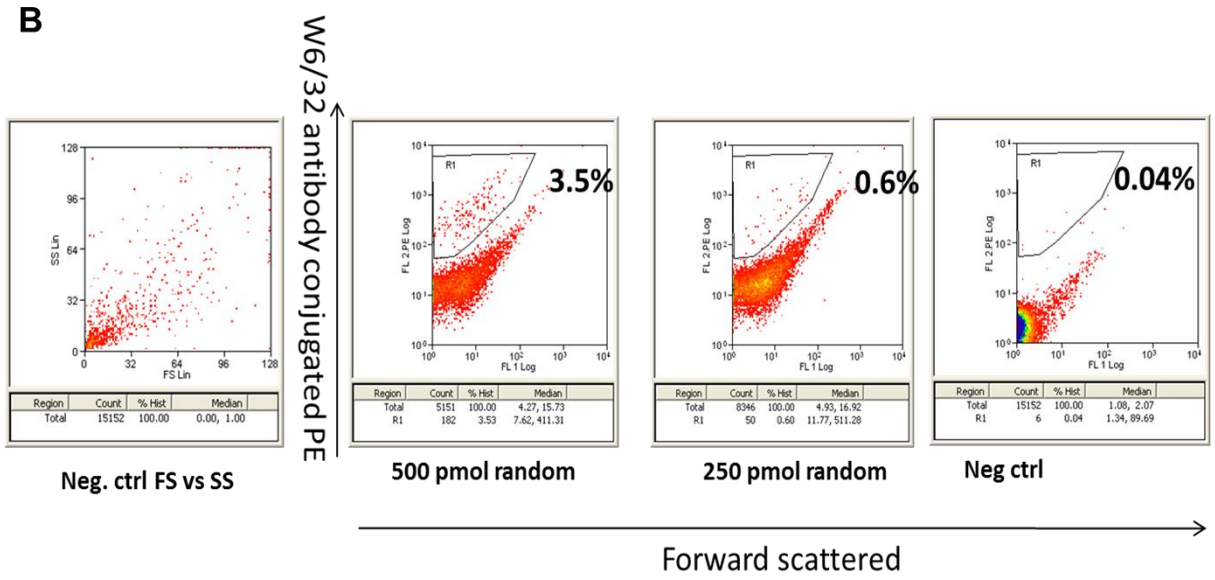


Figure 5.2a The western blot of IVTT beads (the protein-DNA-bead complexes) with the bands showing streptavidin-binding-protein staining from three independent preparations and the western blot of refolded IVTT beads (the HLA-peptide library) with the bands showing the stain of free heavy chain (left: HLA-A2, right: HLA-CD8 null-A2), the recombinant heavy chain was used as a positive control.

Figure 5.2b FACS staining of refolded IVTT beads. The refolded beads containing HLA-peptide library were stained with W6/32 antibody conjugated with PE.

5.2.2 Enrichment in the sequences containing HLA-A2 anchors

A library containing all possible variants of a nine amino acids would comprise 10^{12} sequences. As we used a library containing 10^7 beads, we developed Perl scripts (**Appendix**) to identify HLA-A*0201-binding motifs and shorter (5mer) sequences as such variants would be predicted to be present within a 10^7 library. Of note, however, the methodology can be up-scaled to larger numbers of beads. HLA-A*0201-restricted T cell clones were generated by Dr Tao Dong's team to the peptides GILGFVFTL (influenza A matrix), NLVPMVATV (CMV pp65) and GLCTLVAML (EBV BMLF1) and shown to be specific (**Figure 5.3**). The T cells were incubated directly with the HLA-A*0201 beads or after anti-CD8 antibody treatment. The HLA-A*0201 beads were generated using either wild-type HLA-A*0201 or the CD8 null version. Unbound beads were separated from those associated with the T cells using centrifugation through a 3 μm filter. The DNA on the remaining beads DNA was isolated and sequencing using high-throughput methodology. The unique sequences accounted for 98% ± 0.2% (mean ± STD) of the total and showed a good diversity of the peptide library. After the selection, there is no obvious decrease in the number of total peptide sequences and the unique sequences account for 95% ± 1.4% of the total peptides ($P=0.0002$, Student's t test, pre vs post), indicating there is enrichment after the selection. The percentage of unique sequence decreased from 96% ± 0.5% to 94% ± 1.1% ($P<0.001$) when using anti-CD8+ indicating that CD8 blockade can improve the selection.

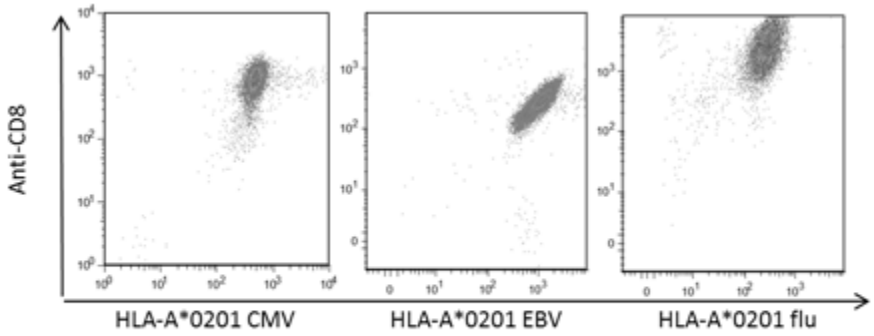


Figure 5.3 HLA-tetrameric complex staining of the HLA-A*0201-restricted T cell clones specific to Influenza matrix (58-66): GILGFVFTL, CMV pp65 (495-503): NLVPMVATV, EBV BMLF-1 (280-288): GLCTLVAML peptides. The staining was performed courtesy of Dr Peng from Dr Tao group.

The sequences containing HLA-A2 anchors were selected based on a pattern search using: XLXXXXXXV for CMV peptide, XLXXXXXXL for EBV peptide and XIXXXXXXL for Flu peptide. The highest frequency and the number of anchor-containing sequences with the frequency ≥2 were normalised to the number of total sequences in each sample and shown in **Figure 5.4**. There is an increase in the number of the anchor-containing sequences in post selection samples suggesting that bead selection has occurred. The CD8-null HLA-A*0201 show a significant enhancement of relevant selection of anchor residues. The prior addition of anti-CD8 antibody further enhanced selection of peptide sequences with correct anchor positions. The frequency of peptides showing I2V6L9, L2V6L9, and L2V6V9 were also enriched in the relevant T cell selected populations (**Table 5.1**). These data show that primary anchor residues at positions 2 and 9 can be selected for HLA-A*0201 but in addition the secondary anchor of position 6 is also important.

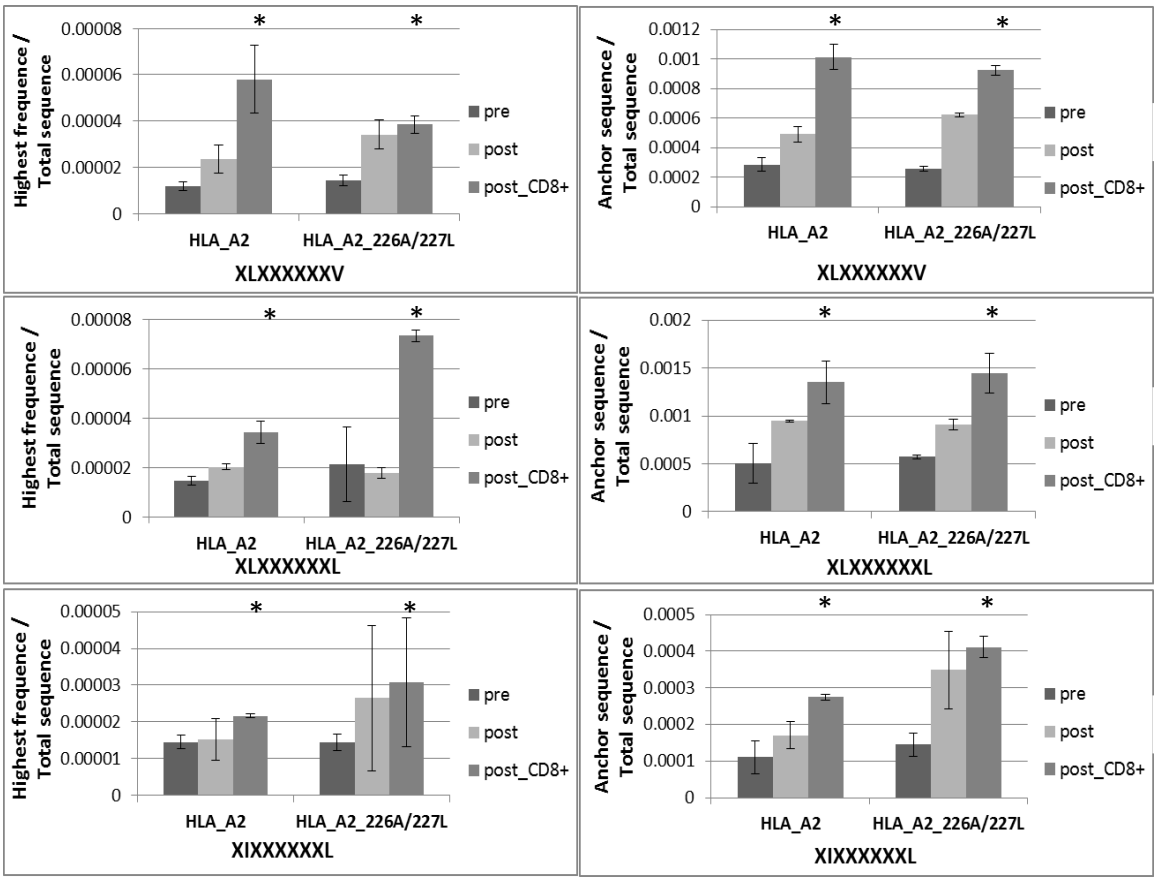


Figure 5.4 Ratio of the number of the sequence containing HLA-A2 anchors to the number of the total sequence from pre-selection, post-selection and anti-CD8 antibody treated post-selection samples. XLXXXXXXV pattern was used to search in samples challenged with CMV viral peptide, XLXXXXXXL for EBV peptide and XIXXXXXXL for flu peptide. **p-values* <0.05 (comparing between unselected samples (pre) against post-selection CD8 treated samples.

specific T cell selection	Search pattern	Freq	Fit to peptide	Heavy chain
pre-selection	XIXXXVXXL	2	N/A	A2
	XLXXXVXXL	2	N/A	A2
	XLXXXVXXV	2	N/A	A2
	XIXXXVXXL	2	N/A	226
	XLXXXVXXL	2	N/A	226
	XLXXXVXXV	1	N/A	226
Flu:GILGFVFTL	XIXXXVXXL	5	33%	226+
EBV:GLCTLVAML	XLXXXVXXL	7	56%	A2+
CMV:NLVPMVATV	XLXXXVXXV	6	33%	226

Table 5.1 HLA-A*0201 anchor analysis with three search patterns (I2V6L9, L2V6L9, L2V6V9). Freq is the highest frequency of the peptides containing the searched pattern. Fit to peptide represents the percentage of the same amino acids in the same position as the viral peptides.

sample	total	unique	unique%	highest freq
preA2	151719	148869	98%	7
preA2	189987	185862	98%	9
pre226	156382	153028	98%	9
pre226	186887	182545	98%	9
NLVPMVATV_A2	154052	147554	96%	13
NLVPMVATV_A2	178529	170381	95%	15
GLCTLVAML_A2	155589	149884	96%	19
GLCTLVAML_A2	190255	182488	96%	29
GILGFVFTL_A2	157144	152025	97%	12
GILGFVFTL_A2	179532	173226	96%	15
NLVPMVATV_226	133990	128619	96%	10
NLVPMVATV_226	155186	148477	96%	14
GLCTLVAML_226	156604	150413	96%	9
GLCTLVAML_226	183870	175731	96%	12
GILGFVFTL_226	160507	153309	96%	20
GILGFVFTL_226	198138	187862	95%	20
NLVPMVATV_A2_CD8	160836	150735	94%	21
NLVPMVATV_A2_CD8	188810	175804	93%	33
GLCTLVAML_A2_CD8	187165	173860	93%	32
GLCTLVAML_A2_CD8	193597	179880	93%	27
GILGFVFTL_A2_CD8	141292	135600	96%	11
GILGFVFTL_A2_CD8	181569	173144	95%	13
NLVPMVATV_226_CD8	170013	158438	93%	22
NLVPMVATV_226_CD8	195617	181508	93%	23
GLCTLVAML_226_CD8	167612	155913	93%	32
GLCTLVAML_226_CD8	200255	185273	93%	19
GILGFVFTL_226_CD8	163436	154340	94%	33
GILGFVFTL_226_CD8	185266	174250	94%	32

Table 5.2 Summary of the number of peptides in the pre and post selection samples

5.2.3 Enrichment of the peptide fragments after T cell selection

Perl scripts were developed to identify sequential 5mer peptides within each of the peptides GILGFVFTL, NLVPMVATV and GLCTLVAML. Both forward and reverse sequences were used as the peptides coupled to β2m could associate with the HLA heavy chain in either direction. The frequencies were further normalised to the total sequence number for each sample. The mean frequency from at least 4 independent samples of each fragment is listed in **Table 5.3**. None of the 5mers were reduced in both the forward and reverse directions, but several directly matching 5mers were enriched in both forward and reverse directions in the post-selection samples. GILGF, LVAML and CTLVA aligned to the relevant T cells and were not enriched in the T cells specific to the other viruses. There were four viral fragments (VLTCL, ALVTC, LMAVL for EBV peptide, FGLIG for flu peptide) showing an increase over 80% and these 5mers attained the top E value in blast searches for the relevant virus. Enrichment in three further viral fragments (AVPMV for CMV virus and GIGLF, FVFTL for flu virus) was observed. The data confirm that 5mers matching the specificity of the responding T cells can be identified using relative rank enrichment from a 10^7 library.

Flu Matrix	Fragments	pre	post	EBV BMLF-1	Fragments	pre	post	CMV pp65	Fragments	pre	post
	GILGF*	0	8		GLCTL	12	10		NLVPM	2	2
	*FGLIG**	16	28		*LTCLG*	52	57		*MPVLN*	10	3
	ILGFV	6	2		LCTLV	10	10		LVPMV	12	13
	VFGLI	4	4		*VLTCL*	6	15		*VMPVL*	12	13
GILGFVFTL	LGFVF	8	0	GLCTLVAML	CTLVA*	6	9	NLVPMVATV	VPMVA	8	2
	FVFGL	4	2		*AVLTC**	4	9		*AVMPV*	0	8
	GFVFT	6	3		TLVAM	14	9		PMVAT	2	2
	TFVFG	8	13		*MAVLT*	10	9		*TAVMP*	2	1
	FVFTL	0	3		LVAML*	6	11		MVATV	4	4
	LTFVF	4	3		*LMAVL**	2	11		*VTAVM*	8	5

Table 5.3 viral peptide fragments analysis. Each viral peptide was broken into 5mers. Pre and post represent the frequency of the fragments appearing in pre-selection and post-selection samples. *These fragments had increased frequencies when selecting both forward (5'-3') and reverse (3'-5')

5.3 Discussion

We have developed a novel HLA-peptide bead display library using emulsion PCR and *in vitro* transcription translation which can be used to probe T cells in order to determine antigen specificity. The TCR does not have to be expressed and refolded, but the T cells themselves can be probed directly with the beads which contain HLA-β2m-peptide and the DNA encoding the construct.

The sequences of HLA-A*0201 binding peptides have been defined (Falk, Rotzschke et al. 1991; Ruppert, Sidney et al. 1993; Parker, Bednarek et al. 1994) with primary anchor residues typically at peptide position 2 and 9 (Fruci, Rovero et al. 1993; Falk, Rotzschke et al. 1994). Therefore, the peptides containing patterns of I2L9, L2V9 and L2L9 based on the relevant viral peptides (GILGFVFTL, NLVPMVATV and GLCTLVAML) were chosen to investigate. The enrichment in A2-anchor containing peptides in all three patterns after selection shows that it is possible to define the pattern of anchor residues. The further improvement in anti-CD8 antibody treated samples indicates that anti-CD8 can be an important factor to enhance specific selection.

The similarity of the selected peptides to the known specificity of the T cells was studied to investigate the ability of the selection system. The variation range of non-viral fragments was 20%-30%. There were four viral fragments (VLTCL, ALVTC, LMAVL for EBV peptide, FGLIG for flu peptide) showing an increase over 80%. Enrichment in three viral fragments (AVPMV for CMV virus and GIGLF, FVFTL for flu virus) were present as well. The enrichment of these specific viral fragments indicates that the identity of unknown antigens can be defined using the HLA bead selection system.

The bead based HLA expression approach solves a number of problems that have hindered the development of HLA-peptide libraries. The problem of T cell cross-reactivity can be

addressed using high-throughput sequence analysis and relative ranking to identify relevant peptide sequences. Using a bead based system allows HLA-peptide molecules to be multiplexed on the bead surface to increase HLA-TCR binding avidity. Through the use of *in vitro* transcription/translation, the bead system avoids expensive requirement to synthesize all peptides within the library using F-moc chemistry. Furthermore, by just having the DNA and HLA molecules on the surface of the beads, interactions with other molecules that might be present on other forms of particle (eg phage or yeast) are not an issue. Non-specific interactions between CD8 and HLA class I can be avoided through the use of anti-CD8 blocking prior to bead binding and selection. Lastly, although the current study has used a random 9 amino acid library it would be straightforward to use DNA sourced from particular targets such as viral, bacterial or tumour cells to target potential T cell epitopes within a known antigen source of interest.

Furthermore, the beads overcome previous challenges to identify specific peptide fragments that are the targets of T cells and avoid the need to express the TCR in recombinant refolded form. Instead the T cells can be directly probed with the bead HLA-peptide library and then identified using high-throughput sequence enrichment of post-selection sequences. Future developments may include the generation of libraries containing restricted residues at HLA anchor positions in order to increase the possible variations at the other residues. It is likely that the technology will have many potential applications for defining the specificity of T cells in T cell associated disease.

Chapter 6 - General Discussion

6.1 OrBIT as a promising alternative technique for ligand selection

We have demonstrated multiple approaches of the ORBIT bead library system. By implementing human β-2m into the construct, the random peptides we synthesized are capable of acting as a peptide library alone, or as a refolded MHC class I antigen library for specific HLA molecules. By creating an artificial compartment, an isolated, clean water-in-oil environment, protein transcription and translation is possible. Our current standard operating procedure produces 10^6-10^7 such isolated micro-environments in 1 ml emulsion mixtures, which are on par with the conventional ligand libraries that utilize in vivo synthesis. It is possible to achieve larger bead library with up to 10^{12} variations by increasing the working volume. However careful optimization of the quality of the water-in-oil emulsion mixture is necessary to ensure the production of homogeneous beads.

We used HIV-1 coat protein gp120 as the target protein and had managed to select and characterize a number of high affinity peptides which not only bound to gp120, but also reduced the rate of HIV-1 replication inhibition activity by two different HIV-1 clades. There was also evidence to suggest the actual gp120 binding site of one of the peptide (peptide GP3 to gp120 V3 loop domain). The sequence of peptides we obtained could potentially serve as a fundamental platform for the design of small drugs or other peptide mimetics that are more adapted for therapeutic uses. Interestingly, to date we only managed to find peptides with 9 random amino acids that could bind to gp120. Despite the fact that we had synthesized a number of random peptides with cysteine residues which mimicked the structure of defensin, we were not able to notice any significant binding. This could be due to the fact that we did not add post-translational modification step to our production line, despite that cysteines should form spontaneous disulphide bridge. It is possible that disulphide crosslinks occur between peptides, therefore disrupting the peptide post-translational structures. However we are now using a readily available IVTT system in which intrachain disulphide bonds are

possible. Careful optimization is required if we decided to use the peptide library in a specific structure for selection.

Furthermore, we showed that the unique construct peptide-β2m-SBP was able to refold into an antigen-presenting MHC class I molecule with two different recombinant HLA molecules (HLA-0201, HLA-B8). The MHC-bead complexes could be used to probe antigen-specific T cell clones and isolate high affinity peptides recognized by T cell receptors. The peptide sequences we analysed suggested that peptides with the correct anchor residue patterns were enriched for HLA-A*0201, and blocking the CD8 co-receptor molecule on T cells increased the specificity of the selection.

6.2 Technical difficulties encountered by emulsion technology

The water-oil condition can prove to be the greatest obstacle for the emulsion technology. Oil surfactants are required to create isolated aqueous compartments, and we were able to synthesize proteins that have simple constructs. However if complex proteins which required specific temperature, enzymes, or pH are to be synthesized, the integrity of the emulsion needs to be evaluated with caution. It has been found that the hydrophobic phase of the surfactants could intervene with complex biochemical processes, and potentially part because the interface can denature newly synthesised proteins (Miller, Bernath et al. 2006). Therefore it is important to test different oil compositions with different aqueous reagents before proper selection. In most cases, surfactant ABIL EM90 proved to be reliable surfactant for all aqueous *in vitro* translation systems, regardless of prokaryotic or eukaryotic nature (Ghadessy and Holliger 2004; Williams, Peisajovich et al. 2006).

Furthermore, our ORBIT system incorporates E. coli lysates for protein synthesis, which limits the system to produce only primary structure proteins. The use of rabbit reticulocyte

lysate can overcome this restriction but at higher cost and lower protein yields. Implementing chaperone proteins such as GroEL and GroES to the E. coli system could solve this problem but whether the added complexity would affect the quality of the library remain to be elucidated.

Thirdly, a good gene expression system should contain means to inhibit nuclease and protease activity. It was reported that the use of biotinylation or phosphorothiolated nucleotides could effectively reduce the DNA degradation by nucleases (Takei, Kadomatsu et al. 2002). Our system uses biotinylated primers and relies on working efficiently on ice to minimize the nuclease activity; however a certain level of template degradation was still inevitable as observed by the presence of fragmented DNA sequences in the sequencing results. This could potentially result in the loss of valuable sequence information. There were also commercially available PURE *in vitro* translation systems (Shimizu, Kanamori et al. 2005) which claimed to have excluded all nucleases in the lysates. Also, the addition of large proteins such as BSA or sodium deoxycholate (a negatively charged surfactant) have been shown to reduce the chance of trapping newly translated proteins by emulsion oil phase (Tawfik and Griffiths 1998; Ghadessy, Ramsay et al. 2004).

Finally, the consistent production of standardized size water-in-oil emulsions need to be mechanized if large size library of up to 10^{12} is to be made in the future. Our current protocol utilizes a magnetic stir bar in a miniaturized volume, and the use of homogenizer and microcapillary had also been reported (Agresti, Kelly et al. 2005; Utada, Lorenceau et al. 2005). While at small volumes these methods produce good quality emulsions, it had been reported that high speed stirring might cause reduction in enzyme activity (Miller, Bernath et al. 2006). With the ever increasing in demands for emulsion PCR technology, bioengineers have developed instruments capable of synthesizing consistent size emulsion at a stable rate of 10^7 emulsions per hour. In addition, it is also capable of merging the two emulsions

together, creating many future possibilities to utilize emulsion technology and the high throughput selection workflow (Kintses, van Vliet et al. 2010).

6.3 Future perspective of ORBIT

Currently we are actively designing new DNA constructs such as single-chain variable fragment (scFv), which was already made possible by phage display (Liu, Hu et al. 2012). Our preliminary lab data had suggested that we were able to produce random scFv library linking with beads at substantial level. The use of scFv constructs would enable us to select and identify antibody fragment sequences without using animal models, which greatly increase the cost and efficiency required for future monoclonal antibody development. In addition to the examples we showed, we had also identified early evidence of peptide ligands that could bind to influenza hemagglutinin coat protein (data not shown); and peptides ligands for Rat sarcoma GTPase (Ras-GTP) tumour-associated protein. Functional experiments of these peptide ligands for both projects are on-going. With the abundance of strong collaborations from labs with extensive knowledge and the ability of synthesizing purified proteins they are interested in, it is possible to obtain peptide ligands binding to the target protein of interest within a week.

Therapeutically, we are also investigating the use of unnatural amino acids for the synthesis of the peptides, which could prove to be important in the potential drug design for example in modifying peptide stability.

In addition we are using a novel IVTT systems that allow the formation of intramolecular disulphide bonds. This theoretically would allow the formation of cyclic peptide structures with constricted secondary structure and potentially greater affinity and other associated characteristics. Overall these further refinements are actively underway and will be used to explore the potential of ORBIT bead display for diagnostic and therapeutic development.

References

Adams, Jarrett J., S. Narayanan, et al. (2011). "T Cell Receptor Signaling Is Limited by Docking Geometry to Peptide-Major Histocompatibility Complex." Immunity **35**(5): 681-693.

Agresti, J. J., B. T. Kelly, et al. (2005). "Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization." Proc Natl Acad Sci U S A **102**(45): 16170-16175.

Aharoni, A., A. D. Griffiths, et al. (2005). "High-throughput screens and selections of enzyme-encoding genes." Curr Opin Chem Biol **9**(2): 210-216.

Alam, S. M., M. Morelli, et al. (2009). "Role of HIV membrane in neutralization by two broadly neutralizing antibodies." Proc Natl Acad Sci U S A **106**(48): 20234-20239.

Andrus, L., A. M. Prince, et al. (1998). "Passive immunization with a human immunodeficiency virus type 1-neutralizing monoclonal antibody in Hu-PBL-SCID mice: isolation of a neutralization escape variant." J Infect Dis **177**(4): 889-897.

Antosova, Z., M. Mackova, et al. (2009). "Therapeutic application of peptides and proteins: parenteral forever?" Trends Biotechnol **27**(11): 628-635.

Appay, V., D. C. Douek, et al. (2008). "CD8+ T cell efficacy in vaccination and disease." Nat Med **14**(6): 623-628.

Baggio, R., P. Burgstaller, et al. (2002). "Identification of epitope-like consensus motifs using mRNA display." J Mol Recognit **15**(3): 126-134.

Barbato, G., E. Bianchi, et al. (2003). "Structural analysis of the epitope of the anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion." J Mol Biol **330**(5): 1101-1115.

Barrick, J. E. and R. W. Roberts (2003). "Achieving specificity in selected and wild-type N peptide-RNA complexes: the importance of discrimination against noncognate RNA targets." Biochemistry **42**(44): 12998-13007.

Bell, C. H., R. Pantophlet, et al. (2008). "Structure of antibody F425-B4e8 in complex with a V3 peptide reveals a new binding mode for HIV-1 neutralization." J Mol Biol **375**(4): 969-978.

Binkowski, B. F., R. A. Miller, et al. (2005). "Ligand-regulated peptides: a general approach for modulating protein-peptide interactions with small molecules." Chem Biol **12**(7): 847-855.

Binley, J. M., T. Wrin, et al. (2004). "Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies." J Virol **78**(23): 13232-13252.

Binz, H. K., P. Amstutz, et al. (2004). "High-affinity binders selected from designed ankyrin repeat protein libraries." Nat Biotechnol **22**(5): 575-582.

Binz, H. K., P. Amstutz, et al. (2005). "Engineering novel binding proteins from nonimmunoglobulin domains." Nat Biotechnol **23**(10): 1257-1268.

Bray, B. L. (2003). "Large-scale manufacture of peptide therapeutics by chemical synthesis." Nat Rev Drug Discov **2**(7): 587-593.

Burton, D. R., J. Pyati, et al. (1994). "Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody." Science **266**(5187): 1024-1027.

Calarese, D. A., C. N. Scanlan, et al. (2003). "Antibody domain exchange is an immunological solution to carbohydrate cluster recognition." Science **300**(5628): 2065-2071.

Callan, M. F., L. Tan, et al. (1998). "Direct visualization of antigen-specific CD8+ T cells during the primary immune response to Epstein-Barr virus In vivo." J Exp Med **187**(9): 1395-1402.

Chaillon, A., M. Braibant, et al. (2011). "The V1V2 domain and an N-linked glycosylation site in the V3 loop of the HIV-1 envelope glycoprotein modulate neutralization sensitivity to the human broadly neutralizing antibody 2G12." J Virol **85**(7): 3642-3648.

Chandrudu, S., P. Simerska, et al. (2013). "Chemical methods for peptide and protein production." Molecules **18**(4): 4373-4388.

Choi, E. M., J. L. Chen, et al. (2003). "High avidity antigen-specific CTL identified by CD8-independent tetramer staining." J Immunol **171**(10): 5116-5123.
Choi, E. M., M. Palmowski, et al. (2002). "The use of chimeric A2K(b) tetramers to monitor HLA A2 immune responses in HLA A2 transgenic mice." J Immunol Methods **268**(1): 35-41.
Craik, D. J., N. L. Daly, et al. (1999). "Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif." J Mol Biol **294**(5): 1327-1336.
Diehl, F., M. Li, et al. (2006). "BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions." Nat Methods **3**(7): 551-559.
Dogo-Isonagie, C., S. Lam, et al. (2012). "Peptides from second extracellular loop of C-C chemokine receptor type 5 (CCR5) inhibit diverse strains of HIV-1." J Biol Chem **287**(18): 15076-15086.
Dressman, D., H. Yan, et al. (2003). "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations." Proc Natl Acad Sci U S A **100**(15): 8817-8822.
Durda, P. J., B. Leece, et al. (1988). "Characterization of murine monoclonal antibodies to HIV-1 induced by synthetic peptides." AIDS Res Hum Retroviruses **4**(5): 331-342.
Eldridge, B., R. N. Cooley, et al. (2009). "An in vitro selection strategy for conferring protease resistance to ligand binding peptides." Protein Eng Des Sel **22**(11): 691-698.
Falk, K., O. Rotzschke, et al. (1994). "Analysis of a naturally occurring HLA class I-restricted viral epitope." Immunology **82**(3): 337-342.
Falk, K., O. Rotzschke, et al. (1991). "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules." Nature **351**(6324): 290-296.
Fernandez-Gacio, A., M. Uguen, et al. (2003). "Phage display as a tool for the directed evolution of enzymes." Trends Biotechnol **21**(9): 408-414.
Flynn, N. M., D. N. Forthal, et al. (2005). "Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection." J Infect Dis **191**(5): 654-665.
Fruci, D., P. Rovero, et al. (1993). "Anchor residue motifs of HLA class-I-binding peptides analyzed by the direct binding of synthetic peptides to HLA class I alpha chains." Hum Immunol **38**(3): 187-192.
Gai, S. A. and K. D. Wittrup (2007). "Yeast surface display for protein engineering and characterization." Curr Opin Struct Biol **17**(4): 467-473.
Gan, R., S. Furuzawa, et al. (2010). "Directed evolution of angiotensin II-inhibiting peptides using a microbead display." J Biosci Bioeng **109**(4): 411-417.
Gan, R., Y. Yamanaka, et al. (2008). "Microbeads display of proteins using emulsion PCR and cell-free protein synthesis." Biotechnol Prog **24**(5): 1107-1114.
Ghadessy, F. J. and P. Holliger (2004). "A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system." Protein Eng Des Sel **17**(3): 201-204.
Ghadessy, F. J., N. Ramsay, et al. (2004). "Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution." Nat Biotechnol **22**(6): 755-759.
Gilbert, P., M. Wang, et al. (2010). "Magnitude and breadth of a nonprotective neutralizing antibody response in an efficacy trial of a candidate HIV-1 gp120 vaccine." J Infect Dis **202**(4): 595-605.
Goverman, J. (2009). "Autoimmune T cell responses in the central nervous system." Nat Rev Immunol **9**(6): 393-407.
Gray, E. S., M. C. Madiga, et al. (2011). "The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection." J Virol **85**(10): 4828-4840.
Griffiths, A. D. and D. S. Tawfik (2006). "Miniaturising the laboratory in emulsion droplets." Trends Biotechnol **24**(9): 395-402.
Guermonprez, P., J. Valladeau, et al. (2002). "Antigen presentation and T cell stimulation by dendritic cells." Annu Rev Immunol **20**: 621-667.

Hanahan, D. and Robert A. Weinberg (2011). "Hallmarks of Cancer: The Next Generation." Cell **144**(5): 646-674.
Hanes, J., L. Jermutus, et al. (1999). "Comparison of Escherichia coli and rabbit reticulocyte ribosome display systems." FEBS Lett **450**(1-2): 105-110.
Hanes, J., C. Schaffitzel, et al. (2000). "Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display." Nat Biotechnol **18**(12): 1287-1292.
He, M. and M. J. Taussig (2005). "Ribosome display of antibodies: expression, specificity and recovery in a eukaryotic system." J Immunol Methods **297**(1-2): 73-82.
Hessell, A. J., E. G. Rakasz, et al. (2009). "Broadly neutralizing human anti-HIV antibody 2G12 is effective in protection against mucosal SHIV challenge even at low serum neutralizing titers." PLoS Pathog **5**(5): e1000433.
Hoet, R. M., E. H. Cohen, et al. (2005). "Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity." Nat Biotechnol **23**(3): 344-348.
Holliger, P. and P. J. Hudson (2005). "Engineered antibody fragments and the rise of single domains." Nat Biotechnol **23**(9): 1126-1136.
Holmberg, A., A. Blomstergren, et al. (2005). "The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures." Electrophoresis **26**(3): 501-510.
Hoogenboom, H. R. (2005). "Selecting and screening recombinant antibody libraries." Nat Biotechnol **23**(9): 1105-1116.
Horisawa, K., S. Tateyama, et al. (2004). "In vitro selection of Jun-associated proteins using mRNA display." Nucleic Acids Res **32**(21): e169.
Jenssen, H., P. Hamill, et al. (2006). "Peptide antimicrobial agents." Clin Microbiol Rev **19**(3): 491-511.
Jose, J. (2006). "Autodisplay: efficient bacterial surface display of recombinant proteins." Appl Microbiol Biotechnol **69**(6): 607-614.
Keefe, A. D. and J. W. Szostak (2001). "Functional proteins from a random-sequence library." Nature **410**(6829): 715-718.
Kenter, G. G., M. J. Welters, et al. (2009). "Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia." N Engl J Med **361**(19): 1838-1847.
Kintses, B., L. D. van Vliet, et al. (2010). "Microfluidic droplets: new integrated workflows for biological experiments." Curr Opin Chem Biol **14**(5): 548-555.
Kitagawa, M., T. Ara, et al. (2005). "Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research." DNA Res **12**(5): 291-299.
Koga, Y., T. Yamane, et al. (2007). "Creation of novel enantioselective lipases by SIMPLEX." Methods Mol Biol **375**: 165-181.
Kojima, T., Y. Takei, et al. (2005). "PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets." Nucleic Acids Res **33**(17): e150.
Labrijn, A. F., P. Poignard, et al. (2003). "Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1." J Virol **77**(19): 10557-10565.
Lee, L. Y.-H., D. L. A. Ha, et al. (2008). "Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals." The Journal of Clinical Investigation **118**(10): 3478-3490.
Leffers, N., A. J. Lambeck, et al. (2009). "Immunization with a P53 synthetic long peptide vaccine induces P53-specific immune responses in ovarian cancer patients, a phase II trial." Int J Cancer **125**(9): 2104-2113.
Lipovsek, D. and A. Pluckthun (2004). "In-vitro protein evolution by ribosome display and mRNA display." J Immunol Methods **290**(1-2): 51-67.

Liu, J. L., Z. Q. Hu, et al. (2012). "Attainment of 15-fold higher affinity of a Fusarium-specific single-chain antibody by directed molecular evolution coupled to phage display." Mol Biotechnol **52**(2): 111-122.

Loffet, A. (2002). "Peptides as drugs: is there a market?" J Pept Sci **8**(1): 1-7.

Lu, S. M. and R. S. Hodges (2002). "A de novo designed template for generating conformation-specific antibodies that recognize alpha-helices in proteins." J Biol Chem **277**(26): 23515-23524.

Mascola, J. R., M. G. Lewis, et al. (1999). "Protection of Macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies." J Virol **73**(5): 4009-4018.

Matsui, K., J. J. Boniface, et al. (1991). "Low affinity interaction of peptide-MHC complexes with T cell receptors." Science **254**(5039): 1788-1791.

Mattheakis, L. C., R. R. Bhatt, et al. (1994). "An in vitro polysome display system for identifying ligands from very large peptide libraries." Proc Natl Acad Sci U S A **91**(19): 9022-9026.

McCafferty, J., A. D. Griffiths, et al. (1990). "Phage antibodies: filamentous phage displaying antibody variable domains." Nature **348**(6301): 552-554.

McCoy, L. E. and R. A. Weiss (2013). "Neutralizing antibodies to HIV-1 induced by immunization." J Exp Med **210**(2): 209-223.

McElrath, M. J. and B. F. Haynes (2010). "Induction of Immunity to Human Immunodeficiency Virus Type-1 by Vaccination." Immunity **33**(4): 542-554.

McPherson, M., Y. Yang, et al. (2002). "Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries." Chem Biol **9**(6): 691-698.

Mehandru, S., T. Wrin, et al. (2004). "Neutralization profiles of newly transmitted human immunodeficiency virus type 1 by monoclonal antibodies 2G12, 2F5, and 4E10." J Virol **78**(24): 14039-14042.

Melief, C. J. and S. H. van der Burg (2008). "Immunotherapy of established (pre)malignant disease by synthetic long peptide vaccines." Nat Rev Cancer **8**(5): 351-360.

Miceli, R., D. Myszka, et al. (1996). "The coiled coil stem loop miniprotein as a presentation scaffold." Drug Des Discov **13**(3-4): 95-105.

Miller, O. J., K. Bernath, et al. (2006). "Directed evolution by in vitro compartmentalization." Nat Methods **3**(7): 561-570.

Moulard, M., S. K. Phogat, et al. (2002). "Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes." Proc Natl Acad Sci U S A **99**(10): 6913-6918.

Mouquet, H., J. F. Scheid, et al. (2010). "Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation." Nature **467**(7315): 591-595.

Netzer, W. J. and F. U. Hartl (1997). "Recombination of protein domains facilitated by co-translational folding in eukaryotes." Nature **388**(6640): 343-349.

Neumann, H., K. Wang, et al. (2010). "Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome." Nature **464**(7287): 441-444.

Odegrip, R., D. Coomber, et al. (2004). "CIS display: In vitro selection of peptides from libraries of protein-DNA complexes." Proc Natl Acad Sci U S A **101**(9): 2806-2810.

Parker, K. C., M. A. Bednarek, et al. (1994). "Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains." J Immunol **152**(1): 163-175.

Parren, P. W., P. A. Marx, et al. (2001). "Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization in vitro." J Virol **75**(17): 8340-8347.

Pitisuttithum, P., P. Gilbert, et al. (2006). "Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand." J Infect Dis **194**(12): 1661-1671.

Poignard, P., R. Sabbe, et al. (1999). "Neutralizing antibodies have limited effects on the control of established HIV-1 infection in vivo." Immunity **10**(4): 431-438.
Precopio, M. L., T. R. Butterfield, et al. (2008). "Optimizing peptide matrices for identifying T-cell antigens." Cytometry Part A **73A**(11): 1071-1078.
Quinones-Mateu, M. E., M. M. Lederman, et al. (2003). "Human epithelial beta-defensins 2 and 3 inhibit HIV-1 replication." AIDS **17**(16): F39-48.
Reiersen, H., I. Lobersli, et al. (2005). "Covalent antibody display--an in vitro antibody-DNA library selection system." Nucleic Acids Res **33**(1): e10.
Roberts, R. W. and J. W. Szostak (1997). "RNA-peptide fusions for the in vitro selection of peptides and proteins." Proc Natl Acad Sci U S A **94**(23): 12297-12302.
Rosen, O., J. Chill, et al. (2005). "Induced fit in HIV-neutralizing antibody complexes: evidence for alternative conformations of the gp120 V3 loop and the molecular basis for broad neutralization." Biochemistry **44**(19): 7250-7258.
Rossi, L. M., P. Rangasamy, et al. (2008). "Research advances in the development of peptide antibiotics." J Pharm Sci **97**(3): 1060-1070.
Rotem, S. and A. Mor (2009). "Antimicrobial peptide mimics for improved therapeutic properties." Biochim Biophys Acta **1788**(8): 1582-1592.
Rothberg, J. M. and J. H. Leamon (2008). "The development and impact of 454 sequencing." Nat Biotechnol **26**(10): 1117-1124.
Rothe, A., R. J. Hosse, et al. (2006). "Ribosome display for improved biotherapeutic molecules." Expert Opin Biol Ther **6**(2): 177-187.
Rungpragayphan, S., T. Yamane, et al. (2007). "SIMPLEX: single-molecule PCR-linked in vitro expression: a novel method for high-throughput construction and screening of protein libraries." Methods Mol Biol **375**: 79-94.
Ruppert, J., J. Sidney, et al. (1993). "Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules." Cell **74**(5): 929-937.
Sato, A. K., M. Viswanathan, et al. (2006). "Therapeutic peptides: technological advances driving peptides into development." Curr Opin Biotechnol **17**(6): 638-642.
Sawata, S. Y., E. Suyama, et al. (2004). "A system based on specific protein-RNA interactions for analysis of target protein-protein interactions in vitro: successful selection of membrane-bound Bak-Bcl-xL proteins in vitro." Protein Eng Des Sel **17**(6): 501-508.
Schnur, E., E. Noah, et al. (2011). "The conformation and orientation of a 27-residue CCR5 peptide in a ternary complex with HIV-1 gp120 and a CD4-mimic peptide." J Mol Biol **410**(5): 778-797.
Seitz, S., C. K. Schneider, et al. (2006). "Reconstitution of paired T cell receptor α- and β-chains from microdissected single cells of human inflammatory tissues." Proceedings of the National Academy of Sciences **103**(32): 12057-12062.
Sheppard, N. and Q. Sattentau (2005). "The prospects for vaccines against HIV-1: more than a field of long-term nonprogression?" Expert Rev Mol Med **7**(2): 1-21.
Shibata, R., T. Igarashi, et al. (1999). "Neutralizing antibody directed against the HIV-1 envelope glycoprotein can completely block HIV-1/SIV chimeric virus infections of macaque monkeys." Nat Med **5**(2): 204-210.
Shields, M. J. and R. K. Ribaudo (1998). "Mapping of the monoclonal antibody W6/32: sensitivity to the amino terminus of beta2-microglobulin." Tissue Antigens **51**(5): 567-570.
Shimizu, Y., T. Kanamori, et al. (2005). "Protein synthesis by pure translation systems." Methods **36**(3): 299-304.
Siewert, K., J. Malotka, et al. (2012). "Unbiased identification of target antigens of CD8+ T cells with combinatorial libraries coding for short peptides." Nat Med **18**(5): 824-828.
Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**(4705): 1315-1317.
Smith, G. P. (1991). "Surface presentation of protein epitopes using bacteriophage expression systems." Curr Opin Biotechnol **2**(5): 668-673.

Stanfield, R. L., M. K. Gorny, et al. (2004). "Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52D." Structure **12**(2): 193-204.
Stevenson, C. L. (2009). "Advances in peptide pharmaceuticals." Curr Pharm Biotechnol **10**(1): 122-137.
Stiegler, G., R. Kunert, et al. (2001). "A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human immunodeficiency virus type 1." AIDS Res Hum Retroviruses **17**(18): 1757-1765.
Sumida, T., N. Doi, et al. (2009). "Bicistronic DNA display for in vitro selection of Fab fragments." Nucleic Acids Res **37**(22): e147.
Takahashi, T. T., R. J. Austin, et al. (2003). "mRNA display: ligand discovery, interaction analysis and beyond." Trends Biochem Sci **28**(3): 159-165.
Takei, Y., K. Kadomatsu, et al. (2002). "5'-,3'-inverted thymidine-modified antisense oligodeoxynucleotide targeting midkine. Its design and application for cancer therapy." J Biol Chem **277**(26): 23800-23806.
Tawfik, D. S. and A. D. Griffiths (1998). "Man-made cell-like compartments for molecular evolution." Nat Biotechnol **16**(7): 652-656.
Trkola, A., H. Kuster, et al. (2005). "Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies." Nat Med **11**(6): 615-622.
Trkola, A., M. Purtscher, et al. (1996). "Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1." J Virol **70**(2): 1100-1108.
Uger, R. A. and B. H. Barber (1998). "Creating CTL targets with epitope-linked beta 2-microglobulin constructs." J Immunol **160**(4): 1598-1605.
Ullman, C. G., L. Frigotto, et al. (2011). "In vitro methods for peptide display and their applications." Brief Funct Genomics **10**(3): 125-134.
Utada, A. S., E. Lorenceau, et al. (2005). "Monodisperse double emulsions generated from a microcapillary device." Science **308**(5721): 537-541.
Vaine, M., S. Wang, et al. (2010). "Profiles of human serum antibody responses elicited by three leading HIV vaccines focusing on the induction of Env-specific antibodies." PLoS One **5**(11): e13916.
Veazey, R. S., R. J. Shattock, et al. (2003). "Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120." Nat Med **9**(3): 343-346.
Vlieghe, P., V. Lisowski, et al. (2010). "Synthetic therapeutic peptides: science and market." Drug Discov Today **15**(1-2): 40-56.
von Garnier, C., M. Astori, et al. (2000). "Allergen-derived long peptide immunotherapy down-regulates specific IgE response and protects from anaphylaxis." Eur J Immunol **30**(6): 1638-1645.
Walker, L. M., M. Huber, et al. (2011). "Broad neutralization coverage of HIV by multiple highly potent antibodies." Nature **477**(7365): 466-470.
Walker, L. M., S. K. Phogat, et al. (2009). "Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target." Science **326**(5950): 285-289.
Walter, U. and P. Santamaria (2005). "CD8+ T cells in autoimmunity." Current Opinion in Immunology **17**(6): 624-631.
Watkins, J. D., N. B. Siddappa, et al. (2011). "An anti-HIV-1 V3 loop antibody fully protects cross-clade and elicits T-cell immunity in macaques mucosally challenged with an R5 clade C SHIV." PLoS One **6**(3): e18207.
Wei, X., J. M. Decker, et al. (2003). "Antibody neutralization and escape by HIV-1." Nature **422**(6929): 307-312.
White, J., F. Crawford, et al. (1999). "Soluble class I MHC with beta2-microglobulin covalently linked peptides: specific binding to a T cell hybridoma." J Immunol **162**(5): 2671-2676.

Widmann, C., P. Romero, et al. (1992). "T helper epitopes enhance the cytotoxic response of mice immunized with MHC class I-restricted malaria peptides." J Immunol Methods **155**(1): 95-99.
Wild, C. T., D. C. Shugars, et al. (1994). "Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection." Proc Natl Acad Sci U S A **91**(21): 9770-9774.
Willats, W. G. (2002). "Phage display: practicalities and prospects." Plant Mol Biol **50**(6): 837-854.
Willey, S. and M. M. Aasa-Chapman (2008). "Humoral immunity to HIV-1: neutralisation and antibody effector functions." Trends Microbiol **16**(12): 596-604.
Williams, R., S. G. Peisajovich, et al. (2006). "Amplification of complex gene libraries by emulsion PCR." Nat Methods **3**(7): 545-550.
Wilson, D. S., A. D. Keefe, et al. (2001). "The use of mRNA display to select high-affinity protein-binding peptides." Proc Natl Acad Sci U S A **98**(7): 3750-3755.
Wittrup, H. H., A. Tybjaerg-Hansen, et al. (1997). "A common substitution (Asn291Ser) in lipoprotein lipase is associated with increased risk of ischemic heart disease." J Clin Invest **99**(7): 1606-1613.
Wu, X., Z. Y. Yang, et al. (2010). "Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1." Science **329**(5993): 856-861.
Yang, H., H. Wu, et al. (2012). "Antiviral inhibitory capacity of CD8+ T cells predicts the rate of CD4+ T-cell decline in HIV-1 infection." J Infect Dis **206**(4): 552-561.
Yu, Y. Y., N. Netuschil, et al. (2002). "Cutting edge: single-chain trimers of MHC class I molecules form stable structures that potently stimulate antigen-specific T cells and B cells." J Immunol **168**(7): 3145-3149.
Zahnd, C., P. Amstutz, et al. (2007). "Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target." Nat Methods **4**(3): 269-279.
Zolla-Pazner, S. (2004). "Identifying epitopes of HIV-1 that induce protective antibodies." Nat Rev Immunol **4**(3): 199-210.

APPENDIX

2.17.1 Perl script – translation script

```
#!/usr/local/bin/perl -w

use strict;
use Bio::Seq;

use Data::Dumper;

$| = 1;


my $input_dir = '/hts/data5/xiaoyanp/NGS_JAN13/';
# my $input_dir = '/home/staff/smcgowan/xiaoyanp/test_seq_data/';

my $min_obsv_num = 0;


#### opendir -------------------------------------
opendir(DIR, $input_dir);
my @files = readdir(DIR);
closedir(DIR);


my %contaminants;
#$contaminants{ASSCYCSYTGGGGGG} = '';
#$contaminants{HRVLSLFRNGGGGGG} = '';
#$contaminants{WCYRLYMWAGGGGGG} = '';
#$contaminants{HVPFHMIPAGGGGGG} = '';
#$contaminants{GCFCDVCTAGGGGGG} = '';
#$contaminants{MDDLPVVCWGGGGGG} = '';
#$contaminants{RRPVFYSLHGGGGGG} = '';
#$contaminants{NLATVSLVYGGGGGG} = '';
#$contaminants{LKLTQHRYVGGGGGG} = '';
#$contaminants{ASSCYCSFTGGGGGG} = '';
#$contaminants{PTGHFLTPLGGGGGG} = '';
#$contaminants{KRYCELFSIGGGGGG} = '';
#$contaminants{RPSVGLFLWGGGGGG} = '';
#$contaminants{CFTARVAKMGGGGGG} = '';



foreach my $file (@files)
{
     if ($file eq "." or $file eq "..") {next;}
     unless ($file =~ m/^(WTCHG.*)_sequence\.txt$/) {next;}
     my $file_id = $1;

     my $target_seq_len = 45;  (sequence 15 amino acids after first
GATATACCATG)
#     if ($file_id =~ m/283_(1|2)$/) {$target_seq_len = 45;}
#     if ($file_id =~ m/284_(1|2)$/) {$target_seq_len = 45;}

     warn "$file_id - target length: $target_seq_len\n";

     my $input_file = $input_dir . $file;

     my %seq_data = ();
```

```
	my $fastq_line_count = 0;	(count the repeat)
	open (FASTQ1, "$input_file") or die "Couldn't open file
$input_file\n\n";		(FASTQ1 next-gen search)
	while (my $file1_line = <FASTQ1>)
	{
		chomp $file1_line;
		$fastq_line_count ++;

		if ($fastq_line_count == 1) {next;}
		elsif ($fastq_line_count == 2)
		{
			if ($file1_line =~
m/GATATACCATG(.{$target_seq_len})/)
			{
				my $amplfd_seq = $1;
				$seq_data{$amplfd_seq} ++;
			}
		}
		elsif ($fastq_line_count == 3) {next;}
		elsif ($fastq_line_count == 4) { $fastq_line_count = 0; }
	}
	close FASTQ1;

	my %counts_to_seq = ();
	foreach my $seq (keys %seq_data)
	{
		my $count = $seq_data{$seq};
		$counts_to_seq{$count}{$seq} = '';
	}

	my $out_file = $input_dir . $file_id . '.data_full';
	open (OUT, ">$out_file") or die "cannot open out file,
$out_file: $!\n";

	my $out_file_arg1 = $input_dir . $file_id . '.data_ARG1';
	open (ARG1, ">$out_file_arg1") or die "cannot open out file,
$out_file_arg1: $!\n";

	my $out_file_arg2 = $input_dir . $file_id . '.data_ARG2';
	open (ARG2, ">$out_file_arg2") or die "cannot open out file,
$out_file_arg2: $!\n";

	if ( ($file_id =~ m/29[1-6]_(1|2)$/) or ($file_id =~
m/287_(1|2)$/) or ($file_id =~ m/289_(1|2)$/))
	{
		my $out_file_IL = $input_dir . $file_id . '.data_IL';
		open (IL, ">$out_file_IL") or die "cannot open out file,
$out_file_IL: $!\n";
	}

	if ( ($file_id =~ m/287_(1|2)$/) or ($file_id =~
m/288_(1|2)$/))
	{
		my $out_file_B8 = $input_dir . $file_id . '.data_B8';
		open (B8, ">$out_file_B8") or die "cannot open out file,
$out_file_B8: $!\n";
```

```
	}


	my $total_count_full = 0;
	my $total_count_arg1 = 0;
	my $total_count_arg2 = 0;
	my $total_count_IL = 0;
	my $total_count_B8 = 0;

	my $seq_count_full = 0;
	my $seq_count_arg1 = 0;
	my $seq_count_arg2 = 0;
	my $seq_count_IL = 0;
	my $seq_count_B8 = 0;


	my %prot_freqs;
	my %prot_to_dna;

	foreach my $count (reverse sort numerically
keys %counts_to_seq)
	{
		foreach my $seq (keys %{$counts_to_seq{$count}})
		{
			# translate the seq: (translating codons)
			my $seq_obj = Bio::Seq->new(-seq => $seq);
			my $prot_obj = $seq_obj->translate(-frame => 0);
			my $prot_seq = $prot_obj->seq;

			# skip stop codons:
			if ($prot_seq =~ m/\*/) {next;}


			# select for SSC
			   if ( ($file_id =~ m/283_(1|2)$/) or ($file_id =~
m/284_(1|2)$/))
			{
			    unless ($prot_seq =~ m/SSC..........C$/) {next;}
			}

			# select for G-linker
			if ( ($file_id =~ m/20[1-9]_(1|2)$/) or ($file_id =~
m/21[0-8]_(1|2)$/))
			#if ( ($file_id =~ m/28[1-2]_(1|2)$/) or ($file_id
=~ m/28[5-9]_(1|2)$/) or ($file_id =~ m/29[0-6]_(1|2)$/) )
			#if ( ($file_id =~ m/283_(1|2)$/) or ($file_id =~
m/284_(1|2)$/))
			{
			unless ($prot_seq =~ m/GGGGGG$/) {next;}
			}

			# get rid of contaminants:
			if (exists($contaminants{$prot_seq})) {next;}

			$prot_freqs{$prot_seq} += $count;
			$prot_to_dna{$prot_seq}{$count}{$seq} = '';
```

```
        }
    }


    my %freqs_to_prot;
    foreach my $prot_seq (keys %prot_freqs)
    {
        my $total_prot_obs = $prot_freqs{$prot_seq};
        $freqs_to_prot{$total_prot_obs}{$prot_seq} = '';
    }

    my %output_prots;
    foreach my $total_prot_obs (reverse sort numerically
keys %freqs_to_prot)
    {
        foreach my $prot_seq
(keys %{$freqs_to_prot{$total_prot_obs}})
        {
            if (exists($output_prots{$prot_seq})) {next;}
            my $prot_seq_count = 0;
            foreach my $dna_count (reverse sort numerically
keys %{$prot_to_dna{$prot_seq}})
            {
                foreach my $dna_seq
(keys %{$prot_to_dna{$prot_seq}{$dna_count}})
                {

                    # write out subset of results to separate
file:
                    if ($prot_seq =~ m/^R/)
                    {
                        print ARG1
"$dna_seq\t$prot_seq\t$dna_count\n";
                        $total_count_arg1 += $dna_count;

                    }
                    elsif ($prot_seq =~ m/^.R/)
                    {
                        print ARG2
"$dna_seq\t$prot_seq\t$dna_count\n";
                        $total_count_arg2 += $dna_count;

                    }

                    if ( ($file_id =~ m/287_(1|2)$/) or
($file_id =~ m/288_(1|2)$/))
                    {
                        if ($prot_seq =~ m/^..K.K/)
                        {
                            print B8
"$dna_seq\t$prot_seq\t$dna_count\n";
                            $total_count_B8 += $dna_count;

                        }
                    }
```

```
                              if ( ($file_id =~ m/29[1-6]_(1|2)$/) or
($file_id =~ m/287_(1|2)$/) or ($file_id =~ m/289_(1|2)$/))
                              {
                                    if ($prot_seq =~ m/^.L......[L|V]/)
                                    {
                                          print IL
"$dna_seq\t$prot_seq\t$dna_count\n";
                                          $total_count_IL += $dna_count;


                                    }
                              }

                              print OUT
"$dna_seq\t$prot_seq\t$dna_count\n";

                              $total_count_full += $dna_count;
                              $prot_seq_count += $dna_count;
                        }
                  }

                  $output_prots{$prot_seq} = $prot_seq_count;
            }
      }

      # print the total num observations for each protein:
      my %order_by_total_num_obs;
      foreach my $protein_seq (keys %output_prots)
      {
            my $total_num_protein_obs = $output_prots{$protein_seq};

      $order_by_total_num_obs{$total_num_protein_obs}{$protein_seq}
= '';
      }
      print OUT "\n\n-------------------------------------------------
-\n\n";
      foreach my $total_num_protein_obs (reverse sort numerically
keys %order_by_total_num_obs)
      {
            foreach my $protein_seq
(keys %{$order_by_total_num_obs{$total_num_protein_obs}})
            {
                  print OUT "$protein_seq: $total_num_protein_obs\n";
            }
      }
      print OUT "\n\n-------------------------------------------------
-\n\n";


      print OUT "\nTOTAL: $total_count_full\n\n";
      close OUT;

      print ARG1 "\nTOTAL: $total_count_arg1\n\n";
      close ARG1;

      print ARG2 "\nTOTAL: $total_count_arg2\n\n";
```

```
      close ARG2;


      if ( ($file_id =~ m/29[1-6]_(1|2)$/) or ($file_id =~
m/287_(1|2)$/) or ($file_id =~ m/289_(1|2)$/))
      {
            print IL "\nTOTAL: $total_count_IL\n\n";
            close IL;
      }

      if ( ($file_id =~ m/287_(1|2)$/) or ($file_id =~
m/288_(1|2)$/))
      {
            print B8 "\nTOTAL: $total_count_B8\n\n";
            close B8;
      }



     # print out the total number of observations for each fragment:
      my $frag_size = 5;
      my %frag_obs;
      foreach my $protein_seq (keys %output_prots)
      {
            my $total_num_protein_obs = $output_prots{$protein_seq};

            # trim seqs of invariant parts:
            if ( ($file_id =~ m/283_(1|2)$/) or ($file_id =~
m/284_(1|2)$/))
            {
                  $protein_seq =~ s/C$//;
                  $protein_seq =~ s/^SSC//;
            }
            else  { $protein_seq =~ s/GGGGGG$//; }

            for (my $i=0; $i<20; $i++)
            {
                  if ($i >= length($protein_seq)){next;}

                  my $frag = substr($protein_seq, $i, $frag_size);
                  unless(length($frag) == $frag_size) {next;}

                  my $rev_frag = reverse $frag;
                  $frag_obs{$frag} += $total_num_protein_obs;
                  $frag_obs{$rev_frag} += $total_num_protein_obs;

                  #print "\t$frag += $total_num_protein_obs\n";
                  #print "\t$rev_frag += $total_num_protein_obs\n";
            }
      }
      my %order_by_frag_obs;
      foreach my $frag (keys %frag_obs)
      {
            my $total_num_frag_obs = $frag_obs{$frag};
            $order_by_frag_obs{$total_num_frag_obs}{$frag} = '';
```

```
	}

	my $out_file_frags = $input_dir . $file_id . '.data_frags';
	open (FRAG, ">$out_file_frags") or die "cannot open out file,
$out_file_frags: $!\n";

	foreach my $total_num_frag_obs (reverse sort numerically
keys %order_by_frag_obs)
	{
		foreach my $frag_seq
(keys %{$order_by_frag_obs{$total_num_frag_obs}})
		{
			print FRAG "$frag_seq: $total_num_frag_obs\n";
		}
	}
	close FRAG;

}

sub numerically
{
	$a <=> $b;
}
```

2.17.2 Perl script – identifying HLA-A*0201 anchor residues specific to influenza M peptide antigen, CMV pp65 peptide, and EBV BMLF-1 peptide.

```
--
#!/usr/local/bin/perl -w

use strict;
use Data::Dumper;

$| = 1;

my $input_dir = '/hts/data5/xiaoyanp/NGS_JAN13/';
#my $input_dir = '/home/staff/smcgowan/xiaoyanp/';

my $min_obsv_num = 0;

my %ids;
$ids{'WTCHG_48572_205'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_206'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_207'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_208'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_209'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_210'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
```

```
$ids{'WTCHG_48572_211'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_212'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_213'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_214'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_215'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_216'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_217'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];
$ids{'WTCHG_48572_218'} = ['.I......LGGGGGG',  '.L......LGGGGGG',
'.L......VGGGGGG'  ];



#### opendir -------------------------------------
opendir(DIR, $input_dir);
my @files = readdir(DIR);
closedir(DIR);

foreach my $file (@files)
{
      if ($file eq "." or $file eq "..") {next;}
      unless ($file =~ m/^(WTCHG.*)\_(1|2)\.data_full$/) {next;}
      my $file_id = $1;
      my $orient = $2;
      unless(exists($ids{$file_id})) {next;}

      warn "$file_id - $orient:\n";

      my $input_file = $input_dir . $file;

      my %seq_data = ();
      my $in_data = 0;
      open (DATA, "$input_file") or die "Couldn't open file
$input_file\n\n";
      while (my $line = <DATA>)
      {
            chomp $line;

            # the data we need is bracketed by a line of hyphens, so
search for these:
            if ($line =~ m/^\-\-\-\-\-/)
            {
                  if ($in_data) {$in_data = 0;}
                  else {$in_data = 1;}
                  next;
            }

            if ($in_data)
            {
```

```
                unless($line =~ m/\:/) {next;}

                # get the peptide:
               my ($peptide, $obsvns) = $line =~ m/^(.*)\:\s+(\d+)/;


                # test against pattern(s) for this sample:
                foreach my $pattern (@{$ids{$file_id}})
                {
                      if ($peptide =~ m/$pattern/)
                      {
                            $seq_data{$pattern}{$obsvns}{$peptide} =
'';
                      }
                }
          }
     }
     close DATA;

     # print results
     foreach my $pattern (@{$ids{$file_id}})
     {
          my $display_pattern = $pattern;
          $display_pattern =~ s/\./X/g;
          my $results_file = $input_dir . $file . '_' . $orient .
'.' . $display_pattern . '.hits';
          open (OUT, ">$results_file") or die "cannot open file for
results: $results_file: $!\n";
          print "\t writing to: $results_file\n";
          foreach my $obsvns_count (reverse sort numerically
keys %{$seq_data{$pattern}})
          {
                foreach my $peptide (sort
keys %{$seq_data{$pattern}{$obsvns_count}})
                {
                      print OUT "$peptide\t$obsvns_count\n";
                }
          }
          close OUT;
     }
}




sub numerically
{
     $a <=> $b;
}
```

2.17.3 Perl script – check the randomness of the library

```
#!/usr/local/bin/perl -w

use strict;
use Bio::Seq;

use Data::Dumper;

$| = 1;


my $input_dir = '/hts/data5/xiaoyanp/NGS_JAN13/';
# my $input_dir = '/home/staff/smcgowan/xiaoyanp/test_seq_data/';



#### opendir -------------------------------------
opendir(DIR, $input_dir);
my @files = readdir(DIR);
closedir(DIR);


my %frags;
$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';


$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';
$frags{XXXXX} = '';


my %frag_data;

foreach my $file (@files)
{
      if ($file eq "." or $file eq "..") {next;}
      unless ($file =~ m/^(WTCHG.*)\.data\_full$/) {next;}
      my $file_id = $1;

      if ( ($file_id =~ m/20[5-9]_(1|2)$/) or ($file_id =~ m/21[0-
8]_(1|2)$/))
      {
            my $input_file = $input_dir . $file;

            open (DATA, "$input_file") or die "Couldn't open file
$input_file\n\n";
            while (<DATA>)
            {
                  my $line = $_;
                  chomp $line;
                  foreach my $frag (keys %frags)
```

```
                {
                        if ($line =~ m/$frag/)
{$frag_data{$frag}{$file_id}{$line} = '';}
                }
            }
            close DATA;
      }
}


foreach my $frag (sort keys %frag_data)
{
      foreach my $file_id (sort keys %{$frag_data{$frag}})
      {
            foreach my $match (keys %{$frag_data{$frag}{$file_id}})
            {
                  print "$frag\t$file_id\t$match\n";
            }
      }
      print "----------------------------------------\n\n";
}
```