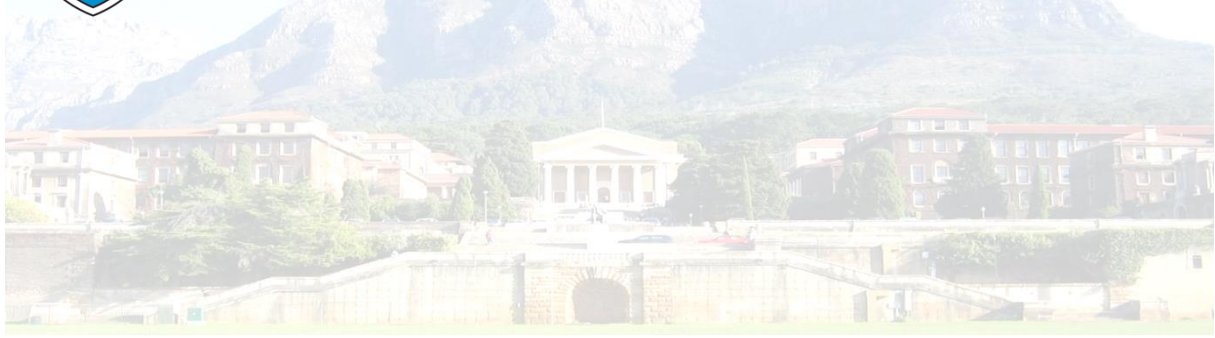




UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



Establishment of Recombinant Antibody Technologies Allowing for the Generation of SNAP-tag Fusion Proteins

by Trishana Nundalall

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Supervisor

Professor Dr. Dr. Stefan Barth

Medical Biotechnology and Immunotherapy Research Unit

Department of Integrative Biomedical Sciences

University of Cape Town

July 2022



Plagiarism Declaration

I, **Trishana Nundalall**, hereby declare that the work contained in this thesis is my own original unaided work. It is being submitted for the degree of Doctor of Philosophy at the University of Cape Town, South Africa. Neither has any part nor this whole thesis been submitted in the past, is being submitted, or will be submitted at any other university.

I have used the numbered Oncotarget referencing style as the convention for citation and referencing. Each significant contribution to, and quotation in, this thesis from the work or works of other authors and collaborators has been attributed, cited and referenced. I give the university the authority to reproduce the content, either partly or in whole for research purposes.

Signature:

Signed by candidate

Trishana Nundalall

Date: 31 July 2022

Acknowledgements

“Appreciation is a wonderful thing: It makes what is excellent in others belong to us as well.”

Francois-Marie Arouet de Voltaire

I would like to express my sincere appreciation to all the amazing humans that contributed to the realization of this thesis, without you it would not have been possible:

Firstly, to Professor Barth, you have been the greatest supporter and a pinnacle of leadership. Your unmatched love and excitement for science is infectious and has inspired me to be a better scientist.

To my mentor and friend Krupa Naran, I will be forever thankful to have met and worked with you. Your never-ending support, guidance, and confidence in me has helped me reach greater heights than I could ever imagine.

To the rest of my friends and colleagues at the MB&I team, your company, friendship, and encouragement has made my time in the lab a far richer and more enjoyable experience. Even on the days when all my experiments failed, having a cup of coffee with Neelakshi, Suresh or Anathi put a smile on my face and having random chats with Sanele or hearing Fleury sing a song I didn't recognise always brightened up day.

To my wonderful, spirit elevators/ besties, Nicole, Munadia, Nazila and Christina thanks for your many words of encouragement despite my continued complaining and for always making time to check in on me. Every shared cup of coffee or bottle of wine with you lovelies, felt like a ray of warm sunshine on a dull and gloomy day.

To my unbelievably amazing family, there are no words that could express how insanely blessed to be loved by each and everyone of you. To my beautiful mama, thanks for always believing in me, motivating me and forcing me to focus even when I didn't want to. To my wonderful Dada, thank you for always been the gentle voice of empathy and encouragement silently willing me along without judgement. To my incredible brother, thanks for inspiring me and giving me the confidence to believe in myself when I needed it the most.

And to my amazing husband, who has been on every high and low of this journey right there with me. From putting up with my horrifically grumpy moods and dealing with my unfairly directed attitude, you have been my comfort and home in the most desperate of times. Thank you for making a big deal of all the little things, like me completing a paragraph of writing or successfully executing an experiment. And for feeding me when I was too busy to feed myself and for loving me especially on the days when I didn't deserve your love. You add colour and light to my every day, without which these last 5 years would have been unbearably dark.

And a very special thanks to my larger-than life Dhada (Grandad) who is unfortunately now watching down on me from heaven. You were the greatest inspiration a kid could ever dream off. I am so proud to be the first Dr. Nundalall and will forever bear your name with pride. You made it possible for me to turn my dreams into a reality and have been such a pillar of strength and motivation in my life, I am changed forever in the best way ever because of it. I am eternally grateful to be born as your granddaughter and to have had so many special years with you present in my life.

List of Figures

Figure 1.1: Overview of the complex interactions among molecular classifications of TNBC based on genomic, transcriptomic, proteomic, epigenomic, and immune characterization of the tumour and its microenvironment.....	4
Figure 1.2: Graphically illustration of the diverse recombinant antibody formats that exist compared to a full length, classical antibody.....	9
Figure 1.3: Antibody-drug conjugate mechanism of action.....	11
Figure 1.4: Schematic representation of M13 filamentous phage life cycle.....	19
Figure 2.1.1: Representation of an assembled scFv containing a 5' <i>Sfi1</i> restriction enzyme recognition site, followed by a antibody heavy chain sequence, a (Gly4)Serine linker sequence, a light chain antibody sequence and a 3' <i>Not1</i> restriction enzyme recognition site.....	35
Figure 2.2.2: Vector map of pCB-Annexin V-SnapF. The <i>Sfi1</i> and <i>Not1</i> restriction sites were used for cloning of the various newly designed ScFv's into the expression plasmid.....	36
Figure 3.1.1: Schematic overview of a PIT2 phagemid vector	49
Figure 3.1.1.1: Enzyme-linked immunosorbent assay (ELISA) results of PIT2-Ki4 propagated libraries	51
Figure 3.1.2.1.1 PIT2-αCSPG4 construct (A) In-silico restriction cloning of αCSPG4 (scFv) into PIT2 using Sfi1 and Not1	53
Figure 3.1.2.1.2 PIT2-αEGFR construct (A) In-silico restriction cloning of αEGFR (scFv) into PIT2 using Sfi1 and Not1	54
Figure 3.1.2.2.1: Agarose gel depicting DNA fragments after <i>sfi1</i> and <i>Not1</i> double digest. The 5400bp PIT2 (vector), 924bp αCSPG4(scFv) and 757bp EGFR(scFv) fragments were all isolated for ligation	55
Figure 3.1.2.3.1: Growth of <i>E. coli</i> cells transformed with potential recombinant (B,D) PIT2-αCSPG4 and (B,E) PIT2-αEGFR plasmid DNA. Following transformation, 100 µl of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C.....	56
Figure 3.1.2.4.1: Analysis of restriction mapping of potential PIT2-αCSPG4 and PIT-αEGFR clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.....	57
Figure 3.2.2.1: Insilico-cloning strategy using αCD3 scFv as a representative target. Sfi1 and Not1 restriction enzymes were used to replace the annexin scFv with other scFvs of interest.....	63

Figure 3.2.2.2 : Open reading frames of (A) α CD3-SNAP (B) α CD4-SNAP (C) α CD8-SNAP (D) α CD19-SNAP (E) α CD20-SNAP (F) α PDL1-SNAP and (G) α B7.1-SNAP. The different elements present on all ORFs include IgK leader – Ig Kappa leader, 6xHIS- N-terminus poly histidine, EKS – enterokinase site, SNAP-tag and stop codon. The scFv in each ORF is flanked by a *Sfi1* and *Not1* recognition site which is used during the cloning of each construct.....64

Figure 3.2.3.2.1: Agarose gel depicting DNA fragments after *sfi1* and *Not1* double digest (DD). 753bp α CD20, 741bp α PDL1 and 7272bp snapF plasmids was isolated for ligation.....66

Figure 3.2.3.2.2: Agarose gel depicting DNA fragments after *sfi1* and *Not1* double digest (DD). 768bp α CD4, 768bp α CD3 and 741bp α CD8 plasmids were isolated for ligation.....67

Figure 3.2.3.3.1: Growth of *E. coli* cells transformed with potential recombinant pCB-CD4(scFv)-snapF, pCB-CD3(scFv)-snapF, pCB-CD8(scFv)-snapF, pCB-PDL1(scFv)-snapF and pCB-CD20(scFv)-snapF plasmid DNA. Following transformation, 100 μ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 μ g/ml. 69

Figure 3.2.3.2 . Analysis of restriction mapping of undigested pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All clones were separated using a 1.2% (w/v) agarose gel ran for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting gel image after electrophoresis.....70

Figure 3.2.3.4: Analysis of restriction mapping of BamH1 digested potential pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All digests were run at 37°C overnight. The DNA was separated using a 1.2% (w/v) agarose gel ran for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting gel image after electrophoresis.....72

Figure 3.2.3.6: Analysis of restriction mapping of Pst1 digested, potential pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All digests were run at 37°C overnight. The DNA was separated using a 1.2% (w/v) agarose gel ran for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting agarose gel after electrophoresis.....73

Figure 3.2.4.1: Alignment map generated following the alignment of trace sequences with the corresponding in silico sequence for (A) pCB- α CD3-snap (B) pCB- α CD4-snap (C) pCB- α CD8-snap (D) pCB- α CD20-snap and (E) pCB- α PDL1-snap. The clones corresponded with 100% homology to the in silico ORF sequences as shown by the dark red arrows.....75

Figure 3.2.5.1: Dot plots depicting the eGFP+ and eGFP- cell populations in (A) pCB- α CD3-snap (B) pCB- α CD4-snap (C) pCB- α CD8-snap (D) pCB- α CD20-snap (E) untransfected HEK293T and (F) pCB- α PDL1-snap cultures.....76

Figure 3.2.5.2: Microscopic visualization of eGFP in transfected HEK293T cells. Transfection was carried out with the XtremeGene™ transfection reagent (Sigma-Aldrich, South Africa) according to manufacturer's instructions. Enrichment of eGFP was performed with 600 µg/ml of Zeocine. The left panel depicts a brightfield (phase contrast) channel which shows all the cells in a specific region. The centre panel depicts the green channel depicting the eGFP expressing cells and the right panel is an overlay of both channels. All images were taken with a ZOE Fluorescent Cell Imager at 100 µm magnification. 77

Figure 3.2.6.1: SDS-PAGE analysis of pre-concentrated IMAC purified pCB-αCD4-SNAP protein fractions. Following IMAC purification, a 10% SDS-PAGE gel was run to confirm the expression of the pCB-αCD4-SNAP protein based on its molecular weight and rate of migration through the gel. The gel was loaded with 15 µl of protein in each well and ran at 100 volts for 95 minutes. The bands were subsequently visualized by Aqua stain solution. Page Ruler loaded in lanes 5 and 11 were used to determine the proteins molecular weight.....78

Figure 3.2.6.2: Assessing the presences and binding activity of pCB-αCD19-SNAP and pCB-αCD20-Snap by SDS and conjugation to BG-Alexa Fluor 488. A ratio of 1:2 of protein to BG-Alexa Fluor 488 was used in the conjugation reaction (5 µM to 10 µM) to αCD20(scFv)-SNAP (51 kDa) and αCD19(scFv)-SNAP(50 kDa). Left panel: Alexa488-conjugated protein ran on a 10% SDS-PAGE gel stained with Aqua staining solution.....79

Figure 3.3.1.1: IgBLAST analysis of αBCK-1 (scFv) extracted (A) heavy-chain and (B) light chain nucleotide sequences compared to previously identified immunoglobulin germline variable region gene sequences. Variations from the germline sequences are depicted in pink.....83

Figure 3.3.1.2: Assembled BCK-1 (scFv) sequence for cloning. The different components of the sequence are highlighted in different colours; blue is the Sfi1 recognition site, green is the heavy chain CDRs, red is the Gly(4)Ser linker, purple are the light chain CDRs and orange is the Not1 recognition site.....84

Figure 3.3.2.1: pCB-BCK1-Snap construct. (A) In-silico restriction cloning of BCK-1(scFv) into pCB-AnnV-SNAP using Sfi1 and Not1. (b) ORF of pCB-BCK1-SNAP construct.....85

Figure 3.3.3.2.1: Agarose gel depicting DNA fragments after sfi1 and Not1 double digest: 7269bp pCB backbone and 745bp αBCK1 insert was extracted from gel for ligation.....87

Figure 3.3.3.3.1: Growth of *E. coli* cells transformed with potential recombinant pCB-BCK1-snap plasmid DNA. Following transformation, 100 µl of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 µg/ml. (A) Contamination control (B) Vector-only plate (C) 1:1 ratio of vector to insert (D) 1:3 ratio of vector to insert (E) 1:5 ratio of vector to insert.....88

Figure 3.3.3.4.1: Theoretical gel simulation using Snapgene and corresponding band sizes for each enzyme.....89

Figure 3.3.3.4.2 . Analysis of restriction mapping of potential pCB-BCK-1-Snap clones using agarose gel electrophoresis.....90

Figure 3.3.3.5.1: Map generated following the alignment of trace sequences with the corresponding in-silico sequence for pCB-BCK1-SNAP. Both clones showed 100% homology with the in-silico ORF sequence as indicated by the dark red sections of the arrows.....	91
Figure 3.3.3.1: Green fluorescent protein expression in pCB_BCK1_Snap transfected HEK293T cells one week post transfection	92
Figure 3.3.4.1: Chromatogram of BCK1-SNAP's purification profile during IMAC. The x-axis represents the AKTA flow-through volume over time while the Y-axis represents the percentage of the elution buffer. The green line indicates the imidazole concentration and the blue line shows the elution profile of the protein.	93
Figure 3.3.4.2: SDS-PAGE analysis of pre-concentrated IMAC eluate BCK1-SNAP protein fractions. 10% SDS-PAGE gel loaded with 15µl of sample and run at 100 volts for 95 minutes were used to confirm expression of BCK1-SNAP proteins. The protein bands were visualized with Aqua staining solution. Page Ruler pre-stain protein ladder was used to determine the protein molecular weights.....	93
Figure 3.3.4.3: SDS-PAGE analysis of concentrated IMAC eluate BCK1-SNAP protein fractions. Amicon concentrated protein fractions were run on an SDS-PAGE gel at 100 volts for 95 minutes. 20 µg of total sample from each peak was loaded on lanes 2 and 3 respectively.....	94
Figure 3.4.1.1: IgBLAST analysis of αBCK-1 (scFv) extracted (A) heavy-chain and (B) light chain nucleotide sequences compared to previously identified immunoglobulin germline variable region gene sequences. Variations from the germline sequences are depicted in pink.....	98
Figure 3.4.1.2: Assembled ZIKA-DIII (scFv) sequence for cloning. The different components of the sequence are highlighted in different colours; dark orange is the Sfi1 recognition site, green is the heavy chain CDRs, light orange is the Gly(4)Ser linker, purple are the light chain CDRs and red is the Not1 recognition site.....	99
Figure 3.4.2.1: pCB-αZIKA-DIII-Snap construct. (A) In-silico restriction cloning of αZIKA-DIII (scFv) into pCB-AnnV-SNAP using Sfi1 and Not1. (b) ORF of pCB-αZIKA-DIII-SNAP construct including Sfi1 and Not1 recognition sites; IgK leader: Ig-Kappa leader sequence important for secretion of the fusion protein during host expression, 6xHis : 6 histidine tags used for protein purification during IMAC and detection in western blot analysis, EKS: enterokinase cleavage site important for enzymatic removal of the N-terminal elements and STOP: stop codon for termination of protein synthesis.....	100
Figure 3.4.3.2.1: Agarose gel depicting DNA fragments after sfi1 and Not1 double digest: 711bp αZIKA-DIII insert was extracted from gel for ligation	102
Figure 3.4.3.3.1: Growth of <i>E. coli</i> cells transformed with potential recombinant pCB-αZIKA-DIII-snap plasmid DNA. Following transformation, 100 µl of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 µg/ml. (A) Contamination control (B) Vector-only plate (C) 1:1 ratio of vector to insert (D) 1:3 ratio of vector to insert (E) 1:5 ratio of vector to insert.....	103

Figure 3.4.3.4.1: Theoretical gel simulation using Snapgene and corresponding band sizes for each enzyme.....	104
Figure 3.4.3.4.2 . Analysis of restriction mapping of potential pCB-ZIKA-DIII-Snap clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.....	105
Figure 3.4.3.4.3 . Analysis of restriction mapping of potential pCB-ZIKA-DIII-Snap clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.....	106
Figure 3.4.3.5.1: Map generated following the alignment of trace sequences with the corresponding in-silico sequence for pCB-ZIKA_DIII-SNAP. Both clones showed 100% homology with the in-silico ORF sequence as indicated by the dark red sections of the arrows.....	108
Figure 3.4.4.1: Green fluorescent protein expression in pCB-ZIKA-DII-Snap transfected HEK293T cells one week post transfection.....	109
Figure 3.4.5.1: Assessing the presences and binding activity of pCB-αZIKA-DIII-snap by SDS and conjugation to BG-Alexa Fluor 488.....	110
Figure 4.3.6.1: Enzyme-linked immunosorbent assay (ELISA) results of pCB-ZIKA-DIII-Snap concentrated cell culture supernatant (CCSN). 100µl of concentrated CCSN was coated onto wells of (A)150 ng or (B) 200ng envelope-protein pre-coated plates and incubated for 1 hour at room temperature. Optical density was measured at OD 450 nm. The red dotted line indicates the OD reading of a positive control plated at 100µg/µl.....	111

Table of Contents

Abstract	1
Section 1: Breast Cancer	1
Introduction.....	1
Causes of Breast Cancer.....	1
Types of Breast Cancers.....	1
Triple Negative Breast Cancer	3
Therapy for TNBC	4
Section 2: Precision Medicine and Immunotherapy	6
Immunotherapy	7
Monoclonal Antibodies	8
Section 3: Antibody-based immunotherapies	11
Antibody conjugates	11
Checkpoint Inhibitors	14
Cytotoxic T-lymphocyte antigen (CTLA-4) and B7.1/B7.2	14
PD1 and PD-L1	15
Clinical Applications	16
Antibody based immunotherapy of TNBC	17
Section 4: Antibody Discovery and applications.....	19
Traditional methods of antibody isolation.....	19
Phage Display	19
Alternative methods of antibody discovery.....	24
Patents as a source of antibody sequences and information	24
IgBlast:	25
Antibody fusion proteins	27
Fusion Protein Domains	27
Linkers.....	27
SNAP-tag Technology	29
Section 5: Applying the Lessons and Tools from Cancer Immunotherapy to Immunotherapy of infectious diseases.....	30
Section 6: Aims and Objectives	32
2.1 Assembly of scFvs from antibody sequences identified in publicly available sources ...	36

2.2 In Silico Cloning.....	36
2.3 Molecular Cloning	39
2.3.1 Transformation of chemically competent DH5 α e.coli cells with plasmid DNA	39
2.3.2 Phenol-chloroform extraction of bulk prepared vector DNA.....	39
2.3.3 DNA Quantification.....	41
2.3.4 Restriction Enzyme Digest	41
2.3.5 Agarose gel Electrophoresis	41
2.3.6 Recovery of DNA fragments from Agarose gels	42
2.3.7 T4 DNA Ligation	42
2.3.8 Mini preparation of transformed colonies.	43
2.3.9 Restriction analysis	43
2.4 Sequence verification	44
2.5 Mammalian cell culture and protein expression	44
2.5.1 Flow Cytometry	45
2.5.2 Purification of recombinant proteins	45
2.5.3 Protein Quantification by Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)	46
2.6 Functional Assays	46
2.6.1 Western blot.....	46
2.6.2 Protein labelling with BG-Alexa Fluor [®] 488.....	47
2.6.3 Enzyme-Linked Immunosorbent Assay (ELISA)	47
Section 3.1: Constructing a two-clone phage library aiming to optimize a live cell selection method	49
3.1.1 Selection of Ki4-scFv from Tomlinson I+J PIT naïve scFv library.....	51
3.1.2 Generation of a two-component phage library	53
3.1.2.1 : In-silico cloning of PIT- α CSPG4 and PIT2- α EGFR.....	53
3.1.2.2 Restriction Enzyme Digest and agarose gel electrophoresis of PIT2-Ki4, pCB- α CSPG4-SNAP and pCB- α EGFR-SNAP	54
3.1.2.3 Ligation of α CSPG4(scFv) and α EGFR(scFv) into the PIT2 backbone	55
3.1.2.4 Restriction mapping of potential PIT2- α CSPG4 and PIT2- α EGFR clones	57
Summary of section 3.1 results	58
Section 2: Developing a differential diagnostic panel for the classification of TNBC tumours	59
3.2.1 Identification and assembly of scFv sequences.....	59

3.2.2 In-silico Cloning.....	62
3.2.3 Molecular cloning of different recombinant SNAP fusion proteins.....	65
3.2.3.1 Bulk preparation of DNA plasmids by transformation.....	65
3.2.3.2 Restriction enzyme digest and agarose gel electrophoresis.....	65
3.2.3.3 T4 DNA Ligation	68
3.2.3 Restriction Mapping of Potential Recombinant Clones	70
3.2.4 Sequence Analysis of Selected Recombinant clones	74
3.2.5: Transfection into HEK293T cells for protein expression	76
3.2.6: SDS and BG Alexa Flour 488 Analysis	78
Section 3.2.7: Summary of Section 3.2 Results.....	81
<i>Section 3.3: Constructing a novel TNBC targeting snap-fusion protein, pCB-αBCK1-snap which has potential diagnostic and therapeutic applications.....</i>	82
3.3.1 Validation and assembly of BCK-1 scFv sequence	82
3.3.2 In-silico cloning of pCB-BCK1-snap.....	85
3.3.3 Molecular cloning of pCB-BCK1-snap	86
3.3.3.1 Bulk preparation of DNA plasmids by transformation.....	86
3.3.3.2 Restriction enzyme digest and agarose gel electrophoresis.....	86
3.3.3.3 Ligation of BCK1(scFv) into the pCB-SNAP backbone	88
3.3.3.4 Restriction mapping of potential recombinant pCB-BCK1-SNAP clones.....	89
3.3.4 Sequencing results of selected pCB-BCK1-SNAP clones.....	91
3.3.3 Transfection and expression of pCB-BCK1-SNAP in HEK293T mammalian expression cells.	91
3.3.4 SDS-PAGE analysis of purified pCB-BCK1-SNAP protein fractions	92
3.3.5 Western blot analysis of purified BCK-SNAP protein.....	94
3.3.6 Summary of section 3.3 results: the construction of BCK1-SNAP protein.....	95
<i>Section 3.4: Applying the tools and technologies used to develop cancer specific snap-fusion proteins to infectious diseases using Zika virus as an example.</i>	96
3.4.1 Validation and assembly of pCB-αZIKA-DIII-Snap scFv sequence.....	97
3.4.2 In-silico cloning of pCB-ZIKA_DIII-snap.....	99
3.4.3 Molecular cloning of pCB-αZIKA-DIII-snap.....	101
3.4.3.1 Bulk preparation of DNA plasmids by transformation.....	101
3.4.3.2 Restriction enzyme digest and agarose gel electrophoresis.....	101
3.4.3.3 Ligation of α ZIKA-DIII (scFv) into the pCB-SNAP backbone.....	103
3.4.3.4 Restriction mapping of potential recombinant pCB- α ZIKA-DII-SNAP clones	104
3.4.3.5 Sequencing results of selected pCB-ZIKA-DIII-SNAP clones	108
3.4.4 Transfection and expression of pCB-ZIKA-DIII-SNAP in HEK293T mammalian expression cells.	109
3.4.5 SDS-PAGE analysis of pCB-ZIKA-DIII-SNAP cell culture supernatant.....	110
4.3.6 Investigating the binding activity of pCB-ZIKA-Snap to target antigens	111

Section 3.4.6: Summary of Results in Section 3.4	112
4.1 Triple Negative Breast Cancer: A need for medical intervention	113
4.2 Identifying potential therapeutic antibody sequences through data mining and patent searching.	115
4.3 Phage Display as a tool for novel antibody identification.....	116
4.3 SNAP-tag Technology as an important tool in development of immunodiagnostic and immunotherapeutic fusion proteins	119
4.4 Expression and first step purification of recombinant SNAP-tag based fusion proteins	121
4.5 SNAP-tag based immunodiagnostics and therapeutics for classification and treatment of TNBC.	123
4.6 Zika Virus : a proxy for the use of SNAP-tag technology in Infectious Diseases.	125
4.7 Conclusions and Future Developments.	127
References.....	129

Abstract

Triple negative breast cancer (TNBC) is a highly aggressive and invasive subtype of breast cancer, typically characterised by the lack of estrogen receptor (ER), progesterone receptor (PR) and Human epidermal growth factor receptor 2 (HER2) with an inexplicable partiality towards African women. The acute heterogeneity and complexity of TNBC tumours, together with a lack of well-defined molecular targets, complicates prognosis of the diseases resulting in patient reliance on traditional therapies, like chemotherapy, radiotherapy, and surgery, which are associated with elevated incidence of adverse effects and relapse. A major contributor to the heterogeneity of TNBCs is the tumour microenvironment which is composed of tumour infiltrating lymphocytes (TILs), tumour cells, healthy cells, and tumour vasculature. TILs have commonly been used as a prognostic marker and show robust predictive value for TNBC. In-depth analysis of the TIL composition within TNBC tumours may provide greatly beneficial information for the development of newer tumour microenvironment changing therapies and could assist treating physicians in understanding what therapies a particular patient maybe susceptible to.

Thus, the diagnosis and therapy of this disease may greatly benefit from improved molecular profiling and patient stratification. Precision medicine seeks to provide such a solution, by dividing patients into subpopulations based on disease-specific profiles. The identification of new molecular targets would provide the basis for development of novel therapies. To this end, one of the major aims of this thesis was to develop a phage display based screening technique which could be utilised to isolate novel TNBC specific cancer antibodies. Once selected these antibodies could be used to generate TNBC specific therapies. Specific monoclonal antibodies (mAbs) and derivatives thereof, have already been established as a revolutionary tool for drug delivery to cancerous cells. Such antibodies have been conjugated to cytotoxic drugs to form antibody-drug conjugates, which may exhibit multiple advantages over their unconjugated counterparts, but their general use in clinical application has been restricted due to developmental deliberations.

Historical conjugation strategies used for the generation of ADCs commonly resulted in heterogeneous mixtures of ADC species, with varying drug-to-antibody ratios resulting in unpredictable pharmacologic characteristics and safety profiles. In more recent time, self-labelling tags such as CLIP-tag, Halo-tag and SNAP-tag have provided a means of developing homogenised recombinant immunotherapeutics. SNAP-tag is a modified version of a human DNA repair enzyme, O⁶-alkylguanine-DNA-alkyltransferase (AGT) which naturally removes alkyl residues from damaged DNA. The enzyme reacts specifically with benzylguanine (BG) derivatives via irreversible transfer of alkyl groups to cysteine residues forming stable end products. In this thesis, SNAP-tag technology, together with other antibody discovery and manipulation tools was used to develop a methodology allowing for the generation of disease specific fusion proteins. Specifically, these fusion proteins consist of single-chain antibody fragments genetically fused to SNAP-tag, allowing for the generation of recombinant ADCs

that could be used as a drug delivery system carrying any BG-modified drug to a disease specific targets. In addition, SNAP-tag interacts with BG in a 1:1 stoichiometry giving rise to homogenised combination products which when fused to a scFv provides a promising target-specific therapeutic option.

In addition to antibody conjugates, one of the most promising of all mAb based therapies currently used, are checkpoint inhibitors. In a balanced immune response, immune activation is counteracted with immunoregulatory pathways such as checkpoint inhibition. These negative regulatory pathways are necessary for maintaining tolerance and preventing hyperactivation, and are governed by cell surface, inhibitory receptors known as "checkpoint inhibitors". Blocking of checkpoint pathways during chronic infections and cancer has been shown to improve T-cell functions leading to reduced viral load and tumour burden. These findings have been translated into clinical application where checkpoint inhibitors, which are monoclonal antibodies targeting CTLA-4, PD1, PD-L1 or other inhibitory ligands, have been used to block these inhibitory interactions.

The main intention of this research was to develop a methodology which could be used to generate SNAP-tag based recombinant fusion proteins with potential diagnostic and therapeutic applications. Several SNAP-tag based fusion proteins were developed using the recommended methodology these included fusion proteins targeting breast cancer specific antigen BCK1, checkpoint inhibitors PDL1, B7.1/CD80 (interacts with CTLA-4), and TIL characterising markers CD3, CD4, CD8, CD19 and CD20. In addition, to demonstrate the versatility and robustness of this methodology we sought to develop a SNAP-tag based fusion protein not targeting breast cancer related antigens. Zika virus, an emerging infectious disease, currently lacking specific therapies was chosen for this purpose. An scFv derived from antibodies targeting the the Zika-DIII envelop protein, which is essential to the viral infection cycle was used in the snap fusion protein. The resulting ZIKA-DII-snap fusion protein demonstrated specific binding to zika virus membrane fractions.

This research demonstrates the feasibility of using SNAP-tag technology as a state-of-the-art conjugation strategy capable of bypassing the challenges previously associated with using antibodies as an effective delivery system for therapeutic drugs. By combining the applicability of SNAP-tag technology with other antibody isolation and manipulation tools we were able to generate several functional SNAP-tag based recombinant fusion proteins. Establishment of this methodology represents an important first step in generating medically necessary, pharmaceutically acceptable immunoconjugates that is instrumental in shifting general therapy towards a more personalized precision medicine approach.

Chapter 1

Section 1: Breast Cancer

Introduction

Breast cancer (BC) is the most common and second most lethal malignancy in females worldwide with a large disparity in survival rates. According to the South African National Cancer Registry, breast cancer makes up 21% of all cancers in females with a corresponding 16% mortality rate making it a significant public health problem in South Africa. While the rate of breast cancer in African women is not as high as the global incidents, African women tend to present with symptoms at an earlier age and have far more aggressive forms of breast cancer than Caucasian women [1]. In addition African women are generally diagnosed later in disease progression than in wealthier parts of the world [1]. This is likely due to the lack of proper health care facilities in Africa resulting in many African women using unorthodox/traditional treatment and prolonging clinical checks until the disease has advanced to a later stage. In contrast, western countries have free health care and regular screening laws which allow early detection of breast cancer that assists with treatment outcomes. Most cancers which are diagnosed late in disease progression are incurable [2]. This is largely due to the inherently robust physiology of cancer, which permits cancerous cells to survive, adapt and proliferate even when treated with anti-cancer drugs [3].

Causes of Breast Cancer

Breast cancers, like most other cancers, are characterised by the accrual of an assortment of genetic changes/mutations and the loss of normal cellular regulatory processes [3]. Cancer causing mutations may be caused by numerous factors; exposure to radiation, being overweight, spending too much time in the sun, certain hormone therapies, alcohol intake, etc [4]. In addition to external causes, cancer causing mutations may also be hereditary. Around 5-10% of breast cancers are caused by gene mutations which are passed from parent to child. The BRCA1 and BRCA2 gene are the 2 most common hereditary genes that drastically increases ones risk of acquiring breast cancer [4]. Depending on the type of mutation and resulting disease, breast cancer has been subdivided into 5 broad categories, ductal carcinoma *in situ*, invasive carcinoma, triple negative breast cancer, inflammatory breast cancer and metastatic breast cancer.

Types of Breast Cancers

Ductal carcinoma *in situ* is the earliest stage of cancer; it is when abnormal cancerous cells are found in the lining of breast milk ducts [4]. These cells are not invasive and do not harm the surrounding breast tissue. In contrast, invasive ductal carcinomas (IDC) are caused by cancerous cells that originate in the lining of breast milk ducts but invade the surrounding breast tissues [4]. In rare cases, cancer cells block lymph vessels in the skin of the infected breast resulting in a swollen, red or inflamed looking breast and is thus classified as

inflammatory breast cancer. Majority of inflammatory breast cancers originate from IDC's. In other cases, IDC cells may spread to tissue in other parts of the body as well; these are known as infiltrative ductal carcinomas or metastatic breast cancer. Metastatic breast cancer is also referred to as stage 4 cancer. Metastasis may occur by tumour invasion of nearby tissue or entry of cancer cells into the circulatory or lymph system allowing for the spread of diseased cells that give rise to metastasized tumours. IDC makes up 70-80% of all diagnosed breast cancers, while the other 10-20% are triple negative breast cancers (TNBC). TNBC are cancers which are classically characterized by a lack of progesterone (PR), oestrogen(ER) and human epidermal growth factor 2 receptors (HER2) in the tumour [4]. PR, ER and HER2 are the most common receptors known to fuel breast cancer growth. Using gene expression profiles of these receptors and other genes, breast cancer has been subdivided into 5 intrinsic molecular subtypes of breast cancer, as shown in Table 1 below.

Table 1: Intrinsic molecular subtypes of breast cancer based on gene expression profiles.[5]

Breast Cancer Subtypes	Gene Expression					
	Oestrogen Receptor	Progesterone Receptor	Human Epidermal Growth Factor Receptor	Cytokeratin 5/6	Epidermal growth factor 1 (EGFR 1)	Unclassified
Luminal A	+/-	+/-	-	-	-	-
Luminal B	+/-	+/-	+	-	-	-
HER2 +	-	-	+	-	-	-
Basal Like	-	-	-	+	+	-
Breast like	-	-	-	-	-	+

The “+” indicates positive expression of the receptor/molecule within the cancerous tumours. The “-” indicates that this subtype lacks certain receptors/molecules in their tumours.

These subgroups are significantly different in their occurrence, risk factors, prognosis and treatment outcomes [6]. Luminal A cancer has better outcomes than any of the other subtypes when diagnose. The distinguishing characteristic of both Luminal cancers is an increase in proliferation/cell cycle processes and Luminal/hormone regulated pathways [6]. In comparison to Luminal A, Luminal B tumours have higher expression of proliferative genes but lower expression of Luminal related genes, like PR [7]. In addition Luminal B tumours have higher numbers of mutations in the genome and more chromosomal copy-number changes than Luminal A tumours [7]. Furthermore, about 6-8% of Luminal A and 16-20% of Luminal B cancers display overexpression of HER2, similar to that of the HER2 subtype.

HER2+ cancers are characterised at an RNA and protein level by high expression of HER2 and proliferation related genes, medium expression of luminal related genes and low basal related gene expression. These tumours also have the highest number of genomic mutations, followed by the basal-like subtype [6]. A small percentage (2-7%) of basal-like cancers also shows over expression of HER2. In addition, basal-like cancer is characterised by elevated expression of proliferation-related genes and keratins, intermediate expression of HER2 and low expression of luminal-related genes [6]. The BRAC1-mutation is associated with Basal-like cancer [8].

Triple Negative Breast Cancer

Majority of basal-like cancers, together with breast-like and others are all collectively classified as triple negative breast cancer (TNBC) due to a lack of/reduced number of surface expression of HER2, PR and ER. This form of cancer disproportionately affects premenopausal women of Hispanic and African descent. TNBC's are biologically more aggressive than most breast cancers resulting in death in a large proportion of patients [9]. Compared to other forms of breast cancers, TNBCs tumours are usually larger and less differentiated and are 2.5 times more likely to metastasized within 5 years of diagnosis [10]. Moreover, TNBC preferentially metastasizes to the viscera, which has been associated with poor prognosis compared to non-TNBC which disseminate to the bones. Therefore, the median time to death is shorter, 4.2 years versus 6 years compared to other breast cancer subtypes. Furthermore, TNBC's are largely heterogeneous, making detection, prognosis and treatment of the tumour challenging [10]. Due to the lack of receptors in TNBC patients, no endocrine or currently available endocrine receptor targeted therapies can be used as treatment options [11]. Additionally, standard treatment options such as surgery, chemotherapy and radiation are associated with an elevated risk of local and systemic relapse [12]. Consequently, TNBC's represent a major challenge in breast oncology [12].

Therapy for TNBC

Being resistant to conventional endocrine treatments, TNBC patients require higher and more frequent doses of radio- and chemotherapy than other cancer patients[12]. This results in a heightened social and economic burden for post-therapy care of TNBC patients. In addition, there is an elevated rate of relapse in TNBC patients after conventional therapies and thus a need for other therapeutic interventions. Over and above the difficulty in treating TNBC, management of the diseases is also challenging due to the extensive molecular heterogeneity [10]. In the past decade, a major effort has gone towards classifying TNBCs into distinct clinical and molecular subtypes which may guide targeted treatment strategies. Figure 1.1 below demonstrates the complexity of TNBC heterogeneity based on classifications by different “-omic” methods.

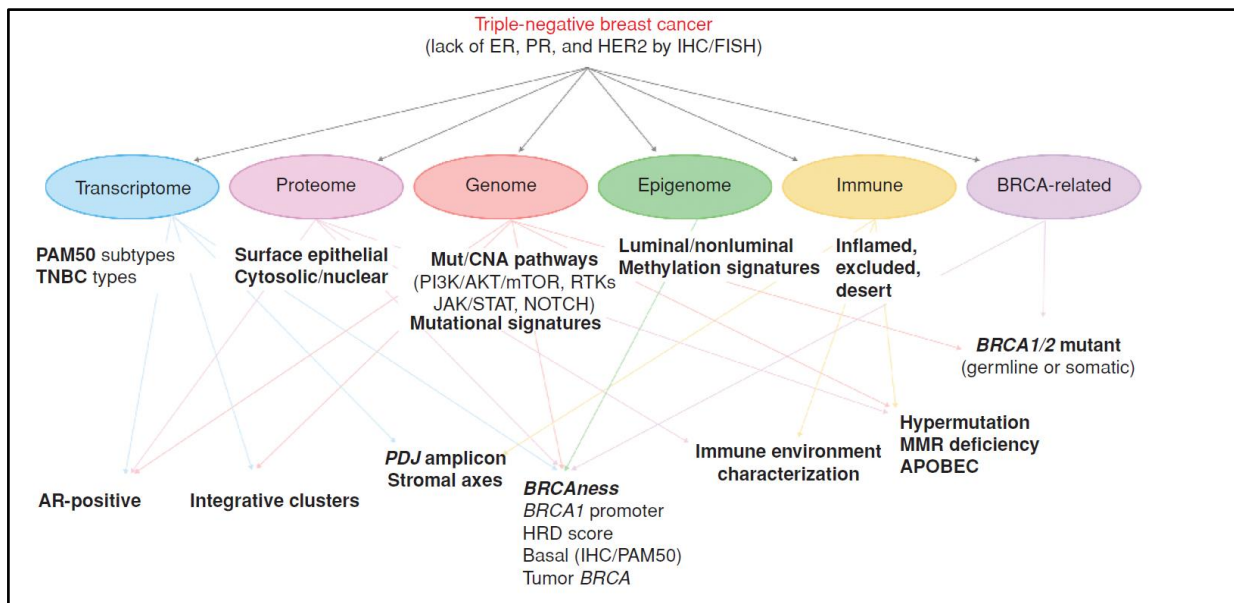


Figure 1.1: “Overview of the complex interactions among molecular classifications of TNBC based on genomic, transcriptomic, proteomic, epigenomic, and immune characterization of the tumor and its microenvironment. ER, estrogen receptor; PR, progesterone receptor; Mut, mutant; RTK, receptor tyrosine kinase, MMR, mismatch repair; CNA, copy-number alteration; AR, androgen receptor; HRD, homologous recombination deficiency; IHC, immunohistochemistry [13].

Even with the omics classification, TNBC tumours of the same subtype may still exhibit variable responses to cancer treatments. One major contributing factor to this variation is the tumour microenvironment, which consists of tumour cells, healthy cells, the tumour vasculature, and tumour infiltrating lymphocytes. Tumour infiltrating lymphocytes (TILs) are commonly used as prognostic markers in breast cancer [14], showing robust and predictive value in TNBC. TILs assist in determining the immune status of a tumour and indicate the potential efficacy of conventional chemotherapy [15, 16]. While CD8+ T-cells have always been known to play an anti-tumour role in cancer, some studies have demonstrated paradoxical results suggesting that some CD8+ T-cell infiltrated tumours still have poor outcomes [17]. In other tumours, efficient CD8+ T-cell infiltration is correlated with good outcomes. However, some patients have poor outcomes due to CD8+ T-cell accumulation in the tumour-associated stroma, with poor infiltration into the tumour epithelium. An abundance of TILs imply that the immune status of a tumour is near an equilibration stage

[18]. In such cases, surgical resection of primary tumours has been shown to skew the balance in favour of the immune system thus resulting in better prognosis with high TILs in TNBC. TNBC tumours with reduced TILs show a correlation with greater clonal heterogeneity and increased mutational load, that may consequently result in tumour escape from immune surveillance [19]. More in-depth analysis of the TIL composition within TNBC tumours may provide greatly beneficial information for the development of newer tumour microenvironment changing therapies. It also has the potential to assist doctors in understanding what therapies a particular patient may be susceptible to.

Section 2: Precision Medicine and Immunotherapy

Given the current lack of therapy and complex heterogeneity of breast cancer, the diagnosis and therapy of such a disease may greatly benefit from improved molecular profiling and patient stratification. Precision medicine seeks to provide such a solution, by dividing patients into subpopulations based on disease-specific profiles [20]. The major aim of precision medicine activities worldwide is to customize healthcare based on such granular stratifications allowing for the identification of 1) disease-specific-pathophysiology, 2) susceptibility to diseases, 3) prognosis/diagnosis markers and 4) response to specific treatments.

Currently, all BC tumours are classified by histopathology which allows for the identification of morphological heterogeneity and immunohistochemistry to phenotypically characterize biomarker expression profiles [21]. These basic classification methods however, nullify the precision medicine paradigm with two major limitations: 1) the large variability in therapeutic response and clinical outcomes of patients, even those with similar clinical and pathological features and 2) this classification does not take into consideration the significance of the biology and molecular pathways that separate BC into distinct subtypes and stages [20]. Thus, an improved classification approach is required. A notable example of improvement in such an area is TNBCs. As previously mentioned TNBCs are classically classified by a lack of ER, PR and HER2 receptors on tumours obtained from patient biopsies. Gene expression assays however, can be used to further differentiate TNBCs into 6 different subpopulations, each with a different set of relevant therapeutic targets. However, there is presently no routine diagnostic assay that distinguishes these subpopulations of TNBCs, and it remains to be determined if treatment approaches informed by these classifications are associated with increased patient survival.

Another method of patient stratification is based on the expression of cell surface receptors (CSRs) [20]. During the malignant process there is a notable difference in expression of CSRs between cancer cells and healthy untransformed cells. CSRs traverse the plasma membrane of cells, providing an important sensory link between the signalling pathways of the cytosol and the extracellular environment. In addition, the function of many CSRs are directly involved in oncogenesis, often making these CSRs effective targets for anticancer-drugs and therapies[22]. Changes in the expression of CSRs during disease development is frequently observed. These changes in expression levels may involve gene mutations, alteration in gene copy number and/or transcriptional changes. Most CSR targets currently used in the treatment of breast and other cancers were identified through the differential expression levels of these receptors in tumours and the adjacent healthy tissues [22]. One limitation of this approach is the lack of consideration for expression levels of CSRs in tissues and organs not associated with the primary tumour. Several clinical trials aiming to evaluate the efficacy of CSR targeting therapies failed due to the dose-limiting toxicity and side-effects associated with expression of the targeted CSR on healthy tissues. It is thus necessary to implement more stringent criteria when choosing a CSR for targeted therapy. Not only should the target CSR

be overexpressed in cancer cells, but it also needs to be under-expressed both in proximal and distal healthy tissue.

Potential examples of such CSR's include CSPG4, EGFR and BCK-1. CSPG4 is a transmembrane proteoglycan expressed by stem cells and is downregulated in terminally differentiated cells [147]. It is transcriptionally associated with a variety of tissues but is overexpressed in melanoma [147]. The protein is also associated with aggressive progression and poor prognosis in several other cancer types, including basal breast cancer [93, 148]. Being a transmembrane proteoglycan, a major part of CSPG4 is expressed extracellularly. This allows for contact with a range of activating ligands, triggering clathrin-mediated endocytosis and recruitment of multiple receptors [149]. This makes CSPG4 a good target for scFv-based ITs, that require internalisation to deploy their toxin function. In addition, its abundant expression on tumour cells and comparatively low expression on normal melanocytes and breast epithelium makes CSPG4 an attractive immunotherapy target [150].

EGFR is also a well-known tumour antigen, over expressed in multiple different solid tumours including more than 50% of TNBCs, while being minimally expressed on healthy tissues [151]. EGFR-specific immunotoxins have previously shown promising results in vitro against embryonal rhabdomyosarcoma cell lines. Previously designed EGFR-specific ADCs containing SNAP-tag demonstrated specific toxicity on target-specific cell lines thus making them a good target to include in our phage libraries [151].

BCK1 is a novel target discovered through phage display [23]. Upon screening of phage libraries generated from the reactive lymph node of a breast cancer patient on membrane fractions of 8701BC (breast carcinoma), 184A1 (mammary epithelial) and L540CY (Morbus Hodgkin derived) cell lines, the authors found that BCK-1 consistently bound to 8701BC but not 184A1 and L540CY cell lines demonstrating the specificity of this clone for breast cancer cells and thus its potential as a target for cancer immunotherapy [23].

Immunotherapy

Immunotherapy utilises different parts of the immune system, including CSRs to fight diseases. This is achieved by stimulating or enhancing pre-existing immune responses in a patient or by using existing immune system components to generate therapeutics that target diseased cells without harming healthy cells.

Following chemotherapy, radiation and surgery, immunotherapy currently represents the fourth frontier in cancer treatment. The editors of Science journal deemed immunotherapy the Breakthrough of the Year in 2013, following the FDA approval of Ipilimumab, a monoclonal antibody-based therapy targeting CTLA-4 for the treatment of metastatic melanomas. Since its laboratory generation was pioneered by C. Milstein and G. Kohler in 1975, such antibodies have completely changed medicine and immunology, as well as other disciplines and is the most utilised tool in immunotherapies to date [24]. The main reason for this is the highly specific selectivity of an antibody for its cognate antigen. This selectivity of antibodies allows

them to bind to a specific target expressed on a pathogen/diseased cell and elicit an array of effector functions without harming healthy cells that lack expression of the same target antigen.

Monoclonal Antibodies

Antibody effector functions include; (1) neutralizing/blocking soluble elements such as growth factors and cytokines by binding to them thereby preventing exhibition of their effects (2) opsonizing diseased cells and pathogens thus alerting the immune system to their presences (3) activation of signalling cascades that result in cell death through apoptosis (4) prevention of ligand-receptor interaction by binding to one of the interacting molecules (5) initiating antibody-dependant cellular cytotoxicity via Fc receptor recruitment of immune effectors like NK cells and T-cells and (6) activation of the complement cascade which leads to the formation of a membrane attack complex that causes pore formation on the surface of cell/pathogen membranes eventually resulting in cell and pathogen death [25].

Antibodies have many advantages over other forms of therapeutic treatments. They belong to a well-established drug class with high success rates from first use in humans to clinical approval [26]. A lot of the information gathered in generating and optimising one antibody is applicable to other antibodies thus streamlining the process of development[27]. Additionally, being commonly found in the human body, antibodies are very well tolerated and their clinical potential can be easily increased by enhancing their already existing properties [27]. The expenses associated with antibody development and productions, like that of drug development in general, highlight the need to readdress the entire process of drug development, not just antibody development. One of the keys way of doing this would be to increase antibody and drug development research at an academic level rather than relying largely on pharmaceutical industries [28, 29]. As with all drugs, side effects are inevitable and can be severe but these events are often rare in antibody therapeutics and increasing knowledge gained by antibody research will assist in reducing adverse reactions [28]. The development process of antibody drugs has become an iterative process. Most of the processes involved in antibody design is based on the immunoglobulin gamma IgG format as this is the format of most clinically approved antibody drugs [26]. However many other antibody formats are also clinically important as they are commonly used in engineering of antibody properties [30].

Antibodies have a distinct structure, being composed of two identical heavy chains and two identical light chains which are brought together by disulphide bonds. The heavy chain is subdivided into 4 parts, 3 constant regions and one variable region while light chains have one constant and one variable region. Within the variable regions of both heavy and light chains are areas of hypervariability, commonly referred to as complementarity determining regions (CDRs). The CDRs are what give each individual antibody its unique antigen specificity and are thus the part of the antibody which is usually exploited in the development of various immunotherapies. The CDRs are located at the N-termini (Fv) of the Fab domain. Since both

Fabs of an antibody have identical heavy and light chains they also have identical antigen binding sites. The Fc region of an antibody interacts with a variety of Fc-receptors expressed on different adaptive and immune cells, dictating their effector functions [31].

The binding region of antibodies have been isolated and modified to form a variety of different antibody-based formats. These recombinant antibody fragments have many advantages over full-length antibodies making them a promising alternative. They are more easily produced leading to faster cultivation and higher yield at lower costs. They also have the potential to overcome other disadvantages commonly associated with full length antibodies, such as Fc-mediated bystander activation, trouble in binding of cryptic epitopes and poor thermodynamic stability[32]. Many of these formats have already been approved for the treatment of various cancers, autoimmune diseases and even infectious disease[33].

The naturally occurring Fab fragment is the most extensively studied format to date. This format is generated from a full-length IgG or IgM molecule, and consists of the V_H, C_{H1} and V_L, C_L regions which are linked together by an intramolecular disulphide bond. Fabs have the same binding specificity as full-length antibodies but due to their lack of Fc domain, they are unable to activate other immune cells. The single-chain variable fragment (scFv) is one of the smallest binding fragments derived from fully length human antibodies. This format consists of one V_H and V_L domain fused together by a flexible linker. The small size of an scFv allows for enhanced cellular penetration which may be advantageous with certain treatments. Moreover, it is simpler to identify and produce than full length antibodies, especially with the use of modern phage display and *E. coli* fermentation techniques. Recently the scFv format was further modified to form disulfide-stabilized variable fragments (dsFv's). The dsFv is created by connecting the V_H and V_L by an interdomain disulphide bond. These disulphide bonds stabilize the dsFv and prevent aggregation issues frequently experienced with scFvs

[34]. Various other formats exist, some of the most relevant formats are summarised in figure 2 below.

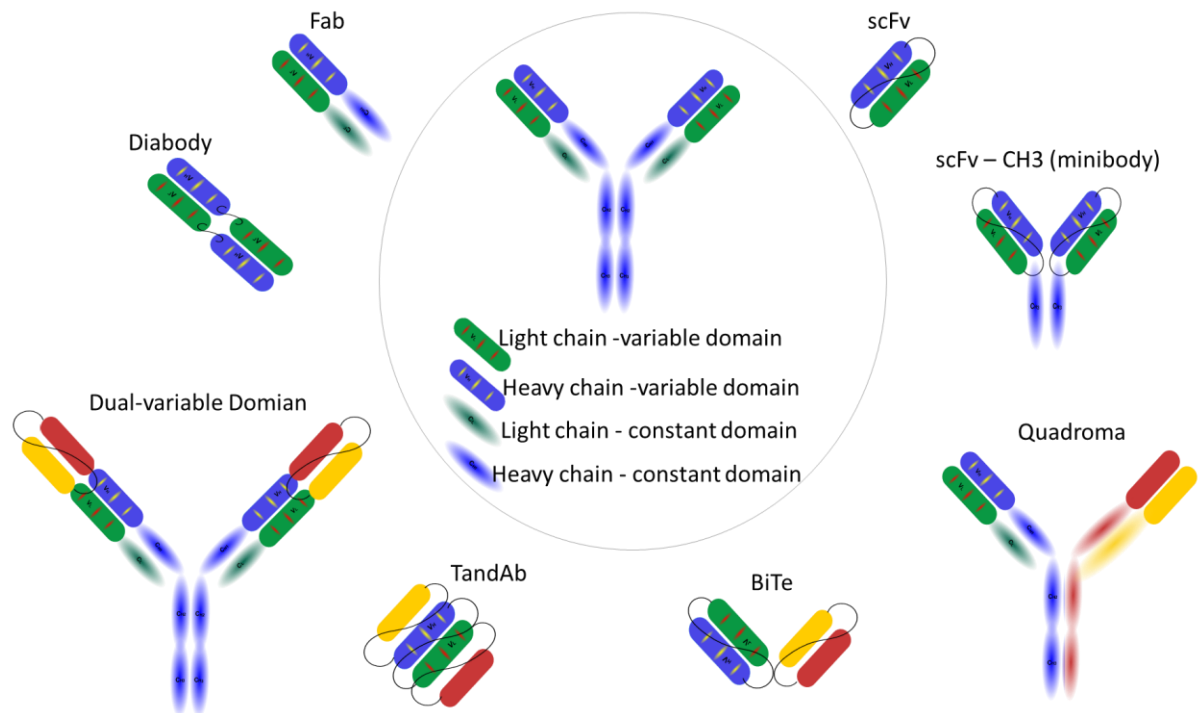


Figure 1.2: Graphically illustration of the diverse recombinant antibody formats that exist compared to a full length, classical antibody.

Section 3: Antibody-based immunotherapies

The aim of ligand-targeted therapeutics is to eliminate/neutralize pathogens and/or malignant cells. The complexity of interactions between antibody/ligands and their receptors makes choosing a target one of the most crucial steps in developing effective targeted therapies. Various considerations should be accounted for, these include the expression density of the target receptor/antigen, the affinity and avidity of binding of the target, the rate of receptor mediated internalization of the target and the format of antibody domain used. In general, the higher the expression of targets on a target cell, the more accurate the selectivity of the therapy will be. Targets should also be well conserved, as higher levels of heterogeneity may result in increased off-target effects. Additionally, it is important to avoid targets which are shed and exist in a cell free state as these cell free targets may compete with cell-bound targets thus reducing the effectiveness of the therapy. The first antibody to be FDA approved was a CD20 specific antibody used for the treatment of non-Hodgkin's lymphoma in 1997. Another antibody worth mentioning is trastuzumab (Herceptin) which binds to a growth factor antigen found on specific breast cancer cells. Trastuzumab prevents the growth and division of these breast cancer cells inevitably leading to cellular death [35].

Antibody conjugates

In addition to utilizing full length monoclonal antibodies for therapies, various antibody formats consisting of an antibody binding fragment have been utilized to direct effector molecules to a specific target. This is done by conjugating the binding fragment of an antibody, such as an scFv or Fab fragment to molecules such as radionuclides (radioimmunotherapy/RITs), cytotoxic drugs (antibody-drug conjugates/ADCs), toxins (immunotoxins) and enzymes (antibody-directed enzyme prodrug therapy/ADEPT) [36].

One of the first forms of targeted therapy was ^{131}I , a β -particle that emits small quantities of γ -radiation[36, 37]. The particle, being highly specific for elemental iodine in the thyroid, was used for treatment of thyroid cancer. The success of this treatment inspired scientists to look for other carrier particles, including mAbs/antibody fragments that could direct radioisotopes to diseased cells. The first RITs generated, and FDA approved were ^{90}Y -ibritumomab and ^{131}I -tositumomab, both specific for CD20 and used for the treatment of lymphoma. One of the main prerequisites for an RIT to be effective is surface expression of the target molecules. The antibody fragment of an RIT binds to the target on infected cells bringing the radioisotope into close enough proximity to the cell to allow for radiation induced killing [37]. While conjugation to an antibody binding domain has significantly improved the specificity of RITs, there are still many issues which limit the widespread application of this therapy. The most significant drawback being that radioisotopes are in a continuous state of decay and are thus capable of harming healthy cells and tissues when in circulation. Furthermore, organs such as the kidneys and liver are affected more drastically by the radiation due to their role in

clearance of the RIT. Additionally the actual manufacturing of RITs is complex and requires experienced nuclear medicine experts due to the serious risks of radiation poisoning[38].

Utilizing a different payload, such as a cytotoxic drug rather than a radioisotope has been highly advantageous in avoiding the many issues related to radiation poisoning. This is the case with antibody-drug conjugates (ADCs) which combine the cytotoxic potential of a drug with the specificity of an mAb. Unlike RITs which only require surface expression of their targets for effective therapy, ADCs require internalizing targets to be effective [39]. Essential an ADC functions as depicted in figure 1.3 below.

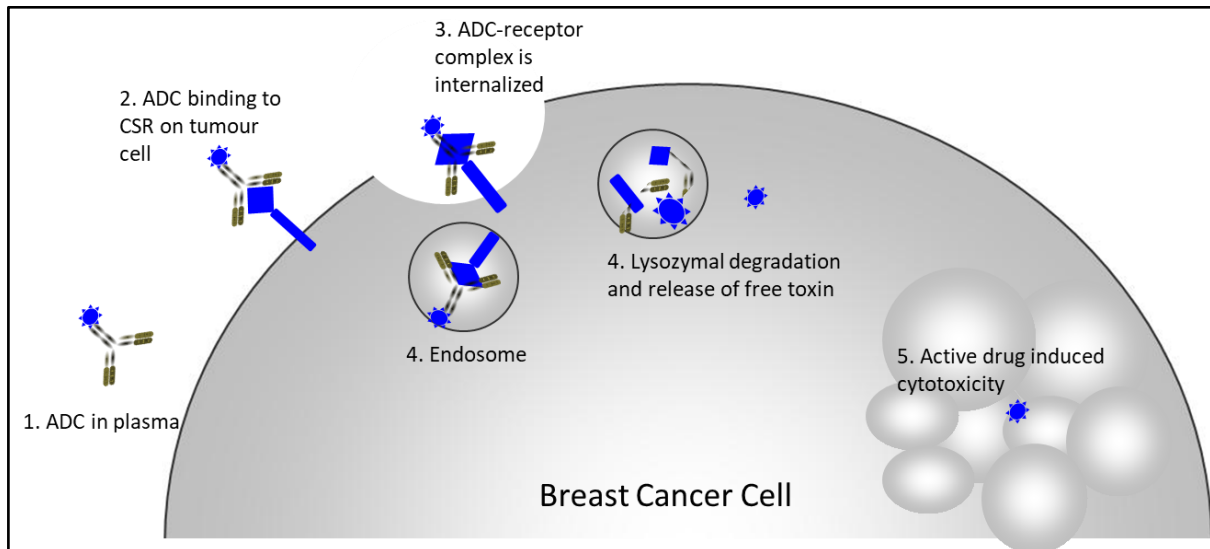


Figure 1.3: Antibody-drug conjugate mechanism of action. 1) ADC needs to reach high enough plasma concentrations to reach cancer cells. 2) The ADC needs to bind its target cell-surface receptor on the tumour cell. 3) Binding to the cell surface receptor initiates internalization of the ADC into endosomes. 4) Endosomes mature into lysosomes with capable of degrading the ADC-receptor complex into its component parts and releases the free active cytotoxin. 5) The cytotoxin initiates by DNA strand breakage or microtubule destruction leading to apoptosis of the cell.

The antibody domain binds to a surface receptor on an infected cell. Once bound the ADC is taken up into the cell via receptor mediated endocytosis. Inside the infected cell, the drug is cleaved from the antibody domain activating the drug. The activated drug then elicits its cytotoxic capacity killing the target cell from the inside. Since, the drug only enters cells expressing its target, non-infected healthy cells remain unaffected by the treatment. There are currently two FDA approved chemotherapeutic ADCs; a HER2 targeting conjugate that inhibits microtubule assembly and a CD30 targeting ADC conjugated to a antimitotic drug (NCT00848926)(NCT00866047) [37, 40].

In addition to targets expressed on infected cells, some antibody-conjugates target pathogens directly. Thiomab, an antibody-antibiotic conjugate, directly targets a cell wall acid residue expressed on *S. aureus* [41]. Thiomab functions in one of two ways; it initiates killing of the infected cells through the intracellular release of its antibiotic dmDNA31 and targets extracellular bacteria through rapid opsonization. The final antibody conjugate that is of importance in precision medicine currently are immunotoxins (ITs). As the name would

suggest ITs are composed of a protein toxin that is conjugated to an antibody. Similarly, to RITs and ADCs, it also requires surface expression of their targets to be effective. Unlike other conjugates however, ITs targets have additional requirements, they need to undergo endocytosis and intracellular trafficking to the cytosol to exhibit its function. Initially ITs were generated from bacterial and plant-based toxins, however due to the immunogenicity of these molecules, most recent ITs utilize human apoptosis-inducing enzymes such as granzyme B, Map-tau and angiogenin [37].

In addition to antibody conjugates, one of the most promising of all mAb based therapies currently used, are checkpoint inhibitors, which are the main component of several drugs that have already been approved for treatment of over 9 types of cancers.

Checkpoint Inhibitors

Using the combined potentials of CD4⁺ T-cells, CD8⁺ T-cells and antibodies the adaptive immune system can inflict major damage to harmful microbes. However, all immune responses require balance to ensure that healthy cells and tissues are unharmed and that self-tolerance is maintained [42]. In a balanced immune response, immune activation is counteracted with immunoregulatory pathways such as checkpoint inhibition. These negative regulatory pathways are necessary for maintaining tolerance and preventing hyperactivation, and are governed by cell surface, inhibitory receptors known as “checkpoint inhibitors”.

Cytotoxic T-lymphocyte antigen (CTLA-4) and B7.1/B7.2

In T-cells when the TCR engages with MHC-peptide complexes leading to positive co-stimulatory signals between CD28 on T-cells and B7 ligands (CD80 and CD86) on APCs, multiple negative pathways are also initiated. Cytotoxic T-lymphocyte antigen 4 (CTLA-4) is the first inhibitory regulator to be induced [42]. CTLA-4 competes with CD28, for binding of target ligands CD80 (B7.1) and CD86 (B7.2) expressed on the surface of antigen presenting cells (APCs). CTLA-4 and CD28 receptors have similar structural and biochemical properties [43, 44], they are found on the same region of chromosome 2 and are selectively expressed in the haemopoietic compartment. However, while CD28 is expressed in high basal levels, CTLA-4 has much lower basal expression but is rapidly induced upon antigen activation. Structurally, both receptors comprise an extracellular immunoglobulin like domain, a transmembrane domain and a cytoplasmic tail, together forming a membrane bound homodimer which recruits signalling proteins and controls surface expression [44-46]. These similarities allow both receptors to bind the same ligands. The major difference between the two is that CTLA-4 has a much higher affinity and avidity for the B7 ligands than CD28 [47, 48]. CTLA-4 therefore out-competes CD28 when expressed in high levels, inhibiting CD28 based T-cell activation and proliferation. In addition, CTLA-4 has a negative modulatory role in helper T-cell activity and is necessary for anti-inflammatory cytokine release by T-regs. Furthermore, CTLA-4⁺ T-regs induce anergy of conventional T-cells by binding to B7 ligands on dendritic cells, initiating trans-endocytosis, a process of internalizing and degrading the ligands [49-51].

PD1 and PD-L1

Similarly to CTLA-4, programmed cell death protein 1 (PD1) expression is initiated during T-cell activation [42]. PD1 counteracts positive T-cell activation signalling by binding to one of its cognate ligands ; programmed death 1 ligand 1 (PD-L1) and/or programmed death 1 ligand 2 (PDL2) [42]. PD1 has become the model for understanding the various functions of inhibitory receptors. PD1 pathway signalling contributes to the regulation of T-cell activation, T-cell tolerance, fine-tuning T-cell fate and function, and return to immune homeostasis. In mice studies, animals deficient in *Pdcd1* (the gene encoding PD1) develop accelerated autoimmunity. Contrastingly, sustained high expression of PD1 and its ligands are found during chronic infection and cancer [42]. The expression of PD1 and its ligands are best understood for T-cells. PD1 is commonly associated with exhausted T-cells and tumour immunosuppression. However, PD1 is not an exhaustion-specific receptor, it is expressed on all T-cells during activation and is thus more of an effector T-cell marker than an exhaustion marker [42]. In a healthy host, if a T-cell activating antigen is acutely cleared PD1 expression decreases, however if the antigen persists, in cases like chronic infection and cancer then PD1 expression is sustained at high levels .

PD1 and its cognate ligand, PD-L1 are both type 1 transmembrane glycoproteins which are part of the immunoglobulin superfamily [52]. Binding of PD1 to PD-L1 induces a conformational change in PD-1 leading to an Src family kinase driven phosphorylation of immunoreceptor tyrosine-based switch motif (ITSM) and immunoreceptor tyrosine-based inhibitory motif (ITIM) . Once phosphorylated, these tyrosine motifs recruit SHP-1 and SHP-2 protein tyrosine phosphates which attenuate T-cell activating signals [52]. SHP-2 dephosphorylates CD28, interfering with its binding and signalling capacity. PD-L1 also interacts with CD80, possibly delivering inhibitory signals to activated T-cells. The engagement of PD-1 and PD-L1 has many other effects on T-cells, inhibiting; T-cell proliferation, cytokine expression, survival, and other effector functions. While PD-1 checkpoint signalling is quite well understood, this is not the case for “reverse signalling” through PD-L1 [52]. However, in the case of cancer, evidence suggests that PD-1 induced stimulation of PD-L1 promotes pro-survival signalling to tumour cells, leading to resistance to Fas and staurosporine-induced apoptosis. Moreover, PD-L1 protects cancer cells from type 1 and type 2 interferon induced cytotoxicity and CTL mediated cytolysis, without any signalling requirement from PD1. While the intercellular signalling factors that participate in PD-L1 induced apoptosis are not well understood, it has been shown that the cytoplasmic domain of the PD-L1 is critical for this signalling [52].

Clinical Applications

Blocking of checkpoint pathways during chronic infections and cancer has been shown to improve T-cell functions leading to reduced viral load and tumour burden [42]. These findings have been translated into clinical application where checkpoint inhibitors, which are monoclonal antibodies targeting CTLA-4, PD1, PD-L1 or other inhibitory ligands, have been used to block these inhibitory interactions. The first of these therapies to demonstrate clinical success was Ipilimumab, an anti-CTLA4 mAb used for treatment of melanoma [42]. Furthermore, the initial trials with PD1/PD-L1 based checkpoint inhibitors showed promising results which lead to the US food and drug administrative (FDA) approval of multiple monoclonal antibodies including PD1 specific nivolumab and pembrolizumab, and PD-L1 specific atezolizumab, avelumab and duravalumab for therapeutic applications in various cancers [42].

Despite the promising clinical results, most patients do not respond to checkpoint inhibition therapy. As the therapy targets ligand/receptor interactions, expression of these targets is an important prerequisite for effective treatment [37]. In the case of cancer, more than 50% of cancers induce a suppressive tumour microenvironment which lacks effector cells and thus have insufficient targets for the therapy to be effective. In chronic infection checkpoint therapy has had contrasting results which suggest that the mechanism of T-cell dysfunction in chronic infection and cancer are intrinsically different [37]. CTLA4 checkpoint therapy in chronically infected LCMV mice did not improve T-cell function or viral load. Moreover, CTLA4 blockade in SIV infected macaques amplified viral replication at mucosal sites and did not enhance T-cell responses. Contrastingly, *in vitro* studies found that CTLA4 inhibitor therapy improve HIV-specific CD4 responses while PD1/PD-L1 blockade enhances HIV-specific CD8 T-cell function as well HCV and HBV-specific CD8-T-cell responses. Combination therapy with CTLA4 and PD1 blockade rescued exhausted HCV-specific CD8 T-cells [37]. In other HIV studies, Ipilimumab (CTLA4 specific mAb) induced the reactivation of latently infected T-cells, rendering them perceptible to the immune system. Furthermore, phase I trials displayed enhanced HIV-specific CD8 T-cell responses following PD-L1 inhibitor therapy. Cumulatively, these studies suggest that checkpoints play a role in inhibiting T-cell function during HIV infection and thus highlight the potential benefit of checkpoint inhibition therapy in this setting. Similar findings were observed with other viral infections as well. Nivolumab, an anti-PD1 blockade therapy significantly reduced viral load in HCV-infected patients [37]. However, the data suggested that pre-existing HCV-specific T-cells need to be present in the liver for treatment to be effective. While there clearly still many issues that need to be addressed , the overwhelming positive results associated with CTLA4 and PD1/PD-L1 therapy has revolutionized cancer treatment and effectively demonstrated its potential to improve treatment outcomes in other diseases as well.

Antibody based immunotherapy of TNBC

The elevated levels of PD-L1 expression on the surface on TNBC cells has functional consequences on infiltrating T-cells resulting in increased T-cell apoptosis and reduced T-cell proliferation, which provides a rationale for using PD-L1 as an immunotherapeutic target [53]. Recently, the glycosylation of PD-L1 was found to be essential for PD1/PD-L1 interactions leading to the development of a modified glycosylated-PD-L1 (gPD-L1) targeting antibody, STM108 which induces PD-L1 lysosomal internalization and degradation [40].

Despite its initial clinical promise, α PD-L1 and α PD1 have yielded mostly unsatisfactory results when used as single therapeutic agents. The expression of these markers may however serve as a biomarker for clinical benefit, as tumours lacking PD-L1 or PD1 expression would be less sensitive to α PD-L1 and α PD1 therapies. Furthermore, high expression of either PD-L1 or PD1 was correlated with increased numbers of infiltrating FoxP3+ T-regs which may work synergistically with PD-L1/PD1 to implement immune evasion [54, 55]. It is still not known for certain if α PD-L1/ α PD1 therapy could affect T-reg function but given that PD-L1 is expressed by tumour cells while PD1 is expressed by immune cells, PD-L1 may be the preferred target of immunotherapies. Unfortunately, not all PD-L1 positive tumours respond to α PD-L1 therapy and some tumours lacking PD-L1, such as certain lung cancers do respond to these therapies [56]. Thus, the efficacy of PD-L1 as a target for immunotherapy in TNBC remains to be proved. There are currently three ongoing phase III clinical trials assessing the validity of these PD-L1 as a target for TNBC immunotherapy.

In TNBC, CTLA-4 antibody mediates anti-tumour immunity through blockade of T-regs in the TME via Akt phosphorylation, resulting in potent T-cell expansion [57, 58]. According to a systematic review, TNBCs encompass the highest incidences of TILs (20% ;range of 4-37%) and FoxP3+ T-regs (70%; range 65-76%) among breast cancers [59]. This elevated level of FoxP3+ CD4+ T-cells, may act as therapeutic targets for CTLA-4 blockade in TNBCs [57]. There is only limited pre-clinical data available, demonstrating that α CTLA-4 therapy has not significant impact on tumour growth unless it is accompanied by Treg depletion [60]. There are currently two ongoing clinical trials, evaluating Ipilimumab and Tremelimumab (α -CTLA-4 therapies) as treatments for TNBCs. However, much more extensive investigations are warranted to determine the relationship between Treg infiltration and treatment responses in TNBCs.

While single checkpoint blockade therapy has shown only limited benefit, combination therapies utilizing checkpoint inhibitors together with targeted therapies have demonstrated increased efficacy of treatment, slowing down primary tumour outgrowth and reducing metastasis especially when neoadjuvants are present. Multiple syngeneic TNBC models also demonstrated that dual therapy could improve tumour-specific lymphocyte responses [61]. There are currently no conserved TSAs for targeted therapy in TNBCs, mainly due to the large

heterogeneity of the disease. Discovery and development of new targets is thus an essential next step in the generation of TNBC immunotherapies.

Section 4: Antibody Discovery and applications.

Traditional methods of antibody isolation.

The initial and most widely used technology for generating monoclonal antibodies were mouse hybridomas. These are cell lines generated by stable fusions of immortalised myeloma cells with B-cells from immunized mice [62]. While mice are the predominant host species used for generation of antibodies, many other animals including rabbits, rats, cows and chicken have also been immunized to generate antibodies. The major issue with murine antibodies for human use is that they are highly immunogenic and have weak interactions with human complement and Fc gamma receptors FcγRs, resulting in inefficient effector functions [26]. They also have a terminal half-life of less than 20 hours due to a lack of binding to the human salvage receptor [63]. Many of these limitations have been surpassed by chimerization and humanisation. Chimerization is the joining of the variable domain of a mouse antibody to a constant domain of a human antibody, thus increasing the efficiency of effector functions [64]. Humanization involves the transfer of CDR's (Complementarity determining regions – these are the antigen binding loops within the variable region of an antibody) from mouse antibodies to a human IgG [64]. This permits high antigen binding within the framework of human antibody thus having efficient effector functions and reduced immunogenicity. To further improve the functions of an antibody, a completely human antibody would be necessary. To this end, most antibodies currently entering clinical trials are fully human and directly derived from human B lymphocytes and/or from a technique called antibody phage-display.

Phage Display

Phage display is a selection technique based on the presentation of ligands (antibodies/peptides/proteins) on the surface of bacteriophages [65]. DNA encoding the ligands of interest is cloned into the phage genome as a fusion with one of the phage capsid proteins [66]. Depending on the selection criteria different capsid proteins can be used for display, most commonly used are pIII and pVIII [65]. There are 5 copies of pIII and 2700 copies of pVIII on each mature phage, therefore large ligands with high affinity binding are usually expressed on pIII while smaller ligands are expressed on pVIII providing high avidity [65]. Once cloned into the phage genome, the coat protein fusion is incorporated into new phage

particles and assembled in the bacterium [66]. Expression of the fusion product on a phage surface provides the physical link of phenotype with genotype.

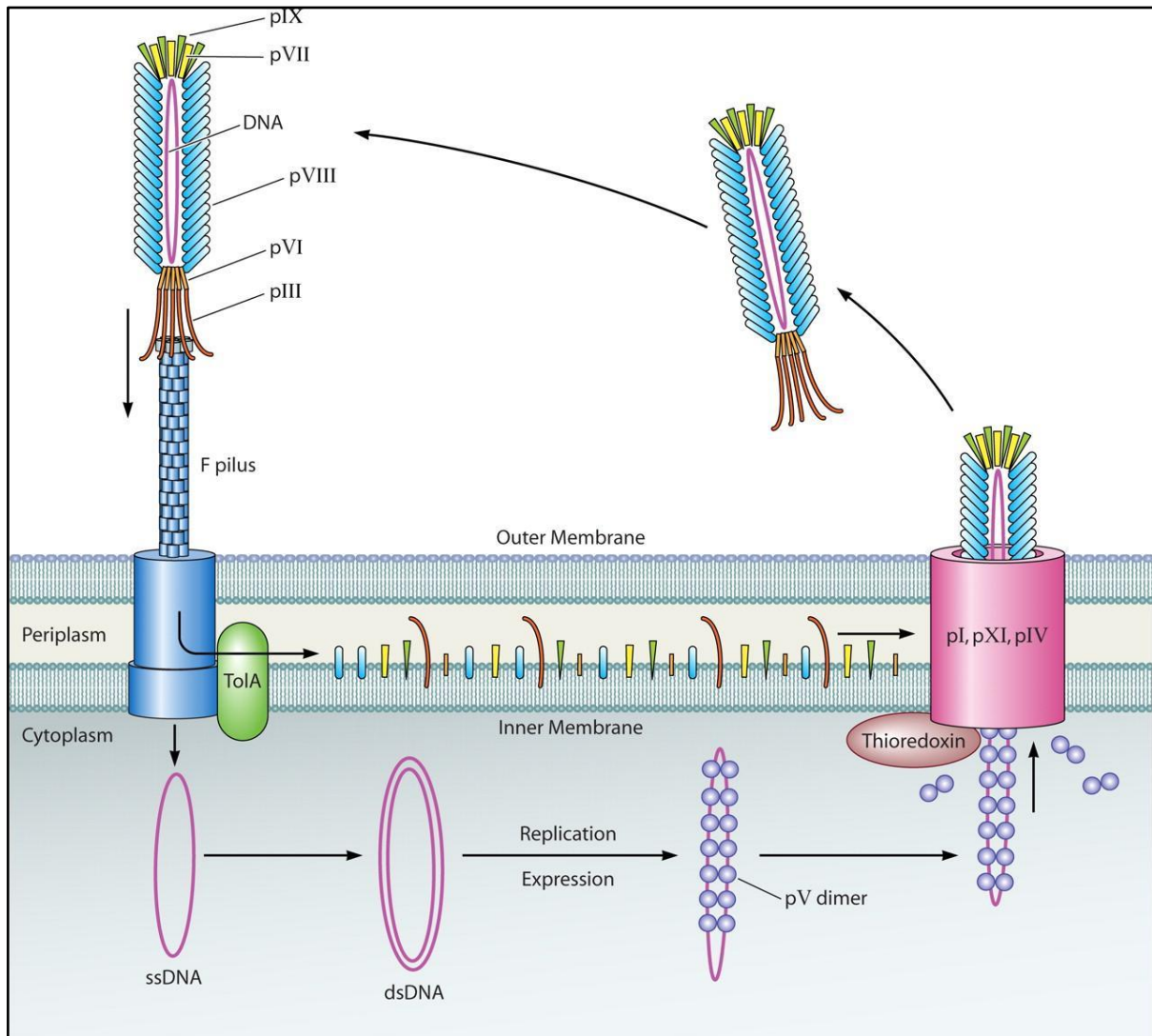


Figure 1.4: Schematic representation of M13 filamentous phage life cycle. The N-terminus of pIII on the phage binds to the tip of the *E. coli* F'-pilus. The phage proteins are subsequently removed by the host protein TolA and the single stranded phage DNA (ssDNA) is converted to a double-stranded (dsDNA), replicative form. Following replication and expression of phage coat proteins by the host enzymes, ssDNA molecules are coated by pV protein dimers to prevent their conversion back to dsDNA. The proteins pI, pXI and pIV work together to form a channel to facilitate phage secretion. Protein pV is replaced with pVIII within this channel and then mature phage particles are assembled and released. Thioredoxin facilitates the removal of pV. Figure obtained from [67].

The most popular vectors for display are nonlytic, filamentous bacteriophages such as fd and M13 phages which infect *E. coli* strains containing the F conjugative plasmid [68]. The infection process involves attachment of phage capsid proteins to protruding bacterial F-pilus [65]. The pilus retracts into the bacterial cell membrane allowing translocation of the ssDNA phage genome into the bacterial cytoplasm. Bacterial enzymes convert the single stranded phage DNA into a replicative dsDNA form. With the assistance of phage proteins pII and pX, multiple copies of the phage proteins are replicated. All replicated phage proteins then undergo transport to the bacterial periplasm, where assembly and extrusion occur. Once released, phages displaying antibodies of interest can be tested for specificity by binding to an

immobilised target antigen [66]. Non-adherent phages are washed away and only specifically bound phages remain and can be used for downstream applications.

The first molecule to be successfully displayed on a phage was a single chain variable fragment (scFv) of an antibody [69]. The scFv was cloned with gene III downstream of gene III signal sequence which normally directs export of proteins [70]. Within the bacterial periplasm, the variable heavy (V_H) and variable light [71], domains fold and pair to form a functional scFv [72]. Initially, phage vectors containing all the genetic material necessary for the phage life cycle were used for phage display [73]. In recent years phagemids have become a more popular vector for display.

Phagemids have a truncated genome containing the replication origins of the phage, an antibiotic resistance gene, the phage gene used for display and a packaging signal [65]. Being smaller than complete phages, phagemids have a much higher transformation efficiency making them suitable for generating large repertoires [66]. Phagemids require helper phages such as R408, VCSM13 and M13KO7 to complete the life cycle of a phage [74, 75]. Similarly to phages, phagemids infect bacteria and undergo DNA replication in a similar manner to plasmids used in recombinant DNA research [74]. Upon completion of transcription and translation, peptidase removes the leader signal peptide and the fusion proteins are transported to the inner membrane of host cells. Helper phages co-infect host cells and provide replication proteins which are missing from phagemids. Helper phages have a defect in their origin of replication thus phagemid DNA is preferentially replicated [75]. In addition to having a higher transformation efficiency, phagemid vectors are also able to accommodate larger foreign DNA fragments, the expression levels of fusion proteins can easily be modulated and controlled, and phagemids are genetically more stable under multiple rounds of propagation than recombinant phages [74].

Screening on phage libraries

Multiple phage libraries are already market approved. Most of these libraries were generated using lymphoid B-cells from immunised and non-immunised donors. One of the major uses of a phage-display library is to screen for novel antibodies against tumour surface antigens. To this end multiple cancer specific antibodies have been isolated from different libraries by panning on purified antigens [76, 77]. There are 3 major aspects to consider when setting up a phage library for the identification of mAbs; 1) the type of library been generated and the source of B-cells that will be used, 2) the selection strategy that will be used during biopanning procedures and 3) the possible need for post-selection phage derived mAb optimisation [78].

The first important step in setting up a phage-display library is to select the source of B-cells that mRNA will be extracted from for subsequent amplification of antibody variable genes. To date, no significant studies have been conducted that compare different B-cell sources in different conditions [79]. Major naïve libraries have been constructed from human bone marrow [80-83], peripheral blood [84-86] and lymphoid tissues [87-89], murine spleen and peripheral blood from other animals [90-92]. Choosing an appropriate source of B-cells is based on the research question at hand. In this study we wish to identify mAbs that will be used for clinical practice, therefore human derived B-cells will be used. In addition to the source of B-cells to be used the format of the mAb required needs careful consideration. Monovalent Fab fragments and single chain Fv fragments (scFv) are the most common formats used in current phage libraries. Fab's constitute 2 chains; a VH1 and CH1 domain of the heavy chain and an entire light chain. scFv's are smaller and constitute the variable regions of both heavy and light chains. The smaller size of the scFv's permit better cellular penetration and internalisation which is essential to the down-stream functioning of our therapeutic applications and is thus the format we chose to use [78].

Once the library is established, an appropriate panning strategy is required. The panning procedure utilises the interaction between specific antibodies presented on bacteriophages and an immobilised target, usually a coat on a microtiter plate. Once bound plates are washed, getting rid of all unbound, non-specific phages. In this study, tumour tissue sections or solubilized tumor cells will be used as the immobilised target, allowing for the capture of tumour specific antibody-displaying phages. The bound phages are then eluted through changing pH and may be used to re-infect bacteria to improve the selection. Multiple rounds of panning can be done to select for strong affinity mAb's. In addition to the positive selection panning, negative selection can also be done to get rid of cross-reactive phages. This is done by selecting phages on healthy tissues, in this instant all unbound phages will subsequently be used for multiple rounds of positive selection [78].

Finally, once specific antibodies are selected, they need to be characterised and modified according to their specific therapeutic intent. With the growing body of knowledge on antibody targets and sequencing being generated through phage display and other technologies, it is now possible to find relevant targets not just through wet lab work but by

simply screening different online platforms. By using sequences already discovered saves time and money and would allow for a much more rapid development of antibody-based therapeutics.

Alternative methods of antibody discovery

The efficiency of generating an antibody based therapeutic agent is critically dependent on selecting an appropriate therapeutic target and binding antibody. The study and acquired knowledge on existing targets and antibodies are very helpful in accelerating drug discovery and development. There are various open access tools and databases that provide comprehensive and complementary data of potential therapeutic targets, especially for antibodies [93].

These tools assist in elucidating structural information that may be necessary to rational antibody design. Various numerical tools including; PylgClassify, DomainGapAlign, AbNum and more, are used to annotate antibody sequences to identify the framework and CDR regions [94]. There are also several alignment databases like IgBlast and SAbDab which allow you to align antibody sequences to various VH and VL domains of different immunoglobulins. And some of the most advanced tools available are tools such as PIGS and FREAd which are able to predict the 3D structure of a CDR loop [94].

The integration of these tools, together with online databases comprising knowledge of genomics, genetics, transcriptomics and literature or patents provide an excellent basis for target identification. Selected targets may already be validated with clinical evidence in patents and literature reports. This is particularly true for targets that have already been approved as therapeutic applications, targets that are currently in clinical trials with a potential targeting drug but no approval and those that have recorded patents and publications.

Patents as a source of antibody sequences and information

Therapeutic antibodies on the market and in late-stage clinical trials have formerly undergone studies by computational and/or experimental methods to identify bio-therapeutic properties which make it a successful therapeutic application. Clinical stage therapeutics are the high-quality end-point results of a lengthy procedure of choosing a molecule from several possible candidates. Thus, single therapeutic molecules only moderately exemplify the complete engineering process of antibody-drug development. Complete public disclosure of processes involved in creating therapeutic antibodies including, intermediary sequences and particulars of selection choices are not commonly shared because of the commercial value of such knowledge which requires legal protection. Legal protection of this knowledge is obtained by patents. Through keywords and patent classification searches one can broadly discern between patents that are based on antibody techniques and novel antibody molecules. Those patents based on novel molecules are of particular interest as they have the potential to find their way into clinical application. However, the antibody sequence and targeting information in such patents is not easily extracted as the purpose of patents is not to convey scientific know-how but rather provide legal protection.

A study published early last year assessed how useful patent data was in antibody drug design [95]. The study quantified the amount of antibody sequences found in patents that were

intended for medical applications and assessed how well these antibody sequences matched those already in clinical application. They found 16526 patent families, reflecting 245109 unique antibody chain sequences. In total antibody sequences made up a non-trivial portion of total amino acid sequences available, up to 11% in USPTO database [95]. The results of this study demonstrated that the volume of patents containing antibody sequences is rapidly growing and they very well reflect the sequences of antibodies used in medical applications. This study provides a solid rationale for the use of patents in rational antibody-based drug design.

While there is a wealth of knowledge to be found in patents, it is not without its caveats. Most patents have sequence depositions which are non-standardized, and many propose wild claims on sequence identity which can obfuscate the extracted clinical sequence [95]. However, once you can understand how to interpret and extract the information required from a patent it can be used to obtain critical knowledge on engineering choices and design that would dramatically accelerate development of clinical biotherapeutics. This thesis utilized a variety of patents to obtain target information and antibody sequences for the design of various fusion proteins.

IgBlast:

IgBlast is an immunoglobulin variable domain sequence analysis tool. The antigen binding domain or variable domain of an antibody is encoded by multiple genes, these include V (variable), D (diversity) and J (joining) genes [96]. These genes are separately encoded within the germline genome but subsequently undergo a process known as VDJ rearrangement in early B-cells. There are multiple variations of each gene type and any of these variations can rearrange with each other resulting in an enormous variety of V domains [96]. In addition, many post gene-rearrangement modifications further enhance the diversity of v-domains these include; imprecise joining of rearranged genes, trimming or adding nucleotides, random heavy chain -light chain pairing and somatic hypermutation making the diversity of immunoglobulins almost unlimited [96].

This makes analysing antibody sequences particularly complex. In many cases a detailed analysis of the antibody sequence including the location of framework and CDR regions is very important to downstream application development. While there are many tools and databases out there, such as the well-known BLAST, many of these programs have only a limited capacity to analyze antibody sequences. The different genes which contribute to a v-domain also each have varying characteristic lengths ranging from 10 bases to 290 bases [96]. Applications such as BLAST require special parameters to identify short matches, but these parameters are not compatible for long matches thus multiple searches would need to be done in BLAST to analyze a single antibody. This would also require a manual assembly of the sequence post blasting which can be very complex and error prone. There are also a number of antibody sequence analysis tools but none of them work as well as IgBlast [96]. IgBlast can search a query sequence against multiple databases at the same time, showing germline gene

matches as well as sequence matches to other rearranged genes. The program also accurately delineates the framework regions and CDRs. Furthermore, it contains other analysis tools which allow you to identify if a sequence is in-frame or out of frame. In-frame sequences are those in which the last coding triple of v-genes is in-frame with the first coding triple of a sequence. Rearranged sequences are only productive when in-frame and lack stop codons. Overall IgBlast is a robust antibody sequence analysis tool. It is also freely available and very user friendly. All antibody sequences used for the design of fusion proteins in this thesis were analysed with IgBlast.

Antibody fusion proteins

Fusion proteins are a type of protein that integrates two or more domains from different proteins into one molecule [97]. Various naturally occurring proteins exist, each with a unique architecture to meet different functional requirements of living organisms at a molecular level. With the use of recombinant DNA technology and different post-translational modification strategies researchers have been able to create synthetic fusion proteins with multiple applications.

Fusion Protein Domains

For the creation of a successful fusion protein, it is important that all fusion domains be physio-chemically compatible in terms of their functional pH, ionic strength, temperature etc [97]. Another key consideration is the order in which domains are fused as the location of one domain can affect the functionality of another. In the case of combining two domains, it may be necessary to try out both possible orientations of fusion and select the one which functions best unless the effect of the domain order is already understood [97]. One of the simplest ways of combining fusion partners is through genetic fusion, where the genes encoding both fusion domains are joined and expressed as a single protein in a suitable host. A direct fusion in this manner is the simplest approach and usually the preferred choice when the N or C terminus of the fusion partners are not necessary to the functioning of a protein and if they allow for sufficient flexibility to permit the correct 3-dimensional folding [97].

Linkers

In the case of an antibody fusion protein, a flexible linker is required to avoid steric hindrance and minimize misfolding. Naturally occurring linkers are ubiquitously found in multidomain proteins where they provide sufficient distance to reduce steric hindrance and enhance domain-domain interactions [97]. Based on these observations many naturally occurring linkers such as the immunoglobulin hinge region have been isolated and utilized in the design of synthetic fusion proteins. Additionally, three categories of artificial linkers have been developed by researchers, these include; rigid, flexible and *in vivo* cleavable linkers. Commonly utilized flexible linkers are (Gly)_n and (Gly-Gly-Gly-Gly-Ser)_n linkers whose copy numbers are adjusted based on the protein application [97]. The one disadvantage of this kind of linker is the number of homologous repeat sequences in the coding DNA sequence, which can reduce protein expression. An appropriate codon pair selection may overcome this. Furthermore, the (Gly-Gly-Gly-Gly-Ser)_n linkers play an important role in connecting the VH and VL chains of an scFv and are thus necessary for the proper assembly of the scFv itself [98]. The choice of linker is an important one as it is one of the deciding factors in a successful fusion protein construction. The length, amino acid sequence, protease sensitivity, hydrophobicity, secondary structures, and possible interaction with fusion partners all need to be considered when choosing an appropriate linker. For the purposes of this thesis a (Gly-

Gly-Gly-Gly-Ser)₃ linker was used for the construction of all fusion proteins. This linker was chosen based on previous data demonstrating its functionality in similar constructs [99].

In addition to a genetic fusion, post-translation protein conjugation is an alternative method to create fusion proteins from fusion partners that are expressed separately. This method is primarily useful when making fusion proteins from domains that originate in different host species. If protein domains from different host species are expressed in a single host the fusion protein may fail to express in sufficient yields, or lack the proper folding and modifications required for full functionality [97]. This method of post-translational coupling is also useful in a multicomponent or interchanging system, where one functional domain may require coupling to different partners to execute different functions. A good example of this is therapeutic fusion proteins, in which one fusion partner usually plays a functional role in molecular recognition while the other partner exerts other advantageous functions such as enhancing half-life and stability or demonstrates cytotoxic effects and improved pharmacodynamics. Specialized enzymes and catalysts are often useful in conferring such adaptability to an antibody binding domain. SNAP-tag is one of these specialized enzymes and is discussed in the section below as a major part of this thesis focused on the design and generation of antibody based diagnostic and therapeutic fusion proteins using SNAP-tag technology.

SNAP-tag Technology

Since the discovery of mAb's, major research has gone into designing methods to arm these antibodies or their binding fragments with cytotoxic drugs and other warheads capable of killing infected cells and microbes. Some of the most utilized methods currently involve reducing disulphide bonds to produce reactive sulfhydryl groups or direct functionalization of lysine side chains using N-hydroxysuccinimide ester derivatives [100]. Both these methods form heterogeneous products with varying safety and efficacy profiles due to the abundance of lysine and cysteine side chains found on an antibody. There are approximately 40 lysine residues alone on a standard antibody, which potentially results in more than a million possible conjugation variants [100]. Similarly, there are four interchain disulfide bonds which give rise to up to 8 reactive cysteine residues upon reduction, yielding about 100 different variants. To avoid the generation of heterogeneous protein populations with unpredictable safety and efficacy profiles, multiple site-specific conjugation methods were developed. Of all the methods currently available, self-labelling proteins like haloalkane dehalogenase, aldehyde-tag and SNAP-tag offer the most promising approach [100]. These proteins are relatively small and can be directly and specifically conjugated to a labelling substrate due to their enzymatic activity.

SNAP-tag is a modified version of a human DNA repair enzyme, known as O⁶-alkylguanine-DNA-alkyltransferase (AGT) which naturally removes alkyl residues from damaged DNA [99]. The modified version of the enzyme reacts specifically with benzylguanine (BG) derivatives via irreversible transfer of alkyl groups to cysteine residues. Under physiological conditions SNAP-tag can be fused to a protein of interest in a rapid, site-specific autocatalytic manner forming a 1:1 stoichiometry. This highly specific conjugation property of SNAP-tag has been used to couple a variety of effector molecules. SNAP-tag based fusion proteins have been used in various experimental applications, such as immobilization of proteins on chips, optical in vivo imaging and in-vivo and live cell biochemical applications [100]. Furthermore, SNAP-tag has been genetically fused to ligands which bind to specific cell-surface molecules like extracellular domains of receptors, different cytokines, growth factors, ligands, and various recombinant antibody binding fragments [100]. The fusion of SNAP-tag to any of these molecules allows for the equipping of the molecule with any BG modified substrate including, fluorescent dyes, cytotoxic drugs, other therapeutic agents and even nanocarriers. Due to its ease of use and highly specific conjugation to BG modified substrates, SNAP-tag was used for site-specific labelling of various scFv's for the generation of different diagnostic/therapeutic fusion proteins in this study. This thesis describes in detail the design and construction of scFv-SNAP-tag fusion proteins, as well as the expression of these proteins in mammalian cells, the labelling of the protein with BG modified derivatives and the applicability of such proteins in diagnostics and therapeutics.

Section 5: Applying the Lessons and Tools from Cancer Immunotherapy to Immunotherapy of infectious diseases

Immunotherapy has been developed and researched for the treatment of cancer more so than any other diseases. Subsequently, the errors, trials and solutions adopted in the cancer immunotherapy field have paved the way for immunotherapeutic application in other fields, with superior advancement in antibody-based therapies of infectious diseases. Antibody therapy of infectious disease can be traced back over 100 years ago, where serum containing antibodies was used to neutralize bacterial toxins [101]. In comparison to toxin neutralization however, using antibodies to eliminate viral diseases is far more complex. Various actions are required to eliminate a viral infection, these include inhibiting: cell infection, viral replication and release, cell-cell transmission and targeted killing of viral infected cells [102]. During an infection, a typical immune response would result in the generation of viral-specific polyclonal antibodies, some of which would be neutralizing and others not. The overall outcome would be a combination of neutralizing, blocking and eliminating antibodies, which together may indeed provide effective protection against viral disease. An anti-RSV monoclonal antibody, Palivizumab has been approved for treatment of high risk infants, and has been shown to be clinically effective prophylactically, but not therapeutically [103]. This led to the development of a new generation of highly potent and/or broadly neutralizing antibodies termed super-antibodies which offer prophylactic and therapeutic possibilities for viral infections. Several of these antibodies are under clinical evaluation for the treatment of highly antigenically variable viruses such as Influenza, Ebola and HIV [37]. In addition to the investigation of super-antibodies for the treatment of infectious diseases, various RITs and ADCs have all been researched as novel therapies for such diseases. For example, an RIT composed of a ^{213}Bi radioisotope conjugated to a gp41-specific (HIV envelope protein) monoclonal antibody was able to significantly decrease the number of HIV-infected cells in vivo [37]. In recent times, flavivirus host proteins have been exploited as targets for the development of virus-specific therapeutics, with particular focus on the development of Zika virus (ZIKV) therapies [104].

Zika virus (ZIKV) was originally isolated in 1947 from the serum of a Rhesus monkey in the Zika forest of Uganda [104, 105]. The following year the virus was found in the *Aedes Africanus* mosquitoes from the same forest [106, 107]. A mere two years later the first human infections were reported in Africa and Asia [106, 108, 109]. The virus remained in these two regions until 2007, after which it spread geographically to the North Pacific, South Pacific and French Polynesia [110-113]. In 2015, 84 countries worldwide reported cases of ZIKV infections despite having prior prevention and control measures in place. ZIKV infections are much more prevalent in areas such as Suriname, Columbia and Brazil where there is large mosquito presence and pose a serious health care issue in these areas [104, 114].

ZIKV is an enveloped flavivirus which is spherical in shape with an icosahedral symmetry [115]. It has a 10.7 kilobase positive-strand RNA genome, containing a single open reading frame that encodes a 3400 amino acid polyprotein [116]. Upon host and viral protease processing, the genome yields seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A and NS5) and

three structural proteins (C, PrM and E) which play an essential role in viral infection and replication [116]. Very little is currently understood about the pathogenesis of ZIKV as its infection cycle involves complex host and viral interactions [115]. ZIKV infection in humans may be initiated after a single bite from an already infected *Aedes* mosquito. Once inside the host, ZIKV infects and replicates in dendritic cells. Most cases of ZIKV infection are self-limiting in an otherwise healthy host, in pregnant women however the infection results in teratogenic effects [117]. There is currently no viral-specific treatment available to treat or prevent ZIKV infection, conveying an urgent need for the development of anti-viral therapies. Most available treatment options are palliative and involve the administration of fluids and plenty of rest.

Glycoprotein E (envelope) , one of the three structural proteins of ZIKV, is a well-established target for antibodies, as it is responsible for viral entry into host cells [115, 118]. As such many of the therapeutics currently being developed act by binding directly to the E protein thus impairing its capacity to mediate membrane fusion with the host cell [118]. Several research groups have sought to identify epitopes on the E protein that are targeted by monoclonal antibodies [119-122]. While antibodies targeting all regions of the protein have been identified, majority of the antibodies are specific for the DIII domain of the E protein [119, 120, 122, 123]. In a mice model study by Liang et al, the authors constructed and purified recombinant envelope protein [124]. The immunogenicity and protective efficacy of the protein was tested in immunocompetent mice. The protein elicited cellular and humoral responses and protected mice from in vivo zika virus challenge. The exact role of envelope protein is not yet fully understood but being present on a large area of the viral surface makes this protein a good immunotherapeutic target [124].

Within this study we sought to demonstrate the versatility and applicability of SNAP-tag and the other recombinant antibody technologies used in this study by applying it to the development of a ZIKV-specific fusion protein. To this end we chose the ZIKV DIII E protein as the target of the fusion protein and utilized the same methodology to generate the ZIKV-specific protein as used in the development of the TNBC-specific fusion proteins.

Section 6: Aims and Objectives

There has always been an interest in understanding the link between the etiology of cancer and its essential biological nature [125]. With the development of molecular and genetic profiling technologies, enhanced insights into the intricacies governing the biochemical differences between healthy and cancerous cells, have been made possible [126]. The growing body of knowledge of the complex nature of cancer cells has initiated a shift from traditional non-targeted therapies to more targeted approaches. A more rational perspective to the one-size-fits-all notion is being assumed to individualize and streamline the development of novel therapies for those who are unlikely to benefit from present treatment options [127]. While this precision medicine approach has the potential to herald a superior future for cancer patients, it remains a research area that is in its infancy [128]. For infamously hostile and tenacious cancers such as TNBCs, the deficiency of druggable targets increases the reliance on regular treatment options (chemotherapy, surgery, radiotherapy) that are frequently associated with elevated risks of relapses [12]. Furthermore, the heterogeneity of TNBCs suggest that the currently available tools are insufficient for the accurate detection and prognosis of the disease [129]. This implies that no single receptor can be targeted in all TNBCs and currently no such effective target exists [130].

Thus, at the Medical Biotechnology and Immunotherapy Research Unit we hope to use SNAP-tag technology in combination with other recombinant antibody technologies to establish a methodology for the development of unique antibody fusion proteins. Using these platform technologies we aim to generate numerous SNAP-tag based antibody-fusion proteins, each labelled with a BG-modified fluorophore or cytotoxic drug. The use of fluorophores allow for sophisticated optical imaging, with immense repercussions for in vitro and in vivo diagnostics, permitting visualization of real-time binding and internalization kinetics [131-134].

Once available, these precision medicine tools would provide a direct link between diagnosis based on differential expression of disease-specific targets and immunotherapy. Patients stratified on their marker profiles would best respond to therapies targeting those markers, significantly reducing the cost of treating patients with therapies they are not susceptible to. Consequently, patient profiling data resulting from precision medicine applications would provide the information necessary to discriminate diseased cells from healthy cells. These differentially overexpressed CSRs are ideal targets for disease-specific antibodies and thus could be used as a target for immunodiagnosis and immunotherapy.

Therefore, in this study we propose developing a companion immunodiagnostic panel which when combined could enable the identification and classification of a vast majority of TNBCs. This will be done with a combination of recombinant SNAP-tag fusion proteins and multiplex immunofluorescence imaging which has already been established by PhD student Natasha Hardcastle. Using this methodology, we would be able to validate disease-targets and provide the incentive required for developing the according targeted therapies. This is important as SNAP-tag based antibody fusion proteins can easily be turned from a diagnostic tool into a therapeutic one by replacing the BG-modified fluorophore with a BG-modified cell-killing warhead. To this end, we hope to demonstrate the versatility of such a technology by developing a fusion protein which targets a disease other than TNBC, in this instance ZIKV was chosen.

Overall, this project intends to use immunoglobulin sequence information from patents, literature and other experimental sources all originating from antibodies able to specifically bind to important cell surface proteins to design single-chain variable fragments. Additionally, we aim to set up technologies such as phage display that would enable the identification of novel disease-specific scFvs. Both novel and pre-existing scFv's of importance will be genetically fused to SNAP-tag allowing for the autocatalytic covalent conjugation to any benzylguanine modified substrate. The SNAP-tag bound fusion proteins will be expressed in a HEK293T mammalian expression system and subsequently purified with ion exchange chromatography. Purified proteins will be tested for functional activity through binding of target antigens on cells or with antigen-based ELISA. The main aim of this methodology is to demonstrate its capacity to streamline the development of potential immunotherapeutic and immunodiagnostic fusion proteins. There are multiple streamlines of methodology that will be run simultaneously within this project as described below:

Streamline 1: To generate a 2-clone phage library for optimization of phage screening.

Phagemids and phage display techniques have multiple applications in antibody production. With these technologies it may be possible to design an antibody from scratch with the options to choose its component parts, such as affinity, size and effector function [135]. The extraordinary role antibodies already play in therapy, diagnosis and research demonstrates how invaluable phage display technology can be in combination with antibody engineering. Thus, the first major aim of this study is to set up a phage display technique that could be used to optimize selection of novel disease-specific antibodies. To address this aim we chose to generate a 2-part phage library containing antibodies targeting the differentially expressed cancer-associated receptors chondroitin sulphate glycoprotein 4 (CSPG4) and epidermal growth factor receptor (EGFR).

This 2-part phage library would then be screened on CSPG4 and EGFR positive and negative cell lines allowing for the controlled optimization of a screening process.

The following aims are intended to address this streamline:

- a) To clone PIT2- α CSPG4-snap and PIT2- α EGFR-snap
- b) To propagate the phage library and express PIT2- α CSPG4-snap and PIT2- α EGFR-snap displaying phages.
- c) To optimize a selection of the phage library on CSPG4 and EGFR positive cell lines.

Streamline 2: Development of a companion differential diagnostic panel for TNBC

Within streamline 2 we aim to develop a companion differential diagnostic panel which is capable of characterising TNBC tumours based on their TIL composition and expression of checkpoint inhibitors. This panel should be used in combination with a TNBC diseases specific panel to assist doctors in predicting what potential immunotherapies an individual patient may be susceptible to. The following are the aims for developing the differential diagnostic panel:

- a) To identify various antibody sequences targeting immune cells and checkpoint inhibitors
- b) To design and codon optimize scFvs sequences from the various identified targets.
- c) To clone the designed scFvs into mammalian expression systems.
- d) To express the scFv-snap fusion proteins in mammalian expression systems.
- e) To confirm the presence of full length SNAP-tag based proteins through western blot or fluorescent conjugation.

Streamline 3: Construction of a breast cancer specific therapeutic fusion protein.

This aim fits into the larger MB&I goals of developing a disease-specific fusion protein panel which can be used to stratify TNBC patients into different groups based on their differential expression of disease-specific targets. This would provide the evidence required for the development of new targeted therapies and serve to inform doctors of which targeted therapies are appropriate for individual patients. BCK1 was chosen as a potential therapeutic target due to its specific expression on breast cancer cells and lack of expression on healthy cells, making it the ideal cancer-specific target for immunotherapy.

The following are the aims for developing the BCK1 specific construct:

- f) To perform *in silico* cloning of the scFv from BCK1 antibody into a SNAP-tag pCB backbone.
- g) To generate a BCK1(scFv)-SNAP fusion protein using molecular cloning techniques.

- h) To express BCK1-snap fusion proteins in mammalian expression systems
- i) To confirm the presence of full length SNAP-tag based BCK1- specific proteins through western blot.

Streamline 4: To demonstrate the flexibility of SNAP-tag technology by generating a Zika-virus specific SNAP-tag protein.

In the final part of this project, we aimed to apply the technology and skills learnt in developing antibody-based diagnostic/therapeutic fusion proteins for breast cancer treatment to zika virus thus demonstrating the versatility and robustness of the utilized SNAP-tag technology. The scFv used for the generation of the zika fusion protein will be sourced from a collaboration partner Dr. Paolo Zanotto based at the University of Sao Paulo (USP). This sequence originated from cells producing antibodies able to neutralize arbovirus infection as a post-attachment step and were shown to have cross reactivity to zika-virus envelope protein. The aims of this streamline are as follows:

- a) To clone α Zika-DIII envelope specific scFv into the pCB vector backbone.
- b) To express α Zika-DIII-SNAP-tag antibody in mammalian expression cells
- c) To purify full length, function α Zika-DIII-SNAP-tag protein from expression cells.
- d) To demonstrate functionality of the full-length protein by antigen-based ELISA.

Chapter 2: Methodology

2.1 Assembly of scFvs from antibody sequences identified in publicly available sources

Using keywords and search filters we identified various patents and publications containing full length antibody sequences. Where possible, the antibody sequences were extracted in FASTA format from GenBank. If FASTA formats were not available, sequences were manually typed. As most antibody sequences were given in their amino acid format, reverse translation was done using the online software “geneinfinity” to obtain the nucleotide sequences (http://www.geneinfinity.org/sms/sms_backtranslation.html). Sequences originating from non-mammalian sources underwent codon optimization for mammalian cell line expression. All extracted nucleotide sequences were then aligned in IgBlast to identify the heavy and light chain sequences and the presence of the CDRs and framework regions within each chain. The identified heavy and light chain sequences were then assembled into an scFv cloning format, containing a *Sfi1* restriction site at the 3’ end followed by the heavy chain, a Gly4Ser linker, the light chain sequence and a *Not1* restriction site as shown in figure 2.1.1 Below.



Figure 2.1.1: Representation of an assembled scFv containing a 5’ *Sfi1* restriction enzyme recognition site, followed by a antibody heavy chain sequence, a (Gly4)Serine linker sequence, a light chain antibody sequence and a 3’ *Not1* restriction enzyme recognition site.

2.2 In Silico Cloning

SnapGene[®] software (version 3.1.1, GSL Biotech, Chicago) was used to design constructs for synthesis and perform in-silico cloning. Newly designed open-reading frames (ORFs) containing scFvs of interest were theoretically cloned into a lab-designed vector backbone, (table 2.2.1) pCB-Annexin V-SnapF (previously generated at MB&I). Various coding elements as described in table below were introduced into the vector together with the scFv nucleotide sequence in SnapGene to ensure proper transcription and translation of the full-length DNA plasmid. Once all the coding elements and scFv were confirmed to be in frame and contain

no stop codons, the scFv inserts were synthesized in pUC57 plasmid vectors by Genscript USA and used in molecular cloning.

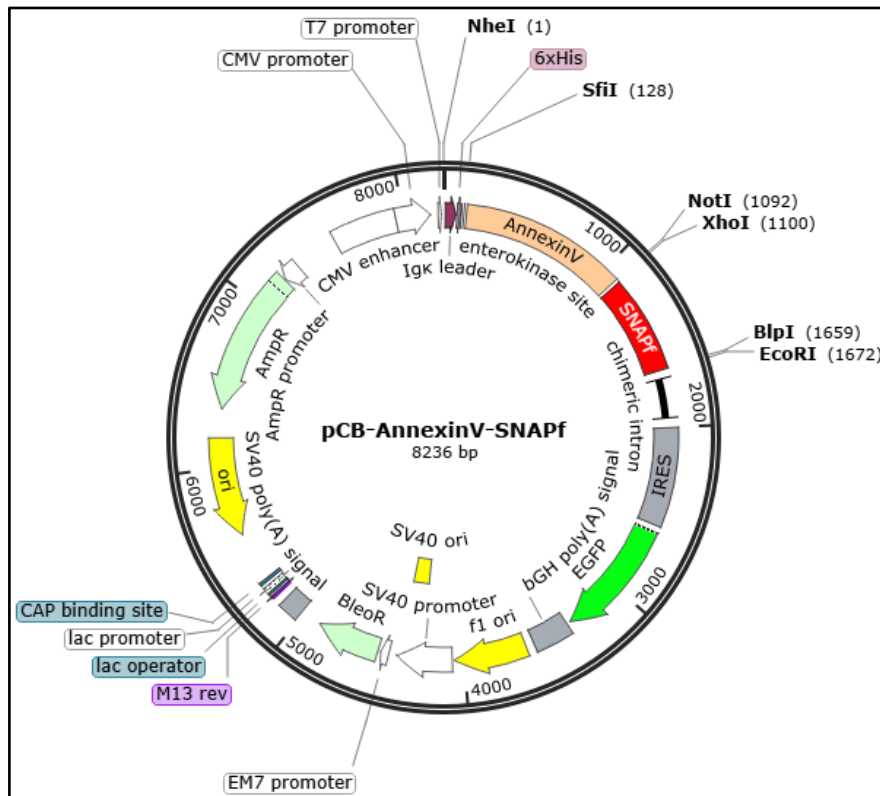


Figure 2.2.2: Vector map of pCB-Annexin V-SnapF. The *Sif1* and *Not1* restriction sites were used for cloning of the various newly designed ScFv's into the expression plasmid. Other important components that are crucial to expression of the plasmid are depicted on the figure and described in table 2.2.1 below.

Table 2.2.1: Description of pCB backbone components

Component	Description
Cytomegalovirus promoter	Allows efficient, high-level expression of recombinant protein.
Igk leader	Essential for the secretion of fusion proteins into HEK293T cell culture supernatant.
His₆-tag	Has high affinity binding to Ni ²⁺ -chelating resin used during purification. In addition, allows detection of fusion proteins with anti-His ₆ -tag antibody.
Enterokinase cleavage site	Used for the removal of the N-terminal his-tags during isolation of fusion proteins.
scFv	N-terminal single-chain variable fragment extracted from relevant antibody
SNAP-tag®	C-terminal self-labelling SNAP-tag protein used for the conjugation of benzylguanine substrates.
c-myc	Allows detection of fusion protein with anti-myc antibody.
IRES	Allows the initiation of translation from any position within an mRNA immediately downstream from where the IRES is located. Mediates expression of green fluorescent protein (GFP) in the pCB vector.
Green Fluorescent protein	Reporter protein used for detection of successful transfection in mammalian cells.
bGH PA terminator	Ensures efficient transcription termination and polyadenylation of mRNA.
F1 origin	Facilitates rescue of single-stranded DNA
SV40 promoter	Allows efficient, high-level expression of the Hygromycin resistance gene and episomal replication in cells expressing the SV40 large T antigen
BleoR	Bleomycin resistance protein used for selection in HEK293T expression cells under Zeocin treatment.
SV400 PA terminator	Allows for efficient transcription termination and polyadenylation of mRNA
pBR322 origin	Ensures high-copy number replication and growth in E. coli
Amp(R)	Selective marker
Amp(R) promoter	Efficient expression of the Ampicillin resistance gene
T7 promoter	Allows for in vitro transcription in the sense orientation and sequencing through the insert
Lac operator	Inhibits the transcription of lac genes in the absence of lactose by interacting with the lac repressor.
Lac promoter	Controls transcription of lac operon genes involved in bacterial absorption and metabolism of lactose.
CAP binding site	Allows for the binding of catabolite activator protein which is necessary for the transcription of various genes including those involved in sugar metabolism.
<i>SfiI/NotI</i>	Restriction sites used to clone in scFv's of interest.

2.3 Molecular Cloning

2.3.1 Transformation of chemically competent DH5 α e.coli cells with plasmid DNA

E.coli NEB[®] 5-alpha competent cells (K12 strain, genotype: fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17, New England Biolabs, USA) were used to propagate exogenous vector DNA (pCB-AnnexinV-SNAP) as well as pUC57 vectors containing assembled scFv inserts into sufficient copy number to permit cloning using the heat shock method. The DH5 α strain has a high propensity in withstanding heat shock making it a suitable option for transformation. For the transformation, the bacterial cells are first thawed on ice for 10 mins. 5 μ l of vector DNA diluted to 200 ng/ μ l in sterile nuclease free water was then added to a sterile 1.5ml Eppendorf tube containing 50 μ l of thawed DH5 α *e.coli* cells. The tube containing the cell-DNA mix was then carefully flicked and incubated for 30 minutes on ice to allow for cooling and proper dispersion of all components. The mixture was then heat shocked at 42°C for 60 seconds in a heating block, resulting in minor temporary breaks within the bacterial cell wall which allows for uptake of vector DNA into the cell. This was followed by snap cooling of the cells on ice for 5 minutes. The mixture was subsequently supplemented with 950 μ l of room temperature SOC (Super Optimal broth with Catabolite repression, containing 2% (w/v) vegetable peptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20 mM glucose) medium and incubated at 37°C for 1 hour. SOC is nutrient-rich containing peptides, vitamins, amino acids and glucose in a low-salt medium providing the optimal metabolic conditions for bacterial growth thereby aiding in rescue of heat shocked cells. The use of SOC maximises transformation efficiency. The entire volume of 1 ml transformed cells is then added to a conical flask with 50ml of Luria Broth (LB, containing 1% (w/v) Casein Peptone, 0.5% (w/v) yeast extract and 1.0% (w/v) NaCl, Thermo Fisher Scientific, South Africa), supplemented with 200 μ g/ml of Ampicillin antibiotic. As the vector DNA contains an ampicillin resistant gene thus only those cells which have successfully taken up the vector DNA will survive and grow while cells not containing the vector plasmid will be killed by the ampicillin. The conical flasks were placed in a shaking incubator overnight at 37°C to allow for sufficient growth and propagation of the vector DNA.

2.3.2 Phenol-chloroform extraction of bulk prepared vector DNA

Overnight transformed DH5 α were pelleted by centrifugation at 4000 rpm (3901 xg) for 15 minutes at 4°C in a Beckmann Allegra X-22 centrifuge. The cell culture supernatant was discarded and the cell pellet was lysed through aspiration in 1ml of lysis solution containing 50 mM glucose, 10 mM EDTA and 25 mM Tris-HCl. All three components of lysis solution one play important roles in recovery of the vector DNA; glucose aids in the osmotic release of cellular content by increasing the osmotic pressure outside the cell, Tris acts as a buffering

agent maintaining a constant pH of 8 and EDTA protects the released DNA from enzymatic degradation by DNAses through binding to divalent cations necessary for DNase activity.

Lysing of cells was followed by the addition of 2ml of solution II which I made with 0.2 M NaOH and 1% sodium dodecyl-sulphate (SDS), a detergent which breaks apart the lipid membrane of cells causing cellular rupture. SDS also solubilizes cellular protein while NaOH denatures double stranded DNA to form single strands. Then ice-cold solution III containing 11.5% acetic acid and 3 M potassium acetate was added to the mixture. Acetic acid neutralizes the pH allowing the single stranded DNA to renature while potassium acetate precipitates SDS along with cellular debris. E.coli chromosomal DNA, which is only partially renatured due to its size also precipitates during this step but the plasmid DNA of interest being much smaller remains in solution. The mixture containing the precipitate and soluble plasmid DNA is then centrifuged at 4000 rpm (3091 xg) for 20 minutes. The supernatant was harvested into 5 x 1.5 ml microfuge tubes and treated with 10 µl of RNA cleaving RNase A at 10mg/ml and incubated at 37°C for 1 hour to allow for degradation of contaminating RNAs.

Following the incubation, the sample is treated with 700 µl of isopropanol and left at room temperature for 10 minutes to allow precipitation of the plasmid DNA. The precipitated DNA was pelleted by centrifugation at 13 000 rpm for 10 minutes. The supernatant was discarded and 200 µl of ice-cold ethanol was added to the tube to remove any residual contaminants. The ethanol was then poured out of the tube without disturbing the pelleted DNA. The cleaned DNA pellets were then each resuspended in 100 µl of nuclease-free water and pooled into a single 500 µl reaction in a clean 1.5 ml Eppendorf tube.

The addition of 50 µl of 3 M Sodium Acetate at this step in the protocol aids in making the DNA less hydrophilic allowing for better precipitation. In solution the sodium acetate breaks up into Na^+ ions and $[\text{CH}_3\text{COO}]^-$. The positively charged Na^+ ions neutralize the negative charge of the PO_3^- groups on the DNA backbone making it less water soluble and easier to precipitate. The mixture is then treated with 700 µl of phenol-chloroform (1:1 v/v) which denatures any contaminating proteins still present in the sample. The sample is subsequently centrifuged at 13000 rpm for 10 minutes to allow for the separation of the liquids into an organic phase containing the denatured protein and an aqueous phase containing the plasmid DNA. The top aqueous phase is transferred into a new tube and treated with 350 µl of chloroform isoamyl alcohol (24:1 v/v) to prevent emulsification of the solution and reduce foaming at the organic-aqueous interphase. The sample is centrifuged again at 13 000 rpm for 10 minutes. Again, the top aqueous phase of the sample is transferred to a clean tube.

The DNA is then allowed to precipitate overnight in 1 ml of ethanol at -20°C. The precipitated DNA is then pelleted by centrifugation at 13000 rpm for 20 minutes at room temperature. The supernatant is discarded, and the pellet is washed with 1 ml of 70% ethanol which is immediately poured off the sample without disturbing the pellet. The purified DNA pellet is

dried in a SpeedVac vacuum (Thermo Fisher Scientific, DE USA). The completely dry pellet is finally solubilized in 30 µl of nuclease-free dH₂O at 42°C for 10 minutes.

2.3.3 DNA Quantification

Following purification all DNA sample were quantified by spectrophotometry on a DeNovix spectrophotometer (Thermo Fisher Scientific, DE USA) and stored at 4°C until required for downstream applications.

2.3.4 Restriction Enzyme Digest

Restriction digests are done with restriction enzymes capable of recognizing and binding to specific DNA sequences only. Once bound to the DNA at the restriction site the enzyme cleaves specific nucleotides either within the recognition site or just outside the site. This type of digests may result in blunt ends which are DNA molecules that end in a base pair, or sticky ends which are DNA molecules that have a nucleotide overhang. For this project the restriction enzyme digest was used for two purposes; to generate compatible vector and insert DNA ends for ligation and for restriction mapping post cloning.

To prepare DNA for cloning, 2µg each of the pCB-AnnV vector DNA and the scFv containing parent vector DNA were individually incubated with 0.5µl of the restriction enzyme *Sfi*1 (NEB) (2000 units/ml) for 3 hours at 50°C. As both plasmids only contain a single *Sfi*1 restriction site the digest allows for linearization of the plasmid DNA. Following this 0.5 µl of *Not*1 (NEB) (2000 units/ml) was added to the DNA and the sample was incubated at 37°C overnight. All digests were done in 20 µl reaction volumes with 10% CutSmart buffer (NEB).

2.3.5 Agarose gel Electrophoresis

Agarose gel electrophoresis was used to resolve restriction digested DNA fragments based on their molecular weight. A 1.2% (W/V) agarose gel was cast with 1xTAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5) and placed into a Wide Mini Sub-Cell systems (Bio-Rad, USA) containing a negative electrode on one end and a positive electrode on the opposite end. DNA samples mixed with 6x Orange DNA loading dye (Thermo Fisher Scientific, USA) were loaded on the negatively charged end of the tank. As DNA is negatively charged, the negative pole of the electrode repels the DNA through the gel matrix towards the positive end. Larger DNA fragments move slower through the gel while smaller fragments migrate faster. The DNA samples were run alongside a 1000 bp molecular weight DNA ladder (New England BioLabs, MA USA) containing DNA fragments of known sizes. Comparison of the distance of migration of a fragment to the migration distances of the ladder was used to determine the size of the different fragments within the DNA samples. 10% SYBR safe nucleic acid stain (Thermo Fisher Scientific, South Africa) supplemented in the agarose gel was used to visualize the DNA fragments under blue light excitation at 509 nm using a Dark Reader Transilluminator (Clare Chemical Research, CO USA). Parts of the gel containing DNA bands with the same molecular weight as expected for the vector and insert were physically cut out

of the gel using a sterile scalpel. The DNA containing gel pieces were transferred to a clean 1.5 ml microfuge tube in preparation of gel extraction.

2.3.6 Recovery of DNA fragments from Agarose gels

DNA fragments of interest (vector and insert) were retrieved from agarose gels using the QIAquick® Gel Extraction Kit (Product number: 28704, Qiagen, Netherlands). Excised gel pieces were incubated with 500 µl of Buffer QC at 50°C for 10 minutes with intermittent agitation. Buffer QC dissolves agarose at 50°C while maintaining a stable pH and DNA integrity. Afterwards 100 µl of isopropanol is added to the sample to precipitate the DNA. The mixture is then loaded onto a QIAquick spin column supported by a 2 ml collection tube. The sample is centrifuged at 10000 rpm allowing the DNA to bind to the column while the rest of the supernatant flows through. The flow-through was discarded and the column containing the DNA was washed by incubation with 750 µl of Buffer PE for 2 minutes followed by a centrifugation at 10000 rpm for 1 minute. The flow-through was discarded and the column was moved to a clean 2 ml tube. 20 µl of pre-warmed nuclease-free water was added to the column and allowed to stand for 5 minutes to elute the DNA. The column was then centrifuged at 10000 rpm for 1 minute and the flow-through now containing the gel extracted DNA was quantified on the Denovix and used for the subsequent ligation reactions.

2.3.7 T4 DNA Ligation

T4 DNA ligation was used to incorporate relevant scFv inserts into the pCB vector backbone to create the required recombinant expression vectors. The ligation reactions were set up as per manufacturer's instructions using 400000 U/ml (New England BioLabs) T4 DNA ligase and 10x ligase buffer in a final reaction volume of 20 µl. All reactions contained 50 ng of PCB vector DNA and 3 different vector: insert molar ratios; 1:1, 1:3 and 1:5. The mass of insert DNA required in each reaction was calculated using the NEBio Calculator (<https://nebiocalculator.neb.com/#!/ligation>). The reactions were incubated at 16°C overnight followed by a 65°C heat inactivation step for 10 minutes. The ligated product was transformed into E.coli DH5α cells as previously described in section 2.3.1. The only amendment made was that 200 µl of the transformed product was spread on an LB agar plates (1% (w/v) Casein Peptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl and 1.2% (w/v) agar) supplemented with 200 ng/ µl of ampicillin rather than being grown in LB broth. A 'vector-only' control (SfiI/NotI double digested pCB-AnnexinV-SNAPf DNA) as well as 'bacterial-only' control was also set up to quantify any vector re-ligation and determine if any contamination

was present respectively. All plates including controls were placed at 37°C overnight and bacterial colonies were picked for restriction mapping thereafter.

The transformation efficiency of each sample was calculated in colony forming units (cfu) per μg of DNA using the following formulae:

Transformation efficiency (cfu/ μg) = Number of transformed colonies / Amount of DNA spread on the plate (μg)

2.3.8 Mini preparation of transformed colonies.

Successfully transformed colonies were picked from Luria agar plates and grown in 2 ml Luria broth overnight at 37°C. DNA was extracted from 1.5 ml of the overnight cultures using a truncated version of the phenol-chloroform DNA extraction method described earlier. The method utilized 10% volumes of all reagents. The purified DNA was used for restriction mapping and the other 500 μl of overnight culture was placed at 4°C until required.

2.3.9 Restriction analysis

All plasmids were screened by restriction analysis using restriction enzymes. The theoretical nucleotide sequences of each plasmid was used to determine which enzyme would digest the successfully ligated construct in a manner that would distinguish it from parent vectors. Snapgene (version 3.1.1, GSL Biotech, Chicago) was used to predict the cutting patterns of different plasmids by different enzymes and thus used to determine the required digestion enzymes for restriction mapping. The software enables virtual plasmids (with the same DNA sequence as our plasmids of interest) to be cleaved and displays the resulting fragments on a theoretically agarose gel. Enzymes which digested the parent vector differently from the ligated product were used to set up restriction digests as described in 2.3.4 above.

Each colony was digested with up to 3 enzymes and those which showed the expected banding pattern on an agarose gel were bulk prepared. Bulk preparations were done from the 500 μl of overnight culture remaining after ligation, these were grown up in 50 ml of Luria broth before undergoing DNA extractions with the phenol-chloroform method as described before. Purified samples were sent to Inqaba Biotechnical Industries (Pretoria, South Africa) for Sanger DNA sequencing. For the sequencing, a universal T7 primer (AATACGACTCACTATAG), and specific internal primers (CMV forward: CGCAAATGGGCGGTAGGCGTG and SNAP internal: CCTCGCCGAACCTTCACCACCTTC) were used in the sequencing of the cloned sample.

2.4 Sequence verification

SnapGene (version 3.1.1, GSL Biotech, Chicago) was used to align the actual sanger sequence results received from Inquaba Biotech with the original in-silico designed sequence to determine if the scFv of interest was successfully ligated into the pCB-SNAP/PIT2 vector backbone. The sequence was also checked for the presence of mutations in the final construct. Sequence verified DNA was then used for transfection into mammalian expression cells.

2.5 Mammalian cell culture and protein expression

Recombinant DNA resulting from molecular cloning was transfected into HEK293T mammalian expression system using XtremeGene™ transfection reagent (Sigma-Aldrich, South Africa). Mammalian expression systems contain all the machinery necessary for post-translational modification of complex proteins such as antibodies and thus are used to generate fully functional recombinant proteins. The X-tremeGene reagent contains lipids and other components which complex with DNA, permitting uptake of the construct into cells with limited toxicity and morphological changes.

For transfection HEK293T cells were seeded at 40-60% confluency into 2x2 mm round cell culture dishes 24-48 hours prior to adding DNA. Cells were grown in RPMI-1640 (Gibco #10566, containing 2 mM L-glutamine, 3.7 g/L NaHCO₃ and 15 mg/L phenol red) supplemented with 10% heat inactivated fetal bovine serum, 100 µg/ml streptomycin (#S-91370 1105 Sigma, USA) and 100U/ml penicillin (#P3032 Sigma, USA) until a confluence of 70-90%. Once the cells were confluent enough, X-tremegene HP reagent and DNA were mixed at a 3:1 (v:v) ratio and added to the cell culture. The DNA mix was spread evenly over the cell culture dish and allowed to grow at 37°C with 5% CO₂ for a few days to allow sufficient uptake of DNA into the cells.

The vectors contained a bleomycin resistance gene for antibiotic selection as well as a green fluorescent protein (GFP) encoding gene. Microscopic viewing of GFP expression in live cells was used to estimate plasmid uptake. For plasmids with low transfection efficiency, 100ng/ml of Zeocin™ (ThermoFisher, USA) was applied to remove untransfected cells.

Zeocin selection pressure was applied until the cell culture showed around a 70% GFP expression under the ZOE™ Fluorescent cell Imager (Bio-Rad Laboratories, UK). Protein expression from these cultures were then assessed by flow cytometry as described in the section below. Once the cell cultures reached a confluency of greater than 90% with over 90% GFP expression the cells were resuspended in media by aspiration and the entire culture was moved to larger culture flasks. Cell culture supernatant (CCSN) containing secreted proteins were harvested from cultures which had reached a T175 culture flask. Harvesting was done every 3-4 days by transfer of the CCSN to 50 ml falcon tubes. The supernatant was centrifuged

at 1000xg for 5 minutes to remove any cellular debris and the clear supernatant was stored at 4°C until enough culture was collected for protein purification.

2.5.1 Flow Cytometry

Flow cytometry was used to determine protein expression in HEK293T cell cultures. Cells were prepared for flow cytometry as follows, cells were stained with trypan blue (1:1 v:v) ration and counted with the BioRad automated cell counter. 5×10^4 cells were then transferred to 15 ml falcon tubes and centrifuged at 1500 x g for 5 minutes. The CCSN was discarded and the cells were resuspended in 200 μ l 1 x phosphate buffer saline (PBS) solution and transferred to FACS tubes (BD). The cells were then treated to 2% paraformaldehyde to fix them and then centrifuged again. The supernatant was discarded and the cells were then resuspended in 200 μ l of FACS buffer. The final samples were then run on an LSRII FACS machine and analysed with Flow Jo 9.9 software.

2.5.2 Purification of recombinant proteins

Cell culture supernatant containing recombinant proteins of interest was diluted 1:3 with incubation (200 mM NaH₂PO₄, 1.2 M NaCl, 40 mM imidazole, pH 8.0) buffer to ensure that all proteins were in optimal binding conditions for purification. This was followed by filter sterilization to remove microcellular debris using the Nalgene™ vacuum filtration system (Sigma-Aldrich, South Africa) containing a 0.45 μ m Durapore® membrane filter (Millipore, USA).

The HIS-tag element present in all constructs was used to purify the proteins using Immobilized Metal Affinity Chromatography carried out on a Ni²⁺ sepharose affinity resin (packed in a HisTrap™ Excel column, GE Healthcare, USA) on an ÄKTA Avant protein purification system (GE Healthcare, USA). Purified cell culture supernatant is applied to pre-equilibrated HisTrap column, followed by 20 column volume washes with equilibration buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0). The bound fusion protein is then eluted with an imidazole containing elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 500 mM imidazole, pH 8.0) which disrupts the bonds between the HIS-tag and Ni²⁺. Imidazole competes with HIS-tags for Ni²⁺ binding thus by gradually increasing the imidazole concentrations in the column, fractional elution of the his-tagged protein is done.

Proteins in the fractions are concentrated with a 10K-sized Amicon filter (Sigma-Aldrich, South Africa). The column is also used to complete a buffer exchange and remove any residual imidazole carried over from purification. For the concentration 15 ml of protein fractions is loaded on the filter at a time and centrifuged at 4000xg for 30 minutes at 4°C. The flow through is discarded and the same filter is used repeatedly , trapping protein of larger than 10KDa in its filter. Once all the fractions have been concentrated the proteins are washed 3 times with 1x phosphate buffered saline (1x PBS, containing 137 mM NaCl, 8.8 mM Na₂HPO₄, 2.7 mM KCl and 1.75 mM KH₂PO₄, pH 7.4) to remove all imidazole prior to downstream

applications. The concentrated purified proteins are quantified by UV spectrophotometry using the DeNovix DS-11 (De-Novix, USA).

2.5.3 Protein Quantification by Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to assess IMAC purified SNAP-tag fusion proteins. A 10% discontinuous and 4% stacking gel was used with SDS which provides the reducing conditions necessary to separate proteins based on. To prevent the tertiary structures and charges on the proteins from interfering with the migration of the protein in the gel, the proteins were first denatured. This makes sure that only the molecular weight of the protein influences its movement on the SDS-PAGE gel. To denature the protein, 15 µl of IMAC purified protein was added to 4 µl of NEB 4X protein-loading buffer and the mixture was heated to 95°C for 5 minutes. The denatured protein as well as a Page Ruler prestained protein ladder (5 µl) (Thermo Fisher Scientific, South Africa), were then loaded on the SDS gel and allowed to run at 100V for 95 minutes on the Mini-Protean Tetra Cell system (Bio-Rad, USA). The proteins were then visualized by the addition of AquaStain (Bulldog Bio, UK) for 30 minutes and compared to a Page Ruler to identify the size.

2.6 Functional Assays

2.6.1 Western blot

Once the size of proteins were verified on SDS gels, their functionality and integrity was assessed with western blot. For western blots an SDS-PAGE gel was run as described in the section above but rather than staining the gel with Aqua-stain the protein bands on the gel were transferred to a nitrocellulose membrane (PVDF transfer membrane, Roche, Switzerland) using a Mini Trans-Blot Cell system (Bio-Rad, USA) set at 100 volts for 75 minutes. After the transfer the membrane was incubated with a solution of PBS containing 2% milk for 1 hour at room temperature to prevent any non-specific binding of antibodies. The membrane was then washed and then incubated with 1:1000 dilution of anti-HIS-tag rabbit primary antibody (Qiagen, Hilden, Germany) at 4°C overnight. A goat-anti-rabbit (Bio-RAD, Hercules, USA) secondary antibody conjugated to horseradish peroxidase (HRP) (Bio-Rad, USA) was then added at a 1:5000 dilution. This was followed by the application of a chemiluminescence substrate called Clarity™ Western ECL substrate (Bio-Rad, USA) which contains enhanced substrate and stable peroxide that interacts with HRP resulting in the production of a chemical substance which emits light at 425 nm. The membrane is then visualized using a Gel Doc™ XR Gel Documentation System (Bio-Rad, USA). A chemiluminescent ladder (5 µl) (SuperSignal™ Molecular Weight Protein Ladder, Thermo

Fisher Scientific, South Africa) pre-loaded with the sample was used to assess the size of the protein bands.

2.6.2 Protein labelling with BG-Alexa Fluor®488

Protein conjugation was used to determine if the snap-tag on the recombinant protein was functional. Conjugation was done with NEB protein labelling in vitro protocol as per manufactures instruction. Reactions were set up in 50 µl final reaction volumes consisting of 42 µl of 1x PBS, 1mM DTT, 5 µM SNAP-tag purified protein and 10 µM SNAP-tag substrate. The mixture was incubated for 30 minutes in the dark to at 37°C to allow for the conjugation reaction and prevent photo-bleaching of the fluorophore. Only functional SNAP-tag would interact with the BG-modified substrate. The conjugated proteins were then run on an SDS-PAGE gel as describe in the section above to determine if the conjugation was successful and the SNAP-tag functional by the presences of a fluorescence band at the correct molecular weight when visualized by blue light excitation, using a Dark Reader Transilluminator (Clare Chemical Research, USA).

2.6.3 Enzyme-Linked Immunosorbent Assay (ELISA)

ELISA was used to assess the binding capacity of purified Zika-virus recombinant SNAP-tag proteins. The ELISA plates used for detection was set up by our collaborates on the study , Prof. Paolo Marinho de Andrade Zanotto (Departamento de Microbiologia do Instituto de Ciências Biomédicas (ICB) da USP). Briefly, two strains of Zika virus were used to set up the assay, ZIKV MR766 low passage and ZIKV Bahia. Both viruses were amplified and purified using the PEG-8000 methodology. The purified viruses were then fixed on an ELISA plate using carbonate-bicarbonate buffer protocol. All plates were set up to have 2 concentrations of virus, columns 1-6 of a 96-well ELISA plate were coated in 150 ng of virus while columns 7-12 were coated with 200 ng of virus. The plates were shipped to us and the ELISA was performed in Cape Town.

For the assay, ELISA plates were first blocked with a solution of 1xPBS containing 0.05% tween 20 , 1% BSA and 2% milk powder. 300 µl of the blocking solution was added to each well and incubated for 2 hours at 37°C. The plates were then washed 4 times with wash buffer (1x PBS+ 0.05% Tween 20). 200 µl of HEK293T cell-culture supernatant containing the ZIKA-SNAP fusion protein was then applied to each well and incubated overnight at 4°C. The plate was washed again 4 times with wash buffer. An anti-HIS-tag secondary antibody was then added to the wells and incubated for 2 hours at room temperature. The plates were then washed for a final 4 washes and treated with an anti-HIS HRP conjugated antibody. TMB ELISA substrate was then added to each well and the reaction was incubated for 10 minutes at room temperature before the addition of H2SO4 which enzymatically inhibits the reaction. The colour produced was measured with an ELISA plate reader and used with a standard curve to determine the concentration of functional snap-fusion protein present.

Chapter 3: Results

The limited targeted therapeutic applications available for TNBCs warrants the need for development of major medical interventions. As a single treatment would unlikely be curative to such a heterogeneous disease, a combination of medical tools would be required to ensure early detection and targeting of the majority of TNBC cases. Thus, the major aim of this thesis pertains to developing novel antibody-based diagnostic and therapeutic applications which form an imperative initial step in addressing the current medical necessity of TNBCs. More specifically, this study aims to harness the versatility of snap-tag technology and other antibody engineering tools used to develop TNBC targeting fusion proteins for the development of ZIKA-virus targeting applications. As such the results of this chapter are divided into different sections each addressing different medical needs; section 1) focuses on the making of a 2-component phage library for optimizing screening of novel antibodies, section 2) describes the development of a differential diagnostic panel for use in TNBC patients, section 3) describes the development and expression of a novel TNBC targeting fusion protein pCB-BCK1-Snap, and section 4) demonstrates the versatility of the various antibody-based tools used in this thesis by applying them to infectious diseases using Zika-virus as an example.

Section 3.1: Constructing a two-clone phage library aiming to optimize a live cell selection method

This section of the project was done in collaboration with Dr. Mehmet Tur at the Justus-Liebig University, Giessen, Germany. Dr. Tur's group set up a bio-panning technique for selection of tumour-specific antibodies by selecting against intact tissue biopsies. For their selection, they used the Tomlinson scFv phage library as described in Fitting et al., 2015. One of the main aims of this collaboration was the advancement of expertise with phage display technology, at the University of Cape Town, South Africa. The establishment of this technique would subsequently create a medical biotechnology platform for researching alternative and novel therapeutic approaches utilizing tumor-specific antibodies. Ultimately, this will lay the foundations to bring cancer therapy in South Africa on par with immunotherapy research that is available in the leading medical institutes in North America and Europe.

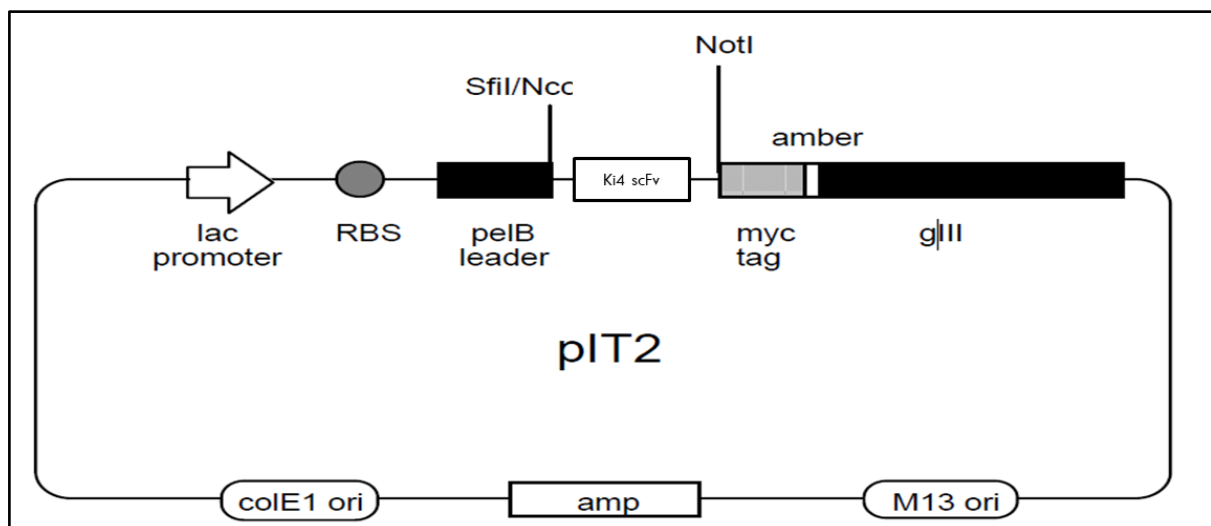


Figure 3.1.1: Schematic overview of a PIT2 phagemid vector containing; a promoter of the bacterial *lac* operon (*lac* promoter); ribosomal binding site (RBS); the *pelB* signal sequence (*pelB* leader) which mediates secretion into the periplasmic space; a *SfiI* and *NotI* restriction enzyme site necessary for cloning of different scFvs into the plasmid; a polypeptide protein tag derived from the c-myc gene (*myc* tag); the *gIII* gene which codes for the protein III; an M13 origin of replication (M13 ORI) which is necessary for replication of the phage genome; an ampicillin resistance gene (*amp*) for selection of positively transformed clones during cloning and a *colE1* ori also necessary for replication of the phagemid.

There were three objectives to address in this section of the project; (1) to have a student learn the phage display technique and transfer this knowledge to Prof. Barths lab at UCT, (2) to generate a two-component PIT2 library which could be used to optimize the bio-panning technique. By mixing in varying ratios of the two components and selecting the libraries on a target specific cell line, one would be able to determine the level of non-specific binding to expect and would be able to improve binding conditions to reduce this background. The selection with just two components would also be informative in determining the sensitivity of the selection strategy, by demonstrating what the minimal concentration of a specific clone would need to be present in a library for binding to occur. For the two-component library, the

targets chosen were chondroitin sulphate proteoglycan 4 (CSPG4) and epidermal growth factor receptor (EGFR).

3.1.1 Selection of Ki4-scFv from Tomlinson I+J PIT naïve scFv library.

As part of the knowledge transfer from Germany to South Africa, I travelled to the Justus-Liebig University in Giessen where I learnt the technique of phage display. As a demonstration of my skills I was tasked with propagating the Tomlinson phage library, which had been spiked with a Ki4-PIT2 phagemid according to the manufacturers instructions. For propagation the library had been superinfected with M13K07 pIII hyperphage allowing for display of polyvalent antibodies. The libraries were then subjected to four rounds of selection on Ki4 coated immunotubes. All non-bound phages were washed off between rounds and the bound phages were eluted of the immunotube, pooled together and reselected. After four rounds of selection a polyclonal Ki4-specific ELISA was run to determine if the propagation and selection strategy was successful. The ELISA was performed on immobilized ki4 cell membrane fragments. Binding was detected using horseradish peroxidase conjugated to anti-M13 antibodies in combination with ABTS as the peroxidase substrate. Six separate propagation reactions were set up simultaneously, three executed by myself and three by another experienced PhD student at Dr.Tur's lab for reference purposes. For the ELISA, 6 dilutions of each library were made and run as biological replicates.

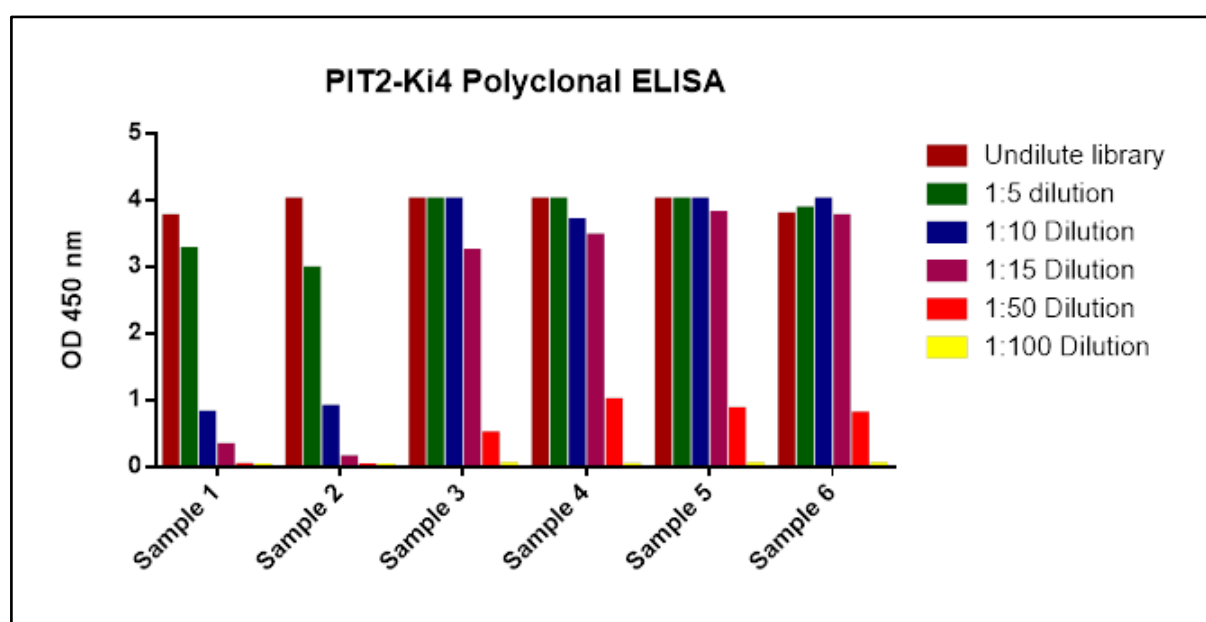


Figure 3.1.1.1: Enzyme-linked immunosorbent assay (ELISA) results of PIT2-Ki4 propagated libraries. Six different propagations and selections were done as biological replicates as denoted by the six samples on the figure. Each sample was plated at six different dilutions. Optical density was measured at OD 450 nm. All OD 450 readings are background subtracted.

As evident in figure 3.1.1.1 the propagation of the Ki4-PIT2 phagemid was successful and resulted in the display of Ki4 scFv. The dilutions showed a dose-dependent binding further demonstrating the specificity of the results. Once we established that the PIT2 phagemid could indeed successfully display scFvs, and that our phage propagation technique was working we sought to use the PIT2 phagemid backbone for optimization of our phage biopanning strategy.

3.1.2 Generation of a two-component phage library

3.1.2.1 : In-silico cloning of PIT- α CSPG4 and PIT2- α EGFR

The first step of generating a two-component phage library involved the cloning of the two desired targets into the PIT2 phagemid backbone. Prior to molecular cloning, snapgene software was used to complete in-silico cloning to make sure the inserts were compatible with the PIT2 phagemid for molecular cloning.

As shown in figures 3.1.2.1.1 and 3.1.2.1.2 below both α CSPG4 and α EGFR scFv were compatible with the PIT2 vector for restriction-based cloning with *Sfi*I and *Not*I enzymes. Thus, we proceeded with molecular cloning of both constructs

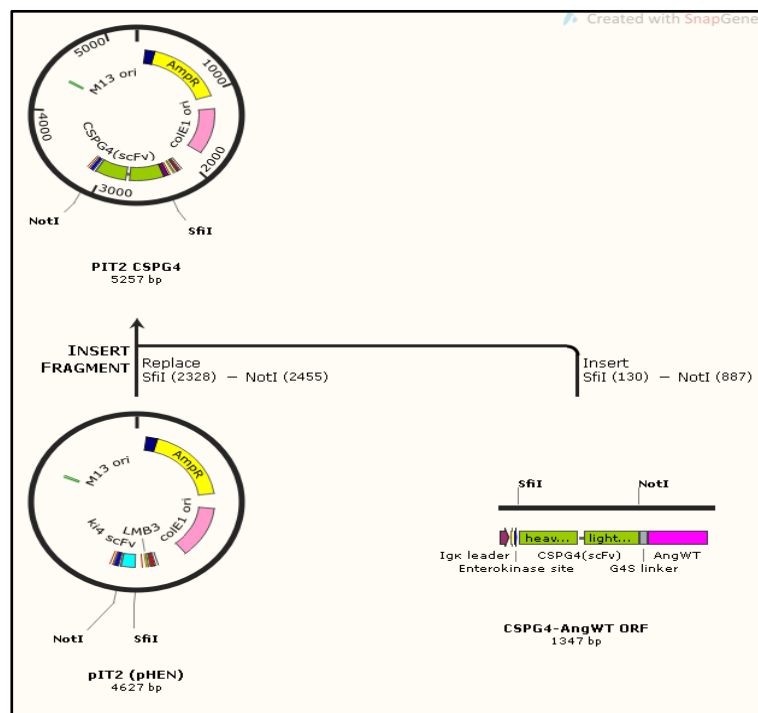


Figure 3.1.2.1.1 PIT2- α CSPG4 construct (A) In-silico restriction cloning of α CSPG4 (scFv) into PIT2 using *Sfi*I and *Not*I. The PIT2- α CSPG4 construct includes *Sfi*I and *Not*I recognition sites; M13 origin of replication; Ampicillin resistance gene, *ColE1* origin of replication; all important in phage propagation.

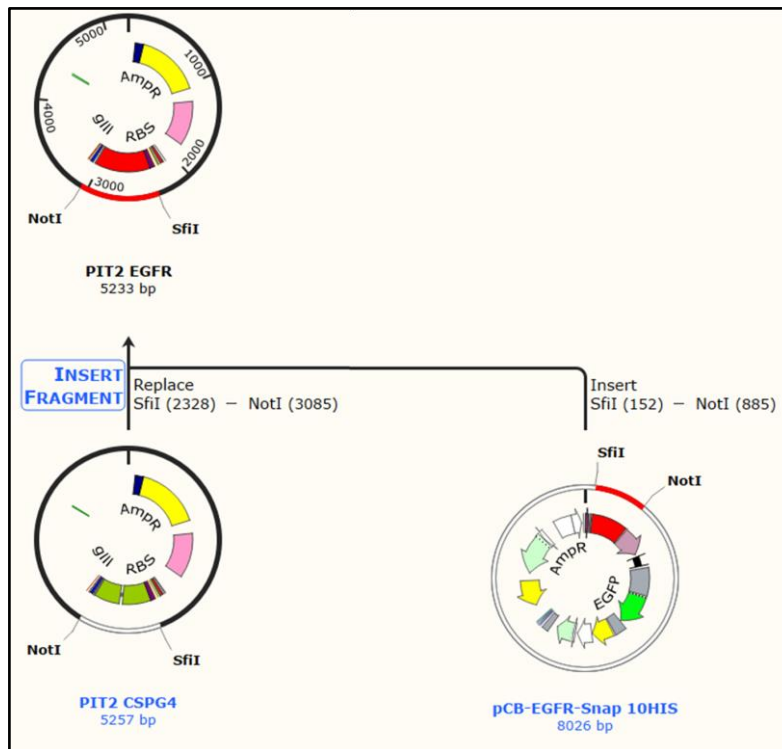


Figure 3.1.2.1.2 PIT2-αEGFR construct (A) In-silico restriction cloning of αEGFR (scFv) into PIT2 using SfiI and NotI. The PIT2-EGFR construct includes *SfiI* and *NotI* recognition sites; M13 origin of replication; Ampicillin resistance gene, ColE1 origin of replication; all important in phage propagation.

3.1.2.2 Restriction Enzyme Digest and agarose gel electrophoresis of PIT2-Ki4, pCB-αCSPG4-SNAP and pCB-αEGFR-SNAP

The Ki4-PIT2 phagemid, as well as pCB-αCSPG4-angiogenin-snap and pCB-αEGFR-snap were double digested with *SfiI* and *NotI* in preparation for ligation. The digested products were run on a 1.2% agarose gel for visualization of the digested fragments as seen in figure 3.1.2.2.1 below.

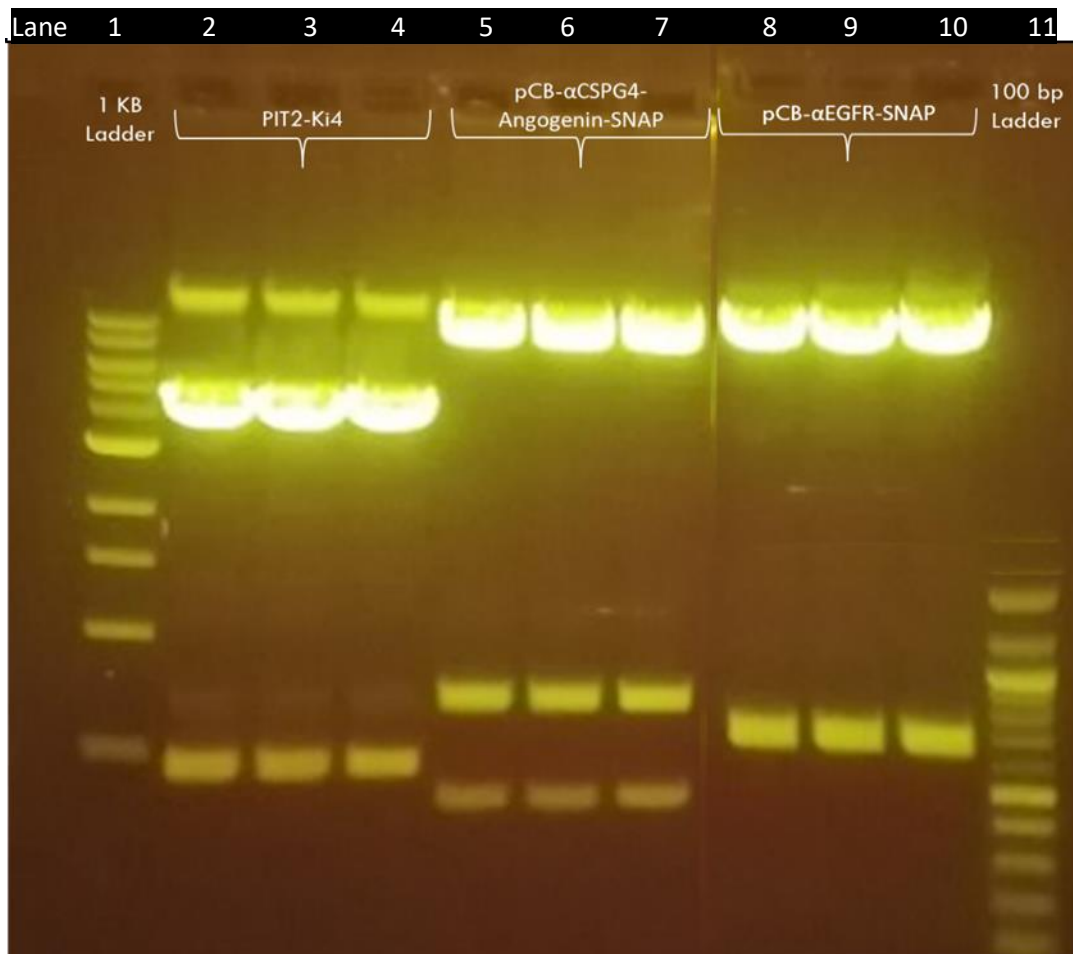


Figure 3.1.2.2.1: Agarose gel depicting DNA fragments after *SfiI* and *NotI* double digest. The 5400bp PIT2 (vector), 924bp α CSPG4(scFv) and 757bp EGFR(scFv) fragments were all isolated for ligation. All digests were run at 37°C overnight. The DNA was separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting agarose gel after electrophoresis

As shown in the agarose gel (figure 3.1.2.2.1) all *SfiI*, *NotI* double digests were successful. The 5400bp PIT2 (vector), 924bp α CSPG4(scFv) and 757bp α EGFR(scFv) fragments were all isolated in preparation for ligations. The DNA fragments were individually purified using the Qiagen quick gel extraction kit. The purified DNA was then quantified with spectrophotometry and used for ligation.

3.1.2.3 Ligation of α CSPG4(scFv) and α EGFR(scFv) into the PIT2 backbone

The previously digested α CSPG4(scFv) and α EGFR(scFv) inserts were ligated to the PIT2 backbone using T4 DNA ligase. The ligation reaction was set up in 2 different vector to insert molar ratios; 1:1 and 1:3. The ligated products were transformed into chemically competent DH5 α *E.coli* cells and plated on LB agar plates containing 200 μ g/ml ampicillin. The plates were incubated overnight to allow individual colonies to grow as shown in 3.1.2.3.1.

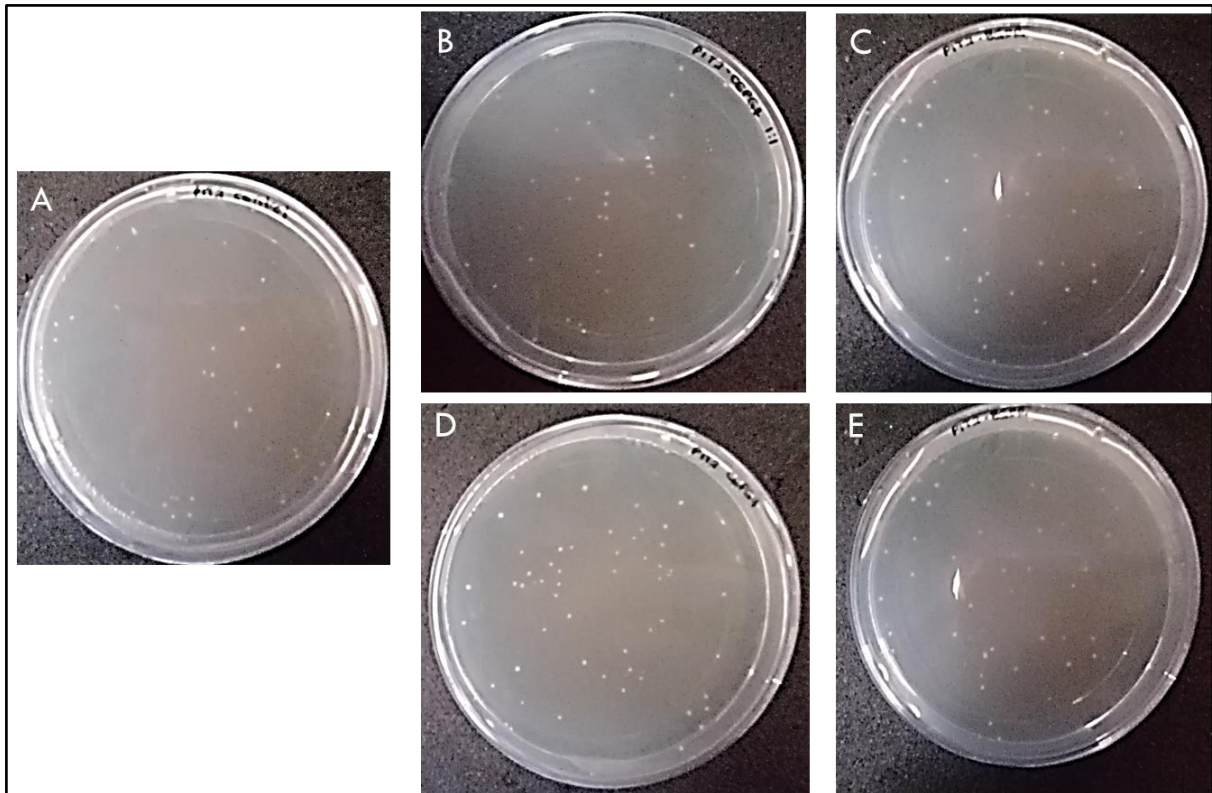


Figure 3.1.2.3.1: Growth of *E. coli* cells transformed with potential recombinant (B,D) PIT2- α CSPG4 and (B,E) PIT2- α EGFR plasmid DNA. Following transformation, 100 μ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 μ g/ml. **(A)** A control vector only plate was also set up.

Colonies were found on the vector only control plate indicating that some vector re-ligation had occurred. A similar number of colonies were found on plates with a 1:1 vector:insert reaction ratio, making it difficult to know which of the colonies would be true transformants. Ligation reactions with a 1:5 vector:insert ratio had a higher transformation efficiency than vector control, thus 2 colonies from each 1:5 reaction plate were chosen for screening with restriction mapping.

3.1.2.4 Restriction mapping of potential PIT2- α CSPG4 and PIT2- α EGFR clones

Selected clones were picked from agar plates and grown overnight in LB broth. Plasmid DNA was then isolated and subjected to restriction mapping with BamH1 and PvuII as these enzymes showed differential banding patterns between the PIT2-Ki4 parental DNA and the possible recombinant clones.

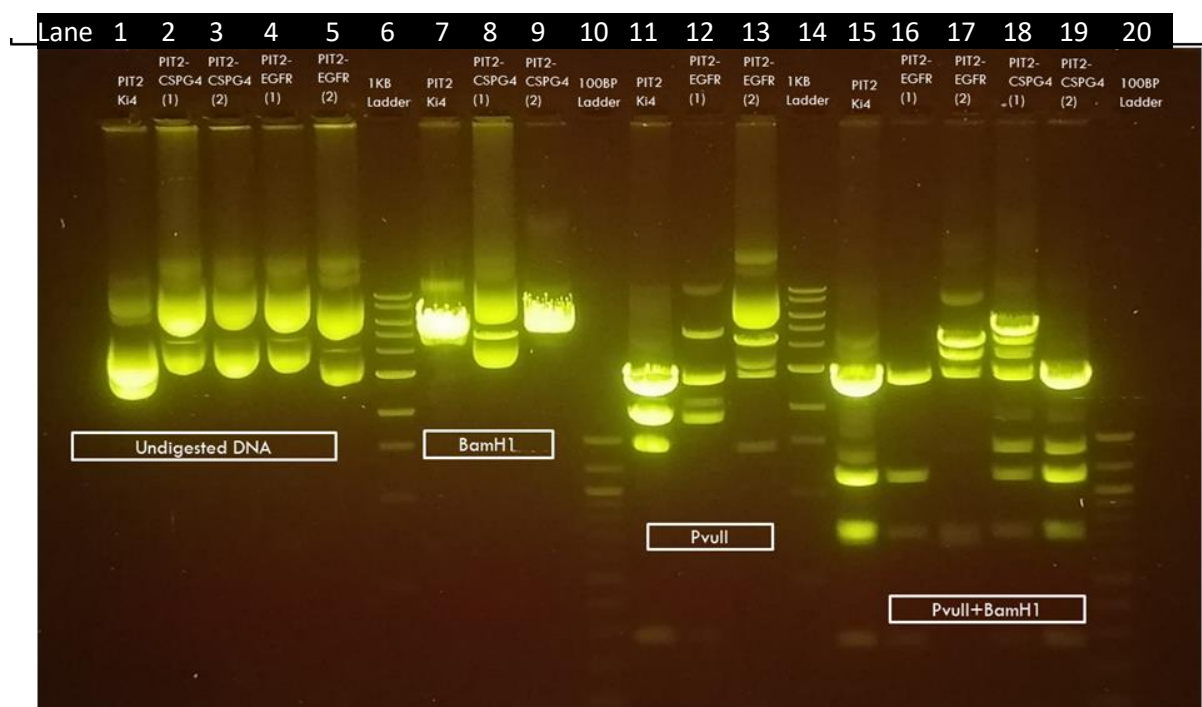


Figure 3.1.2.4.1: Analysis of restriction mapping of potential PIT2- α CSPG4 and PIT- α EGFR clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.

Based on the restriction analysis fragment pattern, one of the PIT2- α CSPG4 and one PIT2- α EGFR colonies seemed to be successfully cloned. These colonies were sent for sequencing to verify that the cloning was successful. The sequencing confirmed that these colonies had indeed been correctly cloned and thus the plasmids could be used for phage display.

Summary of section 3.1 results

We were able to use phage display to selectively isolate Ki4-antibodies demonstrating that the phage propagation method and selection was successful. Furthermore, we were successfully able to generate a two-component phage library, however I was unable to propagate the library due to a lack of infrastructure. One of the initial requirements of this project was to establish an enclosed area of the lab which would be dedicated specifically to phage display. The main reason for this is that my research was based in a bacterial culture lab, and phages easily infect bacteria thus the risk of contamination of doing phage work in a bacterial lab was too high. Unfortunately, building of the phage lab was significantly delayed with the national lockdown in 2019 and the lab was never developed. Once all the equipment and infrastructure are in place for phage work, this two-component library can be propagated and used to optimize a bio-panning strategy for selection of novel antibodies. The identification of novel disease-specific antibodies is forms an essential first step in the development of new therapeutics for TNBCs.

Section 2: Developing a differential diagnostic panel for the classification of TNBC tumours

The main objective of this section was to generate a panel of diagnostic fusion proteins which could be used collectively to characterise TNBC tumours into different treatment categories. Given the cardinal role tumour infiltrating cells play in the progression or inhibition of cancer, various T-cell and B-cell targeting proteins were included in this panel; those were CD3, CD4, CD8, CD19 and CD20 specific targets. Furthermore, with the recent progress made using checkpoint inhibitors, the two most important ones, that is PDL1 and B7.1 (CTLA4) were also included in the panel.

A combination of these targeting proteins would be useful in determining which patients may be susceptible to specific immunotherapies. A patient lacking the expression of PDL1 for example, would probably not benefit from checkpoint inhibition therapy targeting PDL1. Classifying the patients in this manner would increase the likelihood of providing beneficial treatment to the correct patient and reduce the administration of treatments that would likely not work in selected patients. This panel would be combined with the cancer antigen specific diagnostic panel to provide a reliable strategy for targeted therapy using the correct targets in the correct patients.

3.2.1 Identification and assembly of scFv sequences

The first step in developing the differential diagnostic panel was to conduct a comprehensive patent and literature search to find sources of antibody sequences targeting our receptors of interest. Table 3.2.1.1 below summaries the various sequences found and the source from which it was obtained.

Table 3.2.1.1: Summary of targets utilize in Differential Diagnostic Panel

Target	Indication	Construct	Reference	Original ligand type
CD19	B cells	α CD19-SNAP		
CD20 (B cells)	B cells	α CD20-SNAPf	2F2 - Patent US20140093454A1 - Teeling et al., Human monoclonal antibodies against CD20 (2014)	Human antibody
CD3	T cells	α CD3-SNAPf		Humanized antibody
CD4	Helper T cells	α CD4-SNAPf		Humanized antibody
CD8	Cytotoxic T cells	α CD8-SNAPf	CD8 OKT8_H1L1 - Patent WO2016105450A2 - Moore et al., Trispecific antibodies (2016)	Tetavalent trispecific antibody
CD80 (B7.1)	Checkpoint Inhibition - Binds to CTLA-4 on T cells	α B7-1-SNAP	16C10 - Patent EP1914301A1 Anderson et al. Anti-CD80 antibodies (2008)	Primatized antibody (monkey variable domains, human constant sequences)
PD-L1	Checkpoint Inhibition - Binds to PD-1 on T cells	α PD-L1-SNAPf	DrugBank (accession no. DB11595)	Humanized antibody (Atezolizumab) - FDA approved

The final assembled scFv sequences containing restriction sites, linker, CDRs and framework regions are depicted in table 3.1.2.2 below.

Table 3.2.1.2: Assembled scFv sequences. The *Sfi*1 recognition site and *Not*1 recognition site are highlighted in blue and orange respectively, while the linker sequences are highlighted in red.

Construct	Assembled Sequence
α-CD3	GCGGCCAGCCGGCCGAGGTGAAGCTGGTGGAGAGCGGCGGCCTGGTGCAGCCCGGC GGCAGCCTGAGACTGAGCTGCGCCGCCAGCGGCTTCACCTTCAACACCTACGCCATGAAGT GGTGAGACAGGCCCCGGCAAGGGCCTGGAGTGGGTGGCCAGAATCAGAAGCAAGTACAAC AACTACGCCACCTACTACGCCGACAGCGTGAAGGACAGATTACCATCAGCAGAGACGACAG CAAGAGCAGCCTGTACCTGCAGATGAACAACCTGAAGACCGAGGACACCGCCATGTACTACT GCGTGAGACACGGCAACTTCGGCAACAGCTACGTGAGCTGGTTCGCCTACTGGGGCCAGGGC ACCCTGGTGACCGTGAGCAGCGGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGC GGATCCAGGCGCTGGTGACCCAGGAGCCAGCTTCAGCGTGAGCCCCGGCGGCACCGTGAC CCTGACCTGCAGAAGCAGCACCGGCGCCGTGACCACCAGCAACTACGCCAAGTGGGTGCAGC AGACCCCGGCCAGGCCTTCAGAGGCCTGATCGGCGGCACCAACAAGAGAGCCCCGGCGTG CCCGCCAGATTACGCGGCAGCCTGATCGGCGACAAGGCCGCCCTGACCATACCGGCGCCCA GGCCGACGACGAGAGCATCTACTTCTGCGCCCTGTGGTACAGCAACCTGTGGGTGTTGCGCG GCGGCACCAAGCTGACCGTGCTGGCGGCCCACTCGAGTCTAGA
α-CD4	GCGGCCAGCCGGCCGAGGAACAGCTTGTGGAGTCTGGGGGAGGCTTGGTGAAACCCGGAG GTTCTCTGAGGCTCTCCTGTGCAGCCTCGGGTTTCAGTTTCAGTGACTGCCGGATGTACTGGCT TCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGATTGGTGTGATTTCAAGTCAAATCTGAGAATT ATGGAGCAAATTATGCAGAGTCTGTGAGGGGCGAGATTCACTATTTCAAGAGATGATTCAAAA AACACGGTCTATCTGCAGATGAACAGCTTGAAGACCGAAGACACTGCCGTTTATTATTGTAGT GCCTCTATTATAGGTACGACGTGGGGGCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTC ACTGTCTCTTCAAGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGGATCCGACA TCGTGATGACCCAGTCTCCAGACTCCCTGGCTGTGTCTCTGGGCGAGAGGGCCACCATCAACT GCAGGGCCAGCAAAAGTGTCAGTACATCTGGCTACAGTTATATATATTGGTACCAGCAGAAAC CAGGACAGCCTAAGCTGCTCATTTACCTTGATCCATCCTAGAATCTGGGGTCCCTGACCGATT CAGTGGCAGCGGGTCTGGGACAGATTTCACTCTACCATCAGCAGCCTGCAGGCTGAAGATG TGGCAGTTTATTACTGTACGACAGTAGGGAACCTCCGTGGACGTTGCGCCAAGGGACCAAG GTGGAATCAAAAGCGGCCCACTCGAGTCTAGA
α-CD8	GCGGCCAGCCGGCCGAGGTGCAGCTGCAGCAGAGCGGCGCCGAGGTGAAGAAGCCCGGC GCCAGCGTGAAGGTGAGCTGCAAGGCCAGCGGCTTCAACATCAAGGACACCTACATCCACTG GGTGAGACAGGCCCCGGCAAGGGCCTGGAGTGGATGGGCGAATCGACCCCGCCAACGAC AACACCCTGTACGCCAGCAAGTTCAGGGCAGAGTGACCATCACCGCCGACACCAGCACCAA CACCGCCTACATGGAGCTGAGTAGCCTGAGAAGCGAGGACACCGCCGTGTACTACTGCGGCA GAGGCTACGGCTACTACGTGTTGACCACTGGGGCCAGGGCACCACCGTGACCGTGAGCAGC GGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGGATCCGACATCAAGATGACC CAGAGCCCCAGCAGCCTGAGCGCCAGCGTGGGCGACAGAGTGACCATCACCTGCAGAACCA GCAGAAGCATCAGCCAGTACCTGGCCTGGTACCAGGAGAAGCCCGGCAAGACCAACAAGCT GCTGATCTACAGCGGCAGCACCTGCAGAGCGGCATCCCCAGCAGATTACGCGGCAGCGGCA GCGGCACCGACTTCACCCTGACCATCAGCAGCCTGCAGCCCCGAGGACTTCGCCACCTACTACT GCCAGCAGCACAACGAGAACCCCCTGACCTTCGGCGCCGGCACCAAGCTGGAGATCAAGGCG GCCGCACTCGAGTCTAGA
α-CD19	GGCCAGCCGGCCGAGGTGCAGCTGGTGGAGAGCGGCGGCCTGGTGCAGCCCGGCGG CAGCCTGAGACTGAGCTGCGCCGCCAGCGGCTTCACCTTCAGCAGCAGCTGGATGAAGTGGG TGAGACAGGCCCCGGCAAGGGCCTGGAGTGGGTGGGCGAATCTACCCCGGCGACGGCGA

	CACCAACGGCAAGTTCAAGGGCAGATTACCATCAGCAGAGACGACAGCAAGAACAGCCTGT ACCTGCAGATGAACAGCCTGAAGACCGAGGACACCGCCGTGTACTACTGCGCCAGAAGCGGC TTCATCACCAACCGTGCTGGACTTCGACTACTGGGGCCAGGGCACCTGGTGACCGTGAGCAG CGGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGGATCCGAGATCGTGCTGAC CCAGAGCCCCGACTTCCAGAGCGTGACCCCCAAGGAGAAGGTGACCATCACCTGCAGAGCCA GCGAGAGCGTGACACCTTCGGCATCAGCTTCATGAACTGGTTCAGCAGAAGCCCCGACCAG AGCCCCAAGCTGCTGATCCACGCCGCCAGCAACCAGGGCAGCGGCGTGCCAGCAGATTTCAG CGGCAGCGGCAGCGGCACCGACTTCACCCTGACCATCAACAGCCTGGAGGCCGAGGACGCCG CCACCTACTTCTGCCAGCAGAGCAAGGAGGTGCCCTTCACCTTCGGCGGCGGCACCAAGGTG GACATCAAGCGGCCGC
α-CD20	GCGGCCAGCCGGCCATGGAGCTGGGCCTGAGCTGGATCTTCCTGCTGGCCATCCTGAAGGG CGTGCACTGCGAGGTGCAGCTGGTGGAGAGCGGCGGCGGCCTGGTGACCCCGGCAGAAG CCTGAGACTGAGCTGCGCCGCCAGCGGCTTACCTTCAACGACTACGCCATGCACTGGGTGA GACAGGCCCCCGGCAAGGGCCTGGAGTGGGTGAGCACCATCAGCTGGAACAGCGGCAGCAT CGGCTACGCCGACAGCGTGAAAGGGCAGATTACCATCAGCAGAGACAACGCCAAGAAGAGC CTGTACCTGCAGATGAACAGCCTGAGAGCCGAGGACACCGCCCTGTACTACTGCGCCAAGGA CATCCAGTACGGCACTACTACTACGGCATGGACGTGTGGGGCCAGGGCACCACCGTGACCG TGAGCAGCGGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGGATCCATGGAG GCCCCGCCCAGCTGCTGTTCTGCTGCTGTGGCTGCCCCGACACCACGGCGAGATCGTG CTGACCCAGAGCCCCGCCACCCTGAGCCTGAGCCCCGGCGAGAGAGCCACCCTGAGCTGCAG AGCCAGCCAGAGCGTGAGCAGCTACCTGGCCTGGTACCAGCAGAAGCCCGGCCAGGCCCCCA GACTGCTGATCTACGACGCCAGCAACAGAGCCACCGGCATCCCCGCCAGATTTCAGCGGCAGC GGCAGCGGCACCGACTTCACCCTGACCATCAGCAGCCTGGAGCCCCGAGGACTTCGCCGTGTA CTACTGCCAGCAGAGAAGCAACTGGCCCATCACCTTCGGCCAGGGCACCAGACTGGAGATCA AGCGGCCGCACTC
α-PDL1	GCGGCCAGCCGGCCGAGGTGCAGCTGGTGGAGAGCGGCGGCGGCCTGGTGACCCCGGC GGCAGCCTGAGACTGAGCTGCGCCGCCAGCGGCTTACCTTCAGCGACAGCTGGATCCACTG GGTGAGACAGGCCCCCGGCAAGGGCCTGGAGTGGGTGGCCTGGATCAGCCCCTACGGCGGC AGCACCTACTACGCCGACAGCGTGAAAGGGCAGATTACCATCAGCGCCGACACCAGCAAGAA CACCGCCTACCTGCAGATGAACAGCCTGAGAGCCGAGGACACCGCCGTGTACTACTGCGCCA GAAGACACTGGCCCCGGCGGCTTCGACTACTGGGGCCAGGGCACCTGGTGACCGTGAGCAG CGGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGGATCCGACATCCAGATGAC CCAGAGCCCCAGCAGCCTGAGCGCCAGCGTGGGCGACAGAGTGACCATCACCTGCAGAGCCA GCCAGGACGTGAGCACCGCCGTGGCCTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCT GCTGATCTACAGCGCCAGCTTCTGTACAGCGGCGTGCCAGCAGATTTCAGCGGCAGCGGCA GCGGCACCGACTTCACCCTGACCATCAGCAGCCTGCAGCCCCGAGGACTTCGCCACCTACTACT GCCAGCAGTACCTGTACCACCCGCCACCTTCGGCCAGGGCACCAAGGTGGAGATCAAGCGC GCCGA
α-B7.1	GCGGCCAGCCGGCCATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGCAGTCCCAGATGG GTCCTGTCCCAGGTGCAGCTGCAGGAGTCGGGCCCAGGACTGGTGAAGCCTTCGGAGACCCT GTCCCTCACCTGCGCTGTCTCTGGTGGCTCCATCAGCGGTGGTTATGGCTGGGGCTGGATCCG CCAGCCCCCAGGGAAGGGGCTGGAGTGGATTGGGAGTTTCTATAGTAGTAGTGGGAACACCT ACTACAACCCCTCCCTCAAGAGTCAAGTCACCATTTCAACAGACACGTCCAAGAACCAGTTCTC CCTGAAGCTGAACTCTATGACCGCCGCGGACACGGCCGTGTATTACTGTGTGAGAGATCGTCT TTTTTCAGTTGTTGGAATGGTTTACAACAACCTGGTTCGATGTCTGGGGCCCGGGAGTCTGGT CACCGTCTCCTCAGTAGTGGAGGTGGCGGCTCCGGAGGTGGAGGCAGCGGAGGGGGCGG ATCCATGAGGGTCCCCGCCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAGGTGCACGATG TGAGTCTGTCTGACACAGCCGCCCTCAGTGTCTGGGGCCCCAGGGCAGAAGGTACCATCTC GTGCACTGGGAGCACCTCCAACATTGGAGGTTATGATCTACATTGGTACCAGCAGTCCCAGG AACGGCCCCCAAACCTCCTCATCTATGACATTAACAAGCGACCCTCAGGAATTTCTGACCGATT TCTGGCTCCAAGTCTGGTACCGCGGCTTCCTGGCCATCACTGGGCTCCAGACTGAGGATGAG

GCTGATTATTACTGCCAGTCCTATGACAGCAGCCTGAATGCTCAGGTATTCGGAGGAGGGACC CGGCGGCCGCACTC

3.2.2 In-silico Cloning

In-silico cloning was used to simulate the molecular cloning strategy to determine if the assembled scFv's design in section 3.2.1 above would successfully be incorporated into the pCB_snap vector backbone. The CLC genomics workbench application was used to align and compare the reliability of the V genes before proceeding to *in silico* cloning.

The resulting plasmids were then checked to make sure that the scFv would be in frame with the vector promoter after cloning, ensuring proper transcription and translation of the corresponding protein. The resulting constructs were also scanned for inappropriate stop codons to make sure that the full-length protein would be transcribed. The cloning strategy and resulting plasmid maps are depicted below.

Included in the ORF of each construct is an enterokinase recognition sequence which plays an essential role in generating the final active product by removing the N-terminal pre-peptide. The Ig-Kappa leader sequence and poly HIS-tag are located at the N-terminus of the SNAP-tag while the scFv sequence is located at the 3'.

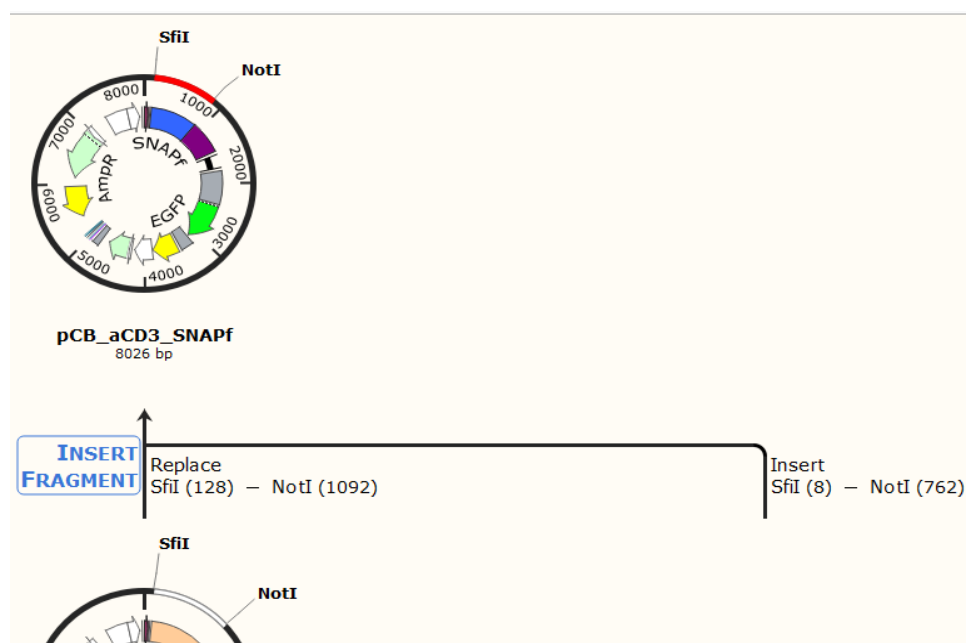


Figure 3.2.2.1: Insilico-cloning strategy using α CD3 scFv as a representative target. Sfi1 and Not1 restriction enzymes were used to replace the annexin scFv with other scFvs of interest.

In-silico cloning in Snapgene software confirmed that the assembled scFv genes would be compatible for ligation into the pCB-snap backbone. After the final assembly each construct was checked for the presence of inappropriate stop codons and translated into its amino-acid sequence which was aligned to its original parental (patent) sequence. This was to validate the integrity of the original CDRs and FRs prior to having the sequences synthesised at Genescript (USA).

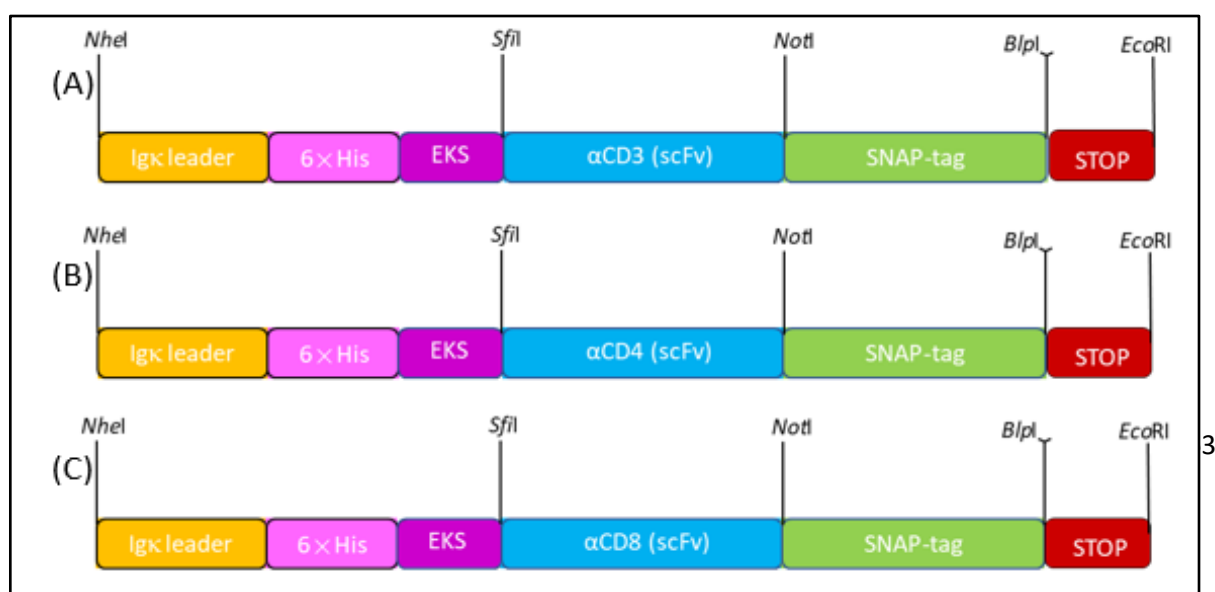


Figure 3.2.2.2 : Open reading frames of (A) α CD3-SNAP (B) α CD4-SNAP (C) α CD8-SNAP (D) α CD19-SNAP (E) α CD20-SNAP (F) α PDL1-SNAP and (G) α B7.1-SNAP. The different elements present on all ORFs include IgK leader – Ig Kappa leader, 6xHIS- N-terminus poly histidine, EKS – enterokinase site, SNAP-tag and stop codon. The scFv in each ORF is flanked by a *Sfi1* and *Not1* recognition site which is used during the cloning of each construct.

3.2.3 Molecular cloning of different recombinant SNAP fusion proteins

A diverse set of experiments were performed collectively to construct and amplify the DNA fragment of interest. The cloning of all constructs were done with restriction enzyme digest and T4 DNA ligation.

3.2.3.1 Bulk preparation of DNA plasmids by transformation

Plasmids received from Genescript containing the scFv genes of interest together with the pCB-AnnV-SNAP backbone were all individually bulk prepared by transformation into DH5 α cells. The cultures were grown overnight at 37 °C and the plasmid DNA from each culture was then individually isolated with the phenol/chloroform DNA isolation method. Extracted DNA was then quantified by spectrophotometry as depicted in table 3.2.3.1.1 below.

Table 3.2.3.1.1: Purity of DNA extracted from bulk preparations.

Plasmid	A260/230	A260/280	Concentration (ng/ul)
Puc57_ α CD3	2.201	1.802	34276
Puc57_ α CD4	2.111	1.875	18689
Puc57_ α CD8	2.001	1.902	29340
Puc57_ α CD19	2.201	1.953	35183
Puc57_ α CD20	2.014	1.875	25910
Puc57_ α B7.1	2.105	1.864	23504
Puc57_ α PDL1	2.138	1.863	20945
pCB_AnnV_Snap	2.159	1.872	48957

The A260/230 ratio is used to determine if DNA samples contain contaminating organic compounds. The generally accepted values for pure DNA samples are between 2.0 and 2.2. All isolated constructs listed in table above were within the acceptable range indicating that the DNA does not contain organic compounds. The A260/280 ratio is more routinely used to determine the purity of a DNA sample. The ideal ratio for DNA is 1.8 but anything between 1.7-2 is acceptable. All extracted constructs seemed to be pure as indicated by A260/280 ratios thus were used for cloning.

3.2.3.2 Restriction enzyme digest and agarose gel electrophoresis

For construction of the recombinant SNAP-tag fusion proteins, puc57 plasmids housing the scFv genes and the pCB_AnnV_snap vector backbone were both double digested using *Sfi*1 and *Not*1. The digestion allowed for the introduction of sticky ends on each scFv gene and the pCB backbone now lacking the AnnV scFv sequence in preparation for ligation. The restriction digested samples were then run on a 1.2% agarose gel to allow for visualization of the different DNA fragments.

Table 3.2.3.2.1: Description of agarose gel lanes depicted in figure 3.2.3.2.1

Lane	Construct	Enzyme/control	Expected fragment size/s (bp)
1	Puc57_αCD20	<i>Sfi1</i> + <i>Not1</i>	2710 + 753
2	Puc57_αCD20	<i>Sfi1</i> + <i>Not1</i>	2710 + 753
3	Puc57_αCD20	<i>Sfi1</i>	3463
4	Puc57_αCD20	<i>Not1</i>	3463
5	1KB molecular weight marker	NA	Multiple
6	pCB_αAnnV_SnapF	<i>Sfi1</i> + <i>Not1</i>	7255 + 978
7	pCB_αAnnV_SnapF	<i>Sfi1</i> + <i>Not1</i>	7255 + 978
8	pCB_αAnnV_SnapF	<i>Sfi1</i> + <i>Not1</i>	7255 + 978
9	pCB_αAnnV_SnapF	<i>Sfi1</i> + <i>Not1</i>	7255 + 978
10	pCB_αAnnV_SnapF	<i>Sfi1</i>	8233
11	pCB_αAnnV_SnapF	<i>Not1</i>	8233
12	1KB molecular weight marker	NA	Multiple
13	Puc57_αPDL1	<i>Sfi1</i> + <i>Not1</i>	2710 + 741
14	Puc57_αPDL1	<i>Sfi1</i> + <i>Not1</i>	2710 + 741
15	Puc57_αPDL1	<i>Sfi1</i>	3451

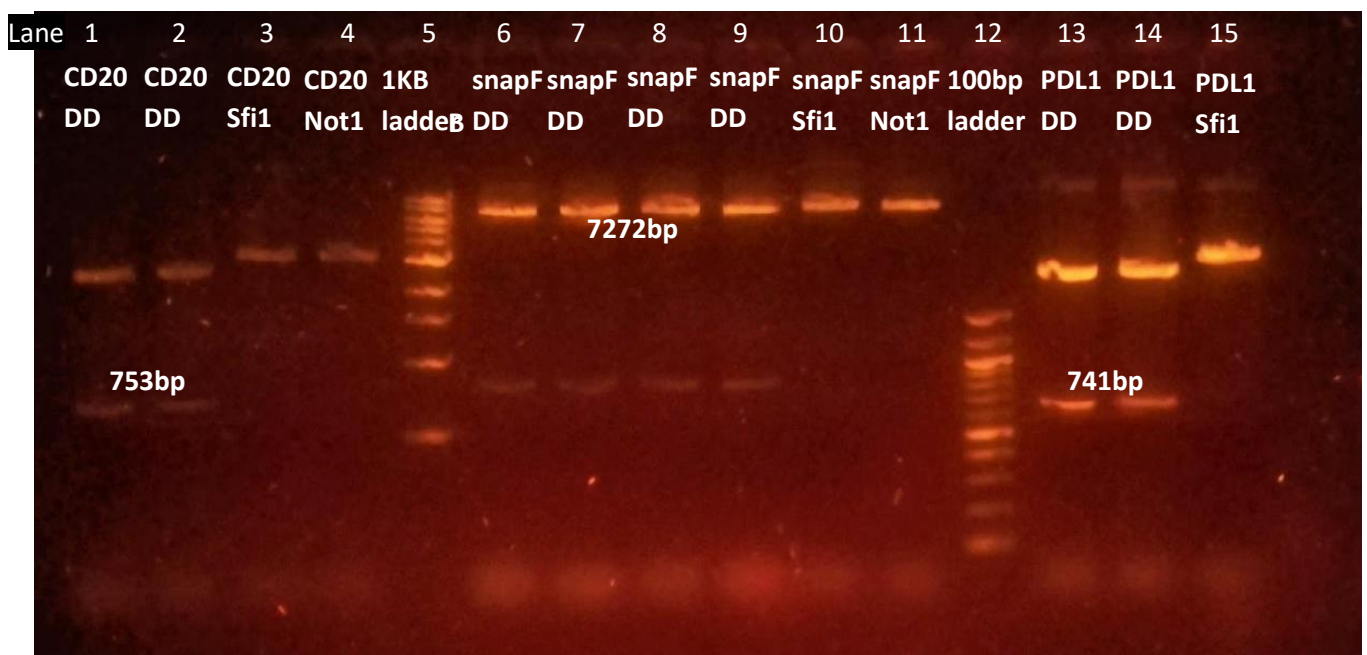


Figure 3.2.3.2.1: Agarose gel depicting DNA fragments after *sfi1* and *Not1* double digest (DD). 753bp αCD20, 741bp αPDL1 and 7272bp snapF plasmids was isolated for ligation.

Table 3.2.3.2.2: Description of agarose gel lanes depicted in figure 3.2.3.2.2

Lane	Construct	Enzyme/control	Expected fragment size/s (bp)
1	Puc57_αCD4	<i>Sfi1</i> + <i>Not1</i>	2710 + 768
2	Puc57_αCD4	<i>Sfi1</i> + <i>Not1</i>	2710 + 768
3	Puc57_αCD4	<i>Sfi1</i>	3478
4	Puc57_αCD4	<i>Not1</i>	3478
5	1KB molecular weight marker	NA	Multiple
6	Puc57_αCD3	<i>Sfi1</i> + <i>Not1</i>	2710 + 768
7	Puc57_αCD3	<i>Sfi1</i> + <i>Not1</i>	2710 + 768
8	Puc57_αCD3	<i>Sfi1</i>	3478
9	Puc57_αCD3	<i>Not1</i>	3478
10	1KB molecular weight marker	NA	Multiple
11	Puc57_αPDL1	<i>Not1</i>	3451
12	Puc57_αCD8	<i>Sfi1</i> + <i>Not1</i>	2710 + 741
13	Puc57_αCD8	<i>Sfi1</i> + <i>Not1</i>	2710 + 741
14	Puc57_αCD8	<i>Sfi1</i>	3451
15	Puc57_αCD8	<i>Not1</i>	3451

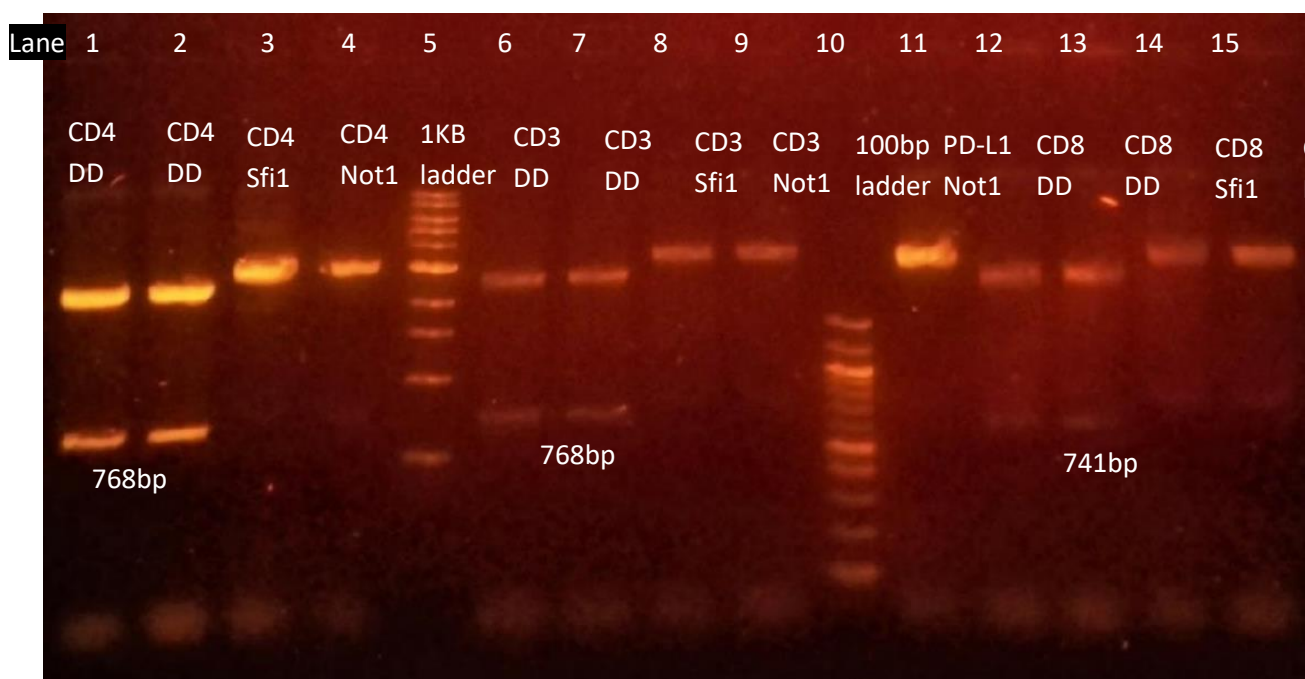


Figure 3.2.3.2.2: Agarose gel depicting DNA fragments after *sfi1* and *Not1* double digest (DD). 768bp αCD4, 768bp αCD3 and 741bp αCD8 plasmids were isolated for ligation.

The *Sfi1* and *Not1* digests were successful for all constructs as depicted in agarose gels in figures 3.2.3.2.1 and 3.2.3.2.2. Following the agarose gel electrophoresis, DNA bands containing the various scFv or pCB-backbone were excised from the gel and purified with Qiagen quick gel extraction kit. The quantity and purity of each DNA sample was then determined by spectrophotometry.

3.2.3.3 T4 DNA Ligation

T4 DNA ligation was used to ligate the scFv fragments and the pCB backbone. T4 DNA ligase was chosen for its capacity to add 3' hydroxyl groups and 5' phosphate groups to the cohesive ends of restriction digested DNA. Once ligated the mixtures were then transformed into chemically competent DH5 α *E. coli* cells and spread on LB-agar plates containing 200 μ g/ml ampicillin.

No bacterial growth was observed on the bacteria only contamination control plate shown in figure 3.2.3.3.1 below, indicating that the colonies observed on the other plates are true transformants. The 'vector only' (digested vector backbone) plate showed presence of a few colonies, indicating that a small amount of vector relegation had occurred. Three colonies were picked at random from each plate for analysis by restriction analysis.

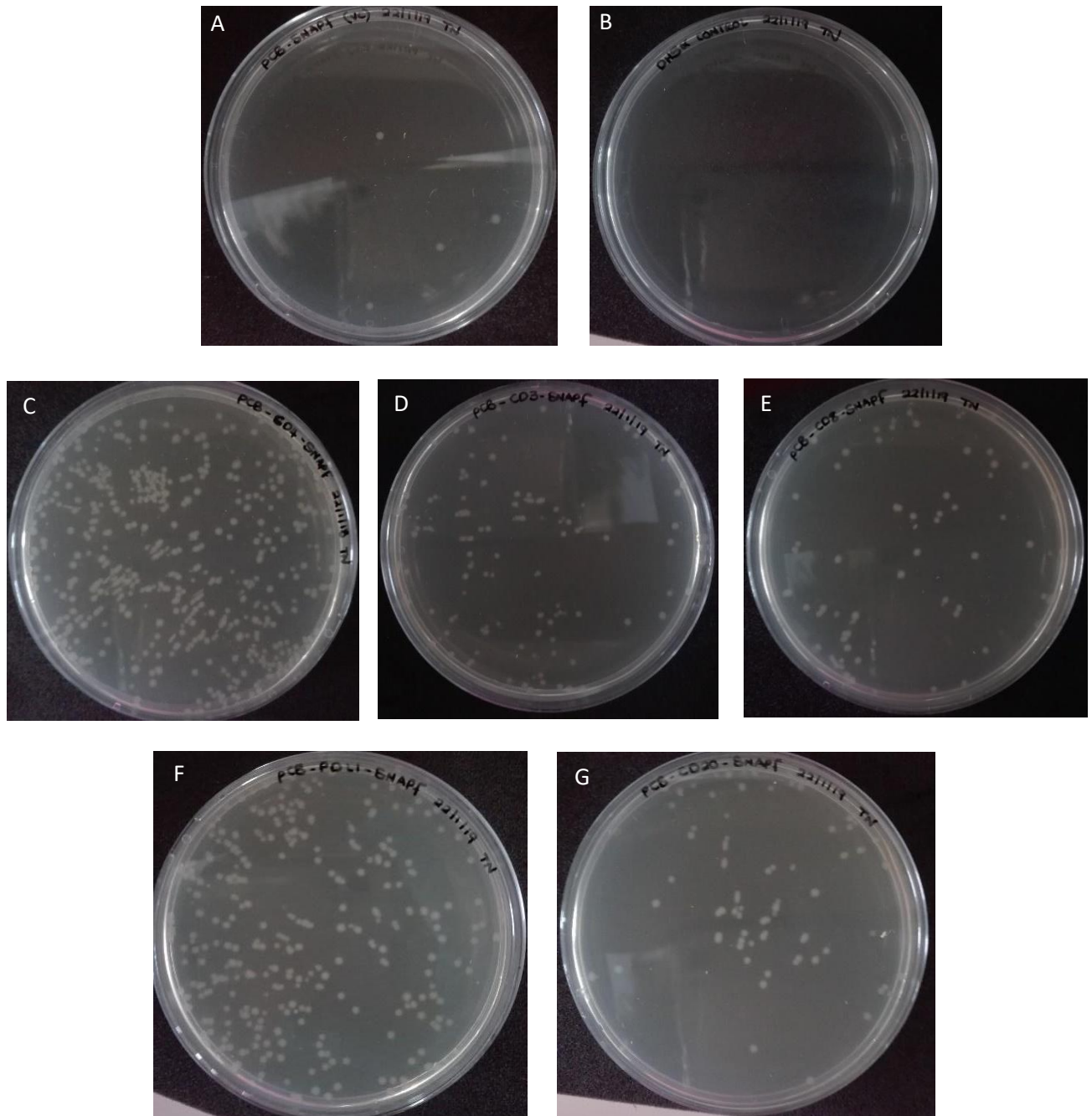


Figure 3.2.3.3.1: Growth of *E. coli* cells transformed with potential recombinant (C) pCB-CD4(scFv)-snapF, (D) pCB-CD3(scFv)-snapF, (E) pCB-CD8(scFv)-snapF, (F) pCB-PDL1(scFv)-snapF and (G) pCB-CD20(scFv)-snapF plasmid DNA. Vector only (A) and (B) bacterial contamination plates are also depicted. Following transformation, 100 μ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 μ g/ml.

3.2.3 Restriction Mapping of Potential Recombinant Clones

Following molecular cloning, colonies picked from LB agar plates were cultured in 12ml culture tubes overnight. The Zyppy™ plasmid miniprep kit was then used to extract DNA from each culture. The purified DNA was then quantified with spectrophotometry and subjected to restriction mapping with different restriction enzyme. The enzymes used for digestion of each construct was chose using an agarose gel simulation on Snapgene

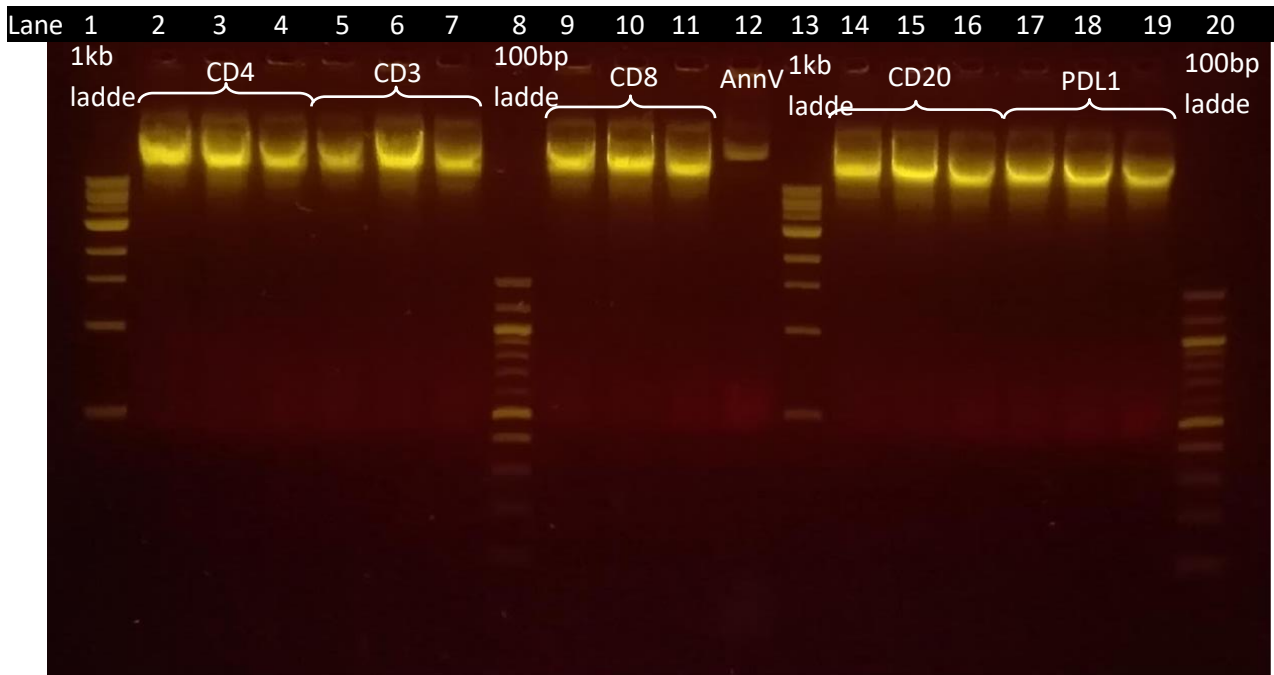


Figure 3.2.3.2 . Analysis of restriction mapping of undigested pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All clones were separated using a 1.2% (w/v) agarose gel ran for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting gel image after electrophoresis.

Table 3.2.3.2 Description of agarose gel depicted in figure 3.2.3.2

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	Clone 1 (CD4)	Undigested	8026
3	Clone 2 (CD4)	Undigested	8026
4	Clone 3 (CD4)	Undigested	8026
5	Clone 1 (CD3)	Undigested	8026
6	Clone 2 (CD3)	Undigested	8026
7	Clone 3 (CD3)	Undigested	8026
8	100 KB DNA ladder	NA	Multiple
9	Clone 1 (CD8)	Undigested	7999
10	Clone 2 (CD8)	Undigested	7999
11	Clone 3 (CD8)	Undigested	7999
12	pCB- α AnnV-snap	Undigested	8236
13	1KB DNA Ladder	NA	Multiple
14	Clone 1 (CD20)	Undigested	8011
15	Clone 2 (CD20)	Undigested	8011
16	Clone 3 (CD20)	Undigested	8011
17	Clone 1 (PDL1)	Undigested	7999
18	Clone 2 (PDL1)	Undigested	7999
19	Clone 3 (PDL1)	Undigested	7999
20	100 KB DNA Ladder	NA	Multiple

Restriction digest reactions containing all reagents except for the actual restriction digest enzyme were set up as controls. Figure 3.2.3.2 shows the undigested DNA run on an agarose gel. This gel is used as a control, to make sure that there is no non-specific digestion with any of our restriction mapping reagents. Figure 3.2.3.3 depicts that a differential banding pattern would be seen between the various newly constructed clones and the original pCB-AnnV-snap DNA. Thus BamH1 was used to digest each of the clones alongside the digestion of the pCB-AnnV-Snap construct.

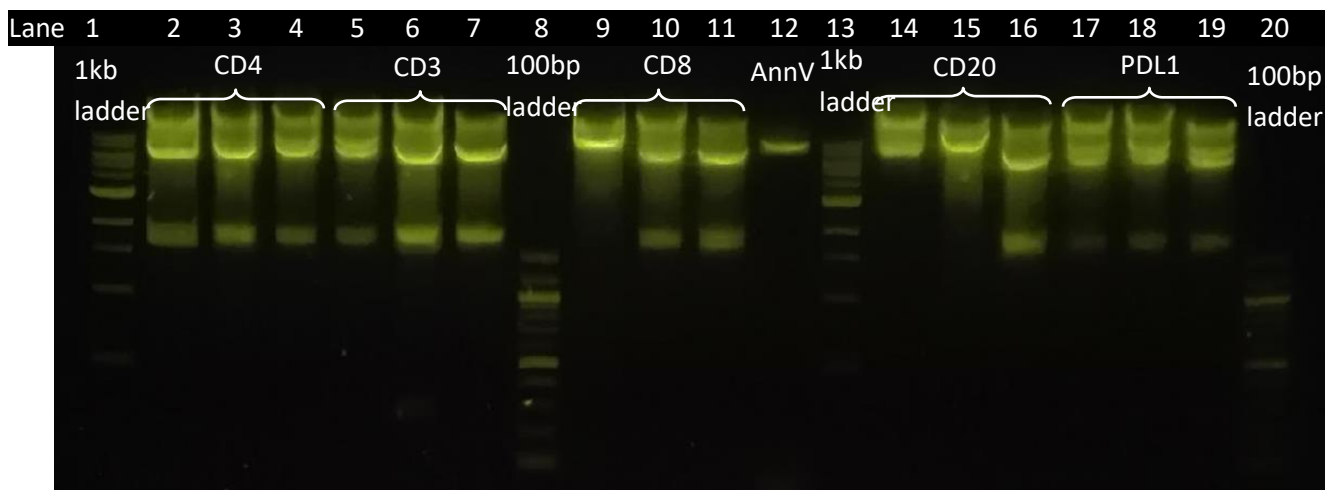


Figure 3.2.3.4: Analysis of restriction mapping of BamH1 digested potential pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All digests were run at 37°C overnight. The DNA was separated using a 1.2% (w/v) agarose gel ran for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting gel image after electrophoresis.

The banding patterns on the agarose gel in figure 3.2.3.4 suggest that all three potential colonies of pCB- α CD4-snap, pCB- α CD3-snap and pCB- α PDL1-snap were all successfully cloned. Two pCB-CD8-snap and just one pCB-CD20-snap colony seemed to be successfully cloned. This data was verified using a second enzyme, Pst1. As with BamH1, Pst1 was chosen based on the differential banding pattern seen between all the newly formed constructs compared to the original pCB-AnnV-Snap construct as depicted in the simulated agarose gel in figure 3.2.3.6 below.

Table 3.2.3.4: Description of agarose gel in figure 3.2.3.4.

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	Clone 1 (CD4)	BamH1	6189, 1837
3	Clone 2 (CD4)	BamH1	6189, 1837
4	Clone 3 (CD4)	BamH1	6189, 1837
5	Clone 1 (CD3)	BamH1	6192, 1834
6	Clone 2 (CD3)	BamH1	6192, 1834
7	Clone 3 (CD3)	BamH1	6192, 1834
8	100 KB DNA ladder	NA	Multiple
9	Clone 1 (CD8)	BamH1	6171, 1828
10	Clone 2 (CD8)	BamH1	6171, 1828
11	Clone 3 (CD8)	BamH1	6171, 1828
12	pCB- α AnnV-snap	BamH1	8236
13	1KB DNA Ladder	NA	Multiple
14	Clone 1 (CD20)	BamH1	6183, 1828
15	Clone 2 (CD20)	BamH1	6183, 1828
16	Clone 3 (CD20)	BamH1	6183, 1828
17	Clone 1 (PDL1)	BamH1	5875, 1828, 296
18	Clone 2 (PDL1)	BamH1	5875, 1828, 296
19	Clone 3 (PDL1)	BamH1	5875, 1828, 296
20	100 KB DNA Ladder	NA	Multiple

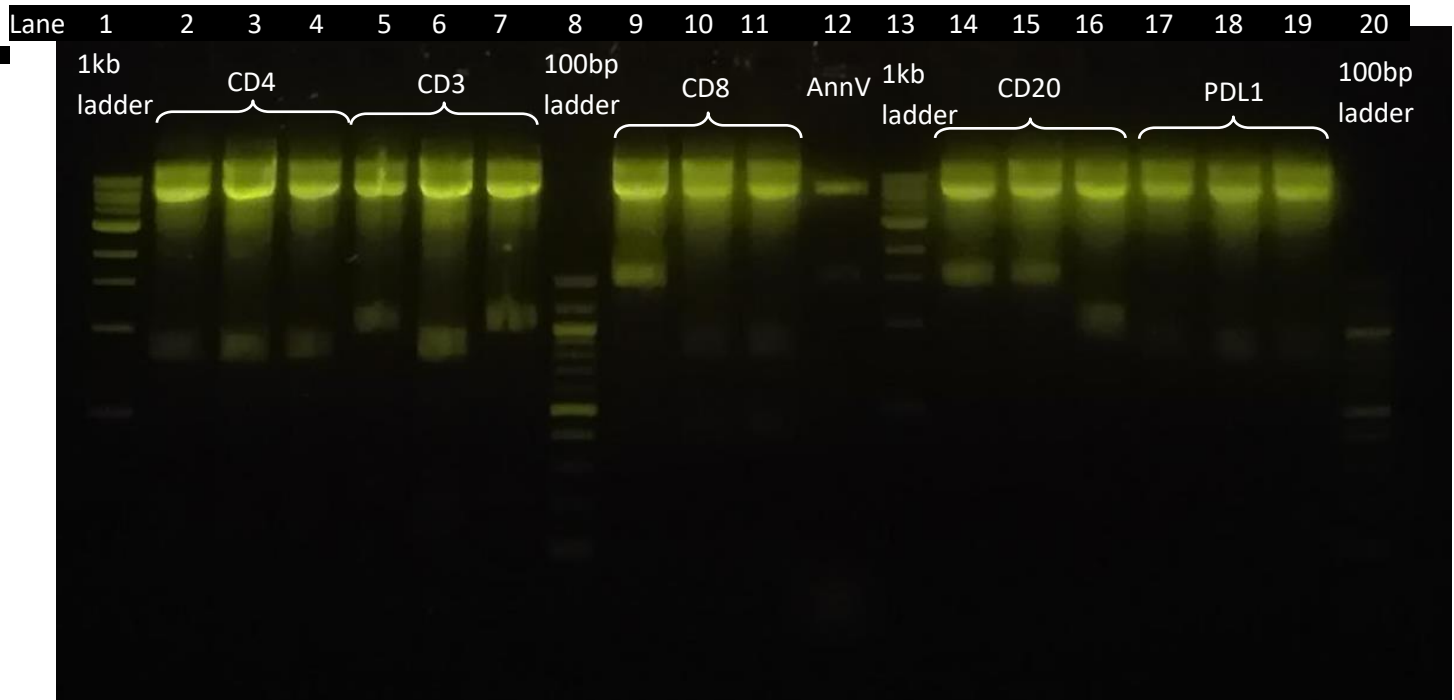


Figure 3.2.3.6: Analysis of restriction mapping of Pst1 digested, potential pCB- α CD4-snap, pCB- α CD3-snap, pCB- α CD8-snap, pCB- α CD20-snap and pCB- α PDL1-snap clones using agarose gel electrophoresis. All digests were run at 37°C overnight. The DNA was separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. The figure depicts the resulting agarose gel after electrophoresis.

Table 3.2.3.6: Description of agarose gel in figure 3.2.3.6:

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	Clone 1 (CD4)	Pst1	6650, 965, 236, 175
3	Clone 2 (CD4)	Pst1	6650, 965, 236, 175
4	Clone 3 (CD4)	Pst1	6650, 965, 236, 175
5	Clone 1 (CD3)	Pst1	6650, 1140, 236
6	Clone 2 (CD3)	Pst1	6650, 1140, 236
7	Clone 3 (CD3)	Pst1	6650, 1140, 236
8	100 KB DNA ladder	NA	Multiple
9	Clone 1 (CD8)	Pst1	6413, 965, 455, 94, 72
10	Clone 2 (CD8)	Pst1	6413, 965, 455, 94, 72
11	Clone 3 (CD8)	Pst1	6413, 965, 455, 94, 72
12	pCB- α AnnV-snap	Pst1	6561, 1675
13	1KB DNA Ladder	NA	Multiple
14	Clone 1 (CD20)	Pst1	6644, 1131, 236
15	Clone 2 (CD20)	Pst1	6644, 1131, 236
16	Clone 3 (CD20)	Pst1	6644, 1131, 236
17	Clone 1 (PDL1)	Pst1	6644, 965, 224, 166
18	Clone 2 (PDL1)	Pst1	6644, 965, 224, 166
19	Clone 3 (PDL1)	Pst1	6644, 965, 224, 166
20	100 KB DNA Ladder	NA	Multiple

One or two potential transformants were chosen based on the restriction mapping for each construct. Chosen clones were bulk prepared and sent for Sanger sequencing to confirm if they were true transformants and contain the correct insert sequence.

3.2.4 Sequence Analysis of Selected Recombinant clones

Recombinant clones which showed restriction digest patterns that most closely resembled the simulation patterns generated on Snapgene were sent to Inqaba Biotechnical Industries (South Africa) for Sanger DNA sequencing. Sanger sequencing is used to determine the specific order of nucleotide bases in a DNA molecule. SnapGene software was then used to confirm the sequence of the cloned products. This was done by aligning the resulting sanger sequence against the original *in silico* designed sequences.

The alignment results are depicted in figure 3.2.4.1 below. The solid arrows indicate that there was a 100% homology between the cloned constructs and the *in silico* designed sequences confirming that the cloning was successful. The sequencing was also checked for point mutations, deletions and insertions prior to using the DNA for transfection.

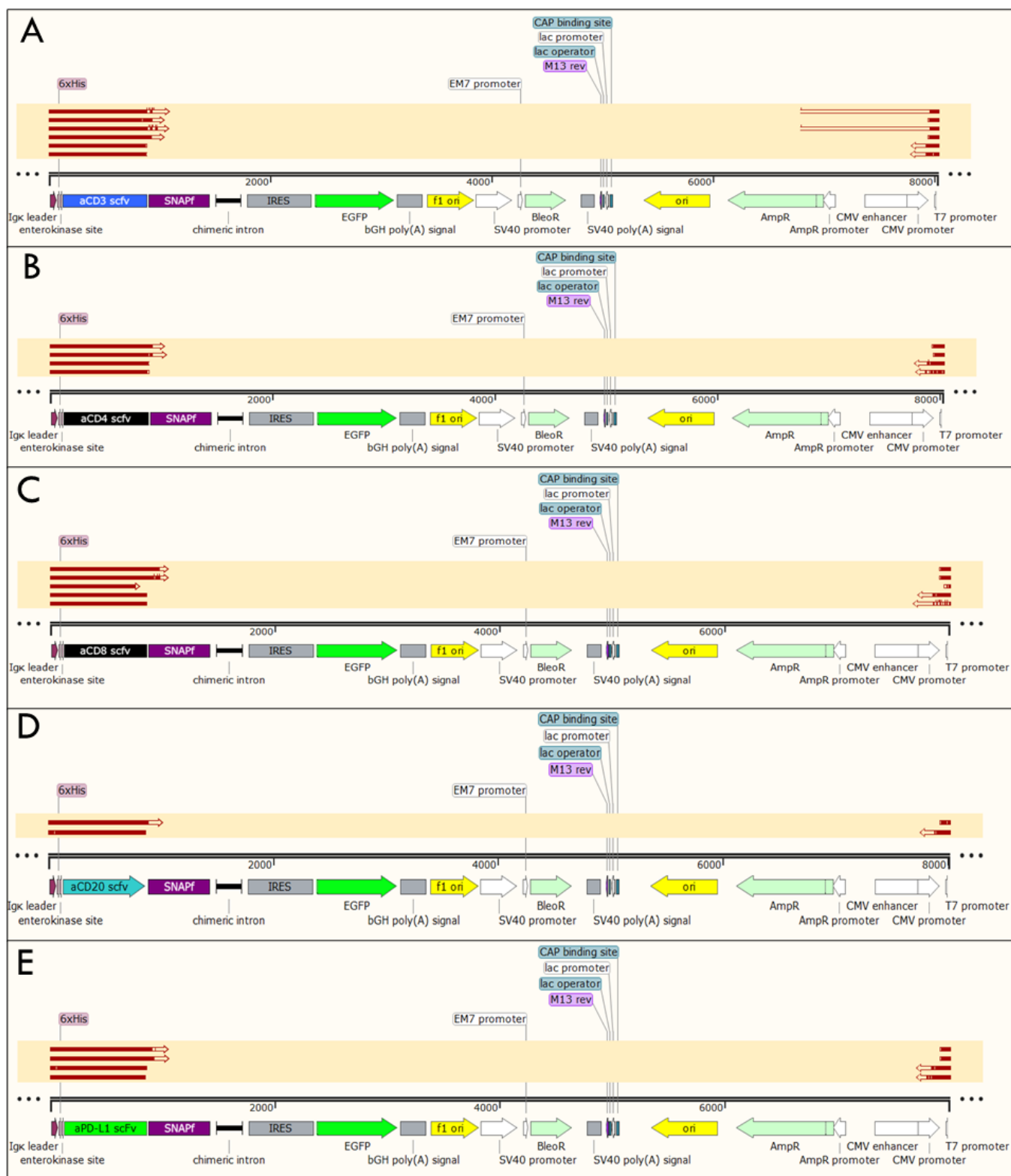


Figure 3.2.4.1: Alignment map generated following the alignment of trace sequences with the corresponding in silico sequence for (A) pCB-αCD3-snap (B) pCB-αCD4-snap (C) pCB-αCD8-snap (D) pCB-αCD20-snap and (E) pCB-αPDL1-snap. The clones corresponded with 100% homology to the in silico ORF sequences as shown by the dark red arrows.

3.2.5: Transfection into HEK293T cells for protein expression

Successfully cloned recombinant plasmids were introduced into HEK293T cells for protein expression. The HEK293T mammalian expression system was chosen as they have a stably transformed SV40 large T-antigen which binds to the SV40 promoter in the plasmid allowing for expression of the recombinant proteins. Successfully transfected cells (potentially expressing the putative fusion proteins) were identified through the visualization of eGFP using a ZOE™ Fluorescent Cell Imager (Bio-Rad Laboratories, UK). Transfection efficiencies of each culture was determined by flow cytometry as shown in figure 3.2.5.1 below.

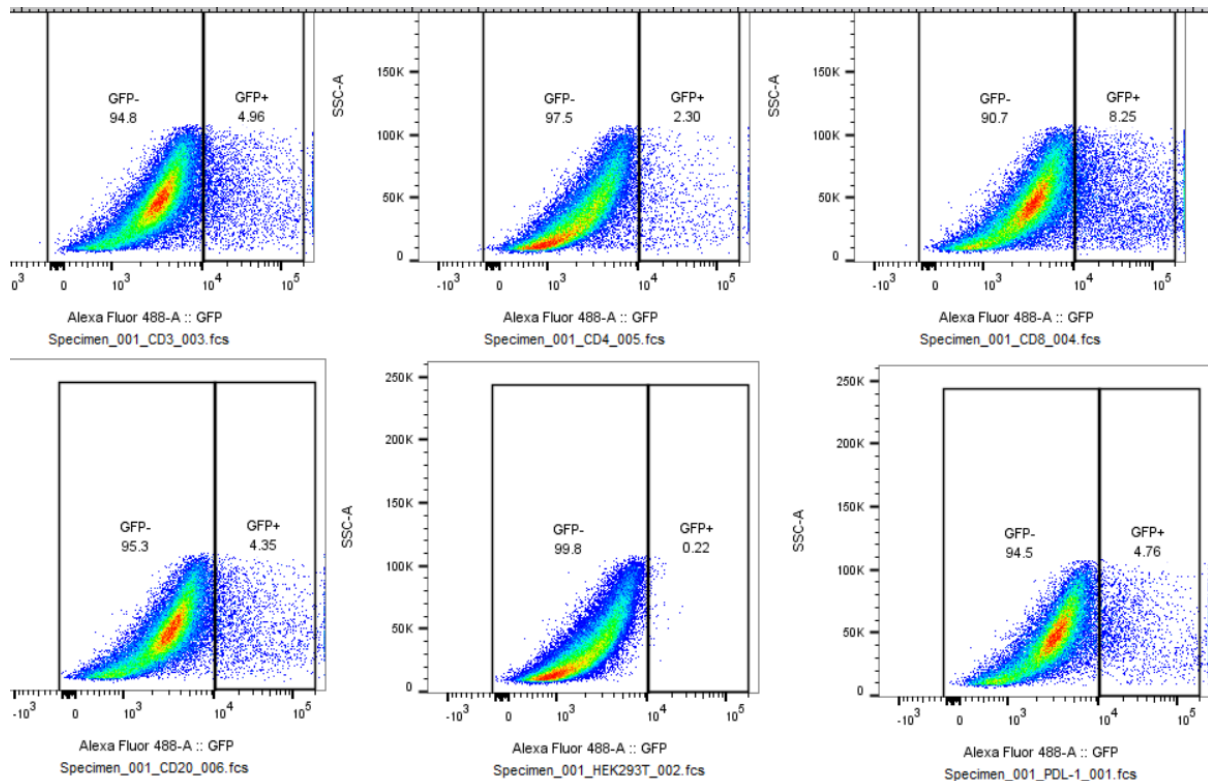


Figure 3.2.5.1: Dot plots depicting the eGFP+ and eGFP- cell populations in (A) pCB- α CD3-snap (B) pCB- α CD4-snap (C) pCB- α CD8-snap (D) pCB- α CD20-snap (E) untransfected HEK293T and (F) pCB- α PDL1-snap cultures.

As seen in figure 3.2.5.1 the transfection efficiencies of all cultures were below 10%. These are low in comparison to standard transfection efficiencies, which are usually around 70%. Due to the low transfection efficiencies an aggressive amount of Zeocin (600 μ g/ml) was applied to the cultures to enrich for the eGFP-positive cell populations carrying the bleomycin resistance gene.

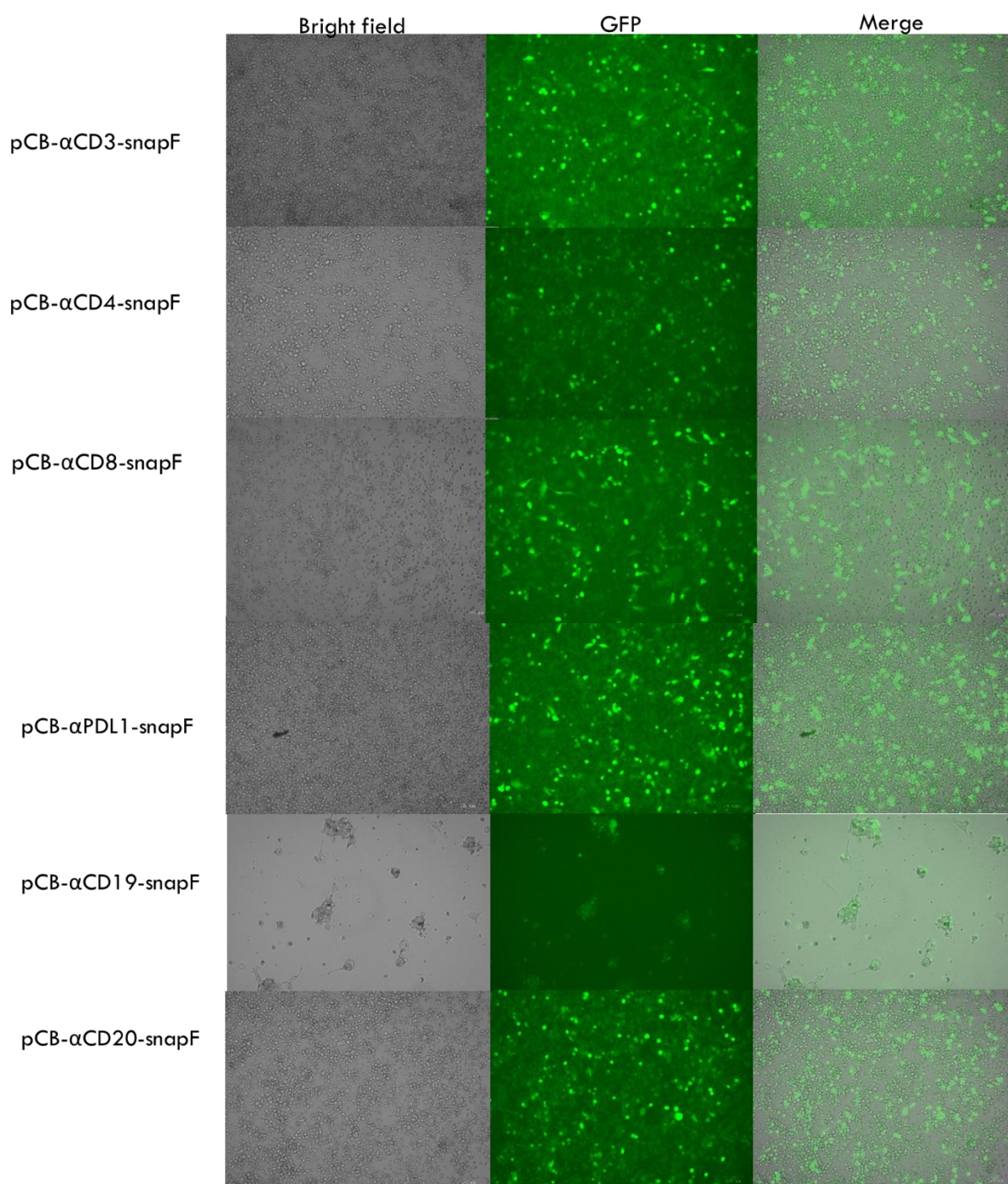


Figure 3.2.5.2: Microscopic visualization of eGFP in transfected HEK293T cells. Transfection was carried out with the XtremeGene™ transfection reagent (Sigma-Aldrich, South Africa) according to manufacturer's instructions. Enrichment of eGFP was performed with 600 μ g/ml of Zeocin. The left panel depicts a brightfield (phase contrast) channel which shows all the cells in a specific region. The centre panel depicts the green channel depicting the eGFP expressing cells and the right panel is an overlay of both channels. All images were taken with a ZOE Fluorescent Cell Imager at 100 μ m magnification.

Due to the initial starting cultures having lower than usual transfection efficiencies it was very difficult to achieve 70% eGFP-positive cells. As such supernatant was harvested from cultures once a 50% eGFP positive culture was reached. Even under 600 μ g/ μ l Zeocin pressure many of the cultures remained eGFP negative and were not harvested. pCB- α CD4-snap and pCB- α CD20 were successfully harvested in sufficient amounts to undergo protein purification. The

pCB- α CD19 culture started of looking very stressed but eventually the cell morphology normalized and sufficient amount of cell culture supernatant was harvested from the culture. The pCB- α CD8-snap culture initially looked promising, but the cultures never reached a confluency of 90% and due to the low cell number the expressed protein of interest was very dilute and not in sufficient amounts to purify. The pCB- α PDL1-snap expressing cells died shortly after transfection, indicating the expressed protein was toxic to the expressing cells.

3.2.6: SDS and BG Alexa Flour 488 Analysis

SDS PAGE gels were used to analyse crude fractions of purified cell culture supernatant. All expressed proteins were expected to be between 48-58 KDa in size. A protein ladder was run with each gel for comparison purposes to determine the size of proteins within each fraction.

Pre-concentrated fractions were first run on the gel to determine which fractions contained proteins and thus should be pooled and concentrated.

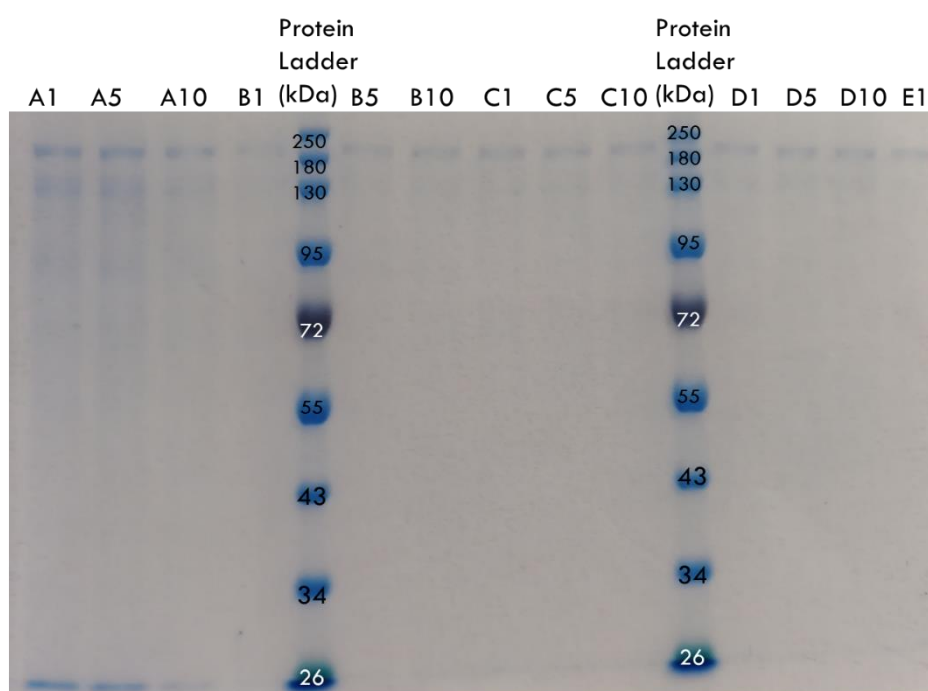


Figure 3.2.6.1: SDS-PAGE analysis of pre-concentrated IMAC purified pCB- α CD4-SNAP protein fractions. Following IMAC purification, a 10% SDS-PAGE gel was run to confirm the expression of the pCB- α CD4-SNAP protein based on its molecular weight and rate of migration through the gel. The gel was loaded with 15 μ l of protein in each well and ran at 100 volts for 95 minutes. The bands were subsequently visualized by Aqua stain solution. Page Ruler loaded in lanes 5 and 11 were used to determine the proteins molecular weight.

As evident in figure 3.2.6.1 no protein bands were visible in any pCB- α CD4-Snap purified fractions indicating that the protein was not present or insufficiently concentrated to visualize. Due to the lack of visual bands, fractions were pooled based on the elution gradient derived during IMAC. The concentrated protein was then re-run on an agarose gel. Figure 3.2.6.2 (A) below shows the presence of a protein band with the correct size indicating that

pCB- α CD4-snap was expressed successfully by the HEK293t cultures but in low concentrations.

In addition to pCB- α CD4-snap; pCB- α CD19-snap, pCB- α B7.1-snap and pCB- α CD20-snap cultures all reached 70% eGFR expression and were harvested for purification. Based on the lack of protein bands observed in the pre-purification SDS, all cultures were directly concentrated after purification using their IMAC elution profiles as a guide on which fractions to pool.

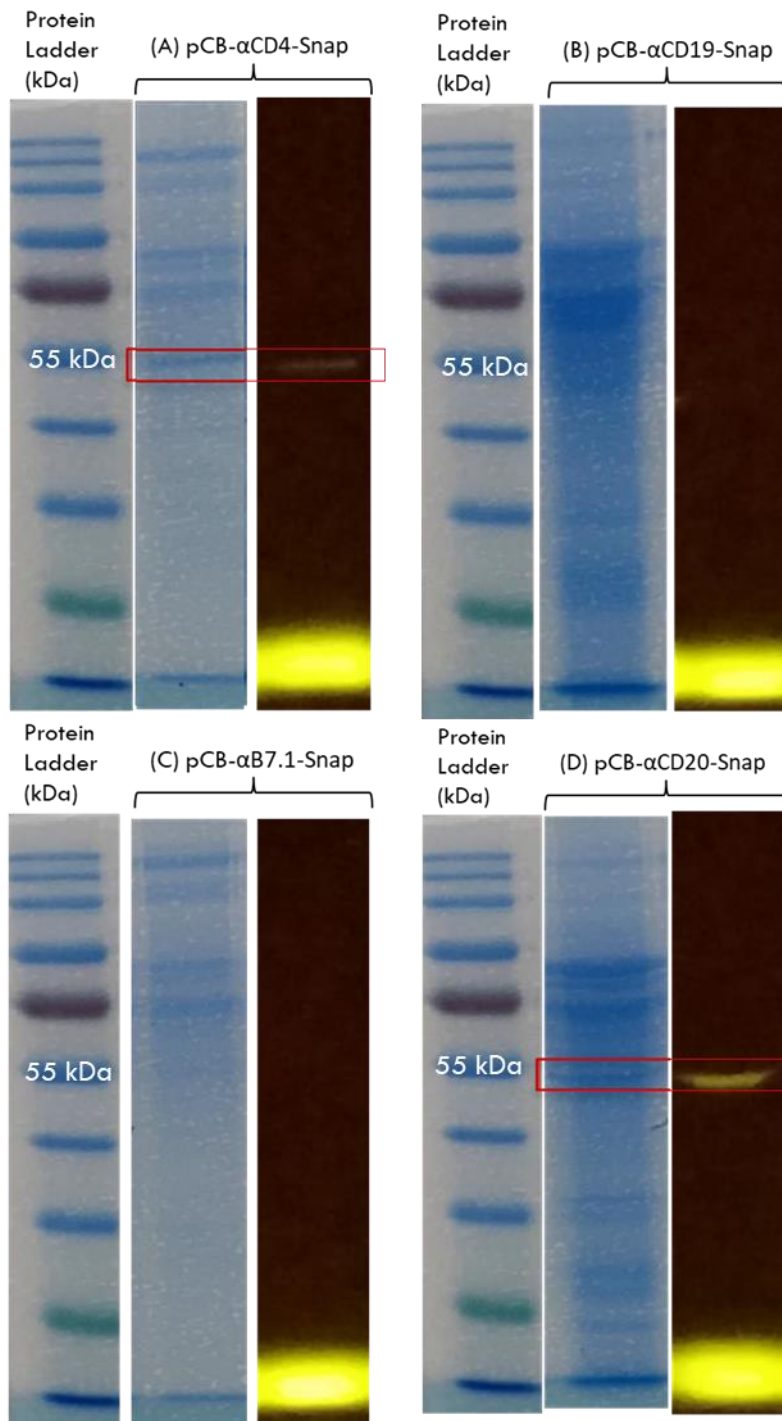


Figure 3.2.6.2: Assessing the presences and binding activity of pCB- α CD19-SNAP and pCB- α CD20-Snap by SDS and conjugation to BG-Alexa Fluor 488. A ratio of 1:2 of protein to BG-Alexa Fluor 488 was used in the conjugation reaction (5 μ M to 10 μ M) to α CD20(scFv)-SNAP (51 kDa) and α CD19(scFv)-SNAP (50 kDa). Left panel: Alexa488-conjugated protein ran on a 10% SDS-PAGE gel stained with Aqua staining solution. Right Results panel: The same SDS-PAGE gel visualized under blue light for potential fluorescence. The SDS-PAGE gel was run at 100 volts for 95 minutes and 20 μ l of conjugated protein samples was loaded per well. Page Ruler prestained protein ladder (Thermo Fisher Scientific, South Africa) was used to confirm the protein sizes. A Dark Reader Transilluminator was used for visualization of the fluorescent signal.

As observed in figure 3.2.6.2, specific bands of the correct size were observed in the pCB- α CD4-snap and pCB- α CD20-snap protein lanes indicating the presence of full length, functional SNAP-tag on the N-terminus of both proteins. This demonstrated that the Hek293T expression system is capable of expressing SNAP-tag based proteins that are functional and capable of conjugation to BG-modified substrates.

No specific bands were observed in the pCB- α CD19-snap and pCB- α B7.1-snap lanes indicating that there is a lack of snap-proteins or that the snap-tag domain of those proteins were not functional. A possible explanation for this is that all cultures (which had >70% eGFP expression at the time) needed to be frozen and stored at -20°C upon the initiation of the nationwide lockdown in 2019 due to the COVID19 pandemic. After three-to-four months at -20°C the cultures were thawed and allowed to propagate. The pCB- α CD19-snap and pCB- α B7.1-snap cultures never got back to 70% eGFP positivity but due to time constraints the cultures were harvested anyway. However, as evident in figure 3.2.6.2 there was insufficient amounts of the putative proteins in both these cultures for harvesting and purification. The pCB- α CD4-snap and pCB- α CD20-snap cultures were the only cultures to recover to 70% eGFP expression. The pCB- α CD3-snap, pCB- α CD8-snap and pCB- α PDL1-snap cultures never recovered from the freeze thawing process having very low numbers of viable cells which died before reaching a harvestable confluency. These cultures would require starting fresh with a new round of transfection and zeocin application. As the cultures initially took more than 6months to reach a 70% eGFP expression, we chose not to re-start these cultures but rather focus on the ones which did recover.

The next steps in assessing the integrity of the pCB-CD4-snap and pCB-CD20-snap proteins would be to test the binding capacity of scFv. Unfortunately, in addition to time delays and having restricted access to the lab after the lockdown, we were also unable to get timely deliveries of the cell lines required to complete our functional assays with these proteins. The proteins have been stored and should be tested as soon as these cell lines are available. In addition to binding of cell lines, the proteins may be tested for binding directly on cancer tissue sections. This method has already been set up by another PhD student in the Barth lab, Natasha Hardcastle, who has shown the specific binding of various snap-tag based proteins on tumour sections. The generation of these proteins is an important first step in setting up a differential diagnostic panel for difficult to diagnose cancers. In conjunction with other proteins generated and expressed by other students at MB&I these targets provide a comprehensive overview of the type receptors expressed by a patient's tumor and thus can be used to determine the most appropriate immunotherapeutic application for a given patient.

Section 3.2.7: Summary of Section 3.2 Results

The overarching aim of this section of the thesis was to establish a methodology to generate SNAP-tag based fusion proteins. Utilizing information available in patents and databases, we were able to extract several antibody sequences necessary for the development of a companion diagnostic panel that could be used to assist in characterizing TNBC tumours based on their TIL distribution. Using information already available on databases allowed us to bypass the need for identifying antibody sequences in a laboratory thus saving large amounts of money and time. By combining this information with sophisticated technologies such as SNAP-tag and mammalian expression cell culture we were successfully able to generate functional SNAP-tag based recombinant fusion proteins targeting CD4 and CD20.

The successful expression of pCB-CD4-snap and pCB-CD20-snap supports this methodology in the generation of possible immunodiagnostics. There are however some major optimizations still necessary with regards to the protein expression system, but the given data does demonstrate an early and important proof-of-concept that this method can be successful in expressing functional snap-tag based proteins.

Section 3.3: Constructing a novel TNBC targeting snap-fusion protein, pCB- α BCK1-snap which has potential diagnostic and therapeutic applications.

3.3.1 Validation and assembly of BCK-1 scFv sequence

This part of the thesis fits into a larger MB&I based scientific goal of developing a panel of tools which enable targeted delivery of cytotoxic drugs. To this end a novel scFv sequence, BCK-1, which demonstrated specific binding to membrane fractions of the 8701BC breast cancer cell line was subjected to various sequence analysis and editing. The original sequence obtained from Rothe et al was provided in an amino acid format. The first step in processing was to obtain the nucleotide sequence of BCK-1 by reverse translation using the geneinfinity online tool (http://www.geneinfinity.org/sms/sms_backtranslation.html). The sequence was then codon optimized for mammalian expression. This was followed by an in-silico quality control analysis using IgBLAST, to confirm the presence of intact full length CDRs and FR regions in the scFv sequence as illustrated in figure 3.3.1.1 below.

GGCCCAGCCGGCCATGGCCCAGGTGCAGCTGCAGGAGTCCGGGCCAGGACTGGTGAAGACTTCGGAGACCC
TGTCCCTCACCTGCAGTGTCTCTGGTGGCTCCATCAGTAGAAACTACCTGAGCTGGCTCCGGCAGCCCGCCGGG
AAGGGGCTGGAGTGGATTGGGCGAATATATACTACTGGGAGTACCGATTACACCCCTCCTTCTGGAGTCGAGT
CACCATGTCGACAGACACGTCCAACACCAAGTTTTCCCTTAGGCTGACGTCTCTGACACCCGCGGACACGGCCA
CATATTACTGTGCGACA GAGGACCGCCTTGGTGGTTATTGGAGAGTCTTTGACACC TGGGGCCAGGGCACACC
GGTCACCGTCTCC TCAGGTGGCGGTGGCTCGGGCGGTGGTGGGTGGGTGGCGGCGATCAGACCTCCAGCT
GACCCAAAGCCCACCCTCCCTGTCTGCATCTGATGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGAGC
ATTAGCAGCTATTTAAAT TGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTACGCTGCATCCAGT
TTGCAAAGTGGGGTCCCATCAAGGTTCAAGTGGCAGTGGATCTGGGACAGATTCTACTATCACCATCAGCAGTCT
GCAACCTGAAGACTTTGCAACTTACTACTGT CAACAGAGTTACAGTACCCCTCGGACGTTTCGGCCAAGGGACA
CAACTGGAGATCAAACGAGCGGCCGE

Figure 3.3.1.2: Assembled BCK-1 (scFv) sequence for cloning. The different components of the sequence are highlighted in different colours; blue is the Sfi1 recognition site, green is the heavy chain CDRs, red is the Gly(4)Ser linker, purple are the light chain CDRs and orange is the Not1 recognition site.

3.3.2 In-silico cloning of pCB-BCK1-snap

An ORF for the pCB-BCK1-snap construct was generated through in-silico insertion of the assembled scFv sequence into the pCB-SNAP backbone as shown in figure 3.3.2.1. Included in the ORF is an enterokinase recognition sequence which plays an essential role in generating the final active product by removing of the N-terminal pre-peptide. The Ig-Kappa leader sequence and poly HIS-tag are located at the N-terminus of the SNAP-tag while the scFv sequence is located at the 3'.

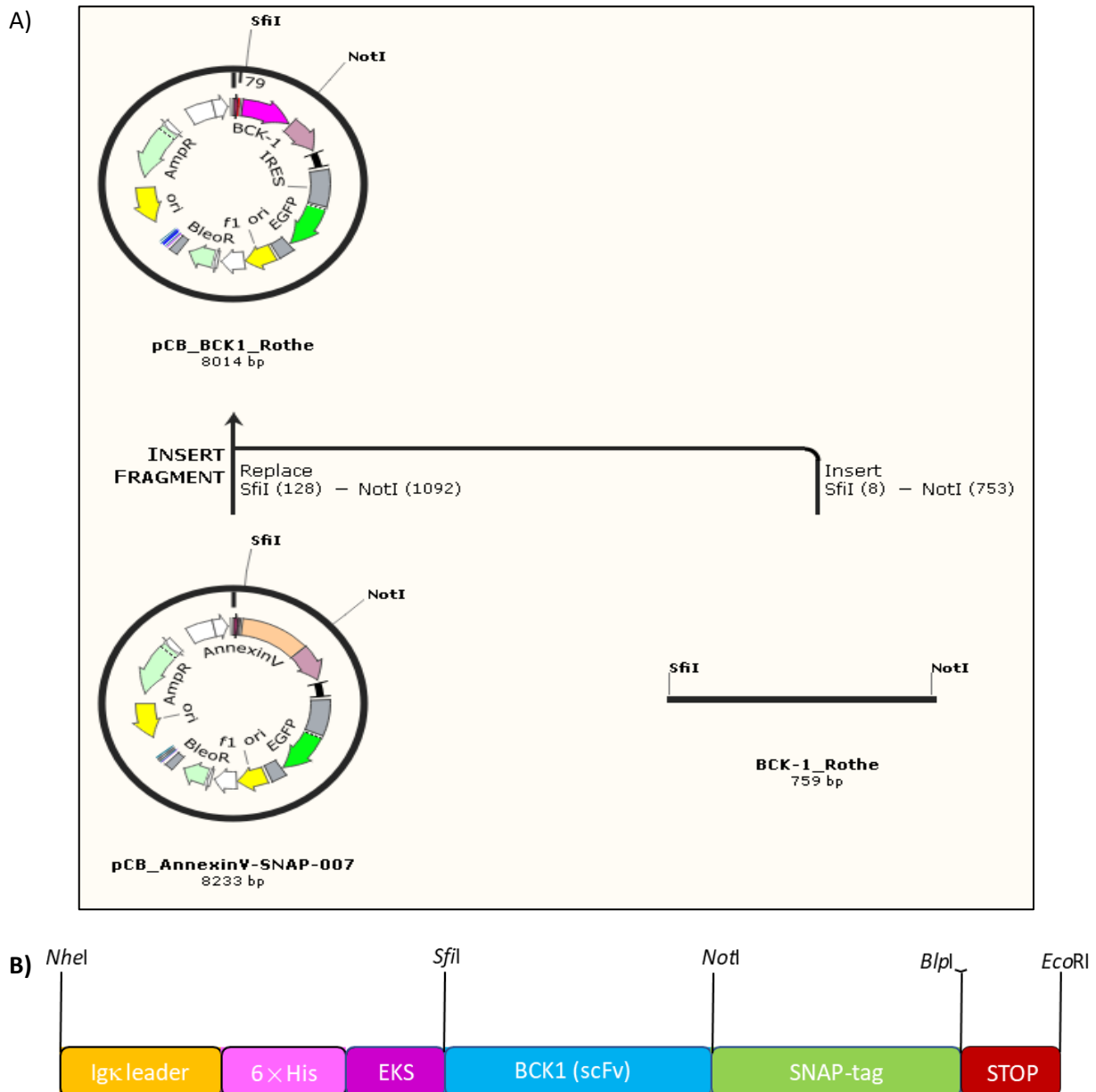


Figure 3.3.2.1: pCB-BCK1-Snap construct. (A) In-silico restriction cloning of BCK-1(scFv) into pCB-AnnV-SNAP using Sfi1 and Not1. (b) ORF of pCB-BCK1-SNAP construct including Sfi1 and Not1 recognition sites; IgK leader: Ig-Kappa leader sequence important for secretion of the fusion protein during host expression, 6xHis : 6 histidine tags used for protein purification during IMAC and detection in western blot analysis, EKS: enterokinase cleavage site important for enzymatic removal of the N-terminal elements and STOP: stop codon for termination of protein synthesis.

In-silico cloning in snapgene software confirmed that the assembled scFv genes would be compatible for ligation into the pCB-snap backbone. After the final assembly the construct was checked for the presence of inappropriate stop codons and translated into its amino-acid sequence which was aligned to its original parental (rothe article) sequence. This was to validate the integrity of the original CDRs and FRs prior to having the sequence synthesised at Genescript (USA).

3.3.3 Molecular cloning of pCB-BCK1-snap

Molecular cloning involves diverse experiments that are performed collectively to construct and amplify a DNA fragment of interest. The cloning of pCB-BCK1-SNAP was done with restriction enzyme digest and T4 DNA ligation.

3.3.3.1 Bulk preparation of DNA plasmids by transformation

Plasmids received from genescript containing the BCK1 scFv gene was bulk prepared together with pCB-AnnV-SNAP by transformation in DH5 α cells. The cultures were grown overnight at 37 °C and the plasmid DNA from each culture was then individually isolated with the phenol/chloroform DNA isolation method. Extracted DNA was then quantified by spectrophotometry as depicted in table 3.3.3.1.1 below.

Table 3.3.3.1.1: Purity of DNA extracted from bulk preparations.

Plasmid	A260/230	A260/280	Concentration (ng/ul)
Puc57_αBCK1	2.021	1.760	27629
pCB_AnnV_Snap	2.159	1.872	48957

The A260/230 ratio is commonly used to determine if DNA samples contain contaminating organic compounds such as phenol. The generally accepted values for pure DNA samples are between 2.0 and 2.2. Both DNA sample were within the acceptable range indicating that the extracted DNA was not contaminated with organic compounds. The A260/280 ratio is more routinely used to determine the purity of a DNA sample. The ideal ratio for DNA is 1.8 but anything between 1.7-2 is acceptable. Both extracted constructs seemed to be pure as indicated by A260/280 ratios thus were used for cloning.

3.3.3.2 Restriction enzyme digest and agarose gel electrophoresis

To construct the recombinant pBC-BCK1-snap construct, the puc57 plasmid housing the BCK1 scFv gene, as well as the pCB_AnnV_snap vector backbone were digested. This was done using *Sfi1* and *Not1* double digest which allowed for the extraction of the scFv gene from the puc57 plasmid. The pCB backbone was also double digested with *Sfi1* and *Not1* to remove the AnnV fragment and introduced sticky ends into the backbone in preparation for ligation. Both Restriction digested samples were then run on a 1.2% agarose gel to allow for visualization of the different DNA fragments as depicted in figure 3.3.3.2.1 below.

Table: Description of agarose gel lanes depicted in figure 3.3.3.2.1.

Lane	Construct	Enzyme/control	Expected fragment size/s (bp)
1	1KB molecular weight marker	NA	Multiple
2	pCB_αAnnV_Snap	Undigested	8233
3	pCB_αAnnV_Snap	No buffer control	8233
4	pCB_αAnnV_Snap	<i>Sfi</i> 1	8233
5	pCB_αAnnV_Snap	<i>Not</i> 1	8233
6	pCB_αAnnV_Snap	<i>Sfi</i> 1 + <i>Not</i> 1	7255 + 978
7	pCB_αAnnV_Snap	<i>Sfi</i> 1 + <i>Not</i> 1	7255 + 978
8	pCB_αAnnV_Snap	<i>Sfi</i> 1 + <i>Not</i> 1	7255 + 978
9	Puc57_αBCK1	Undigested	3455
10	Puc57_αBCK1	No buffer control	3455
11	Puc57_αBCK1	<i>Sfi</i> 1	3455
12	Puc57_αBCK1	<i>Not</i> 1	3455
13	Puc57_αBCK1	<i>Sfi</i> 1 + <i>Not</i> 1	2710 + 745
14	Puc57_αBCK1	<i>Sfi</i> 1 + <i>Not</i> 1	2710 + 745
15	1KB molecular weight marker	NA	Multiple

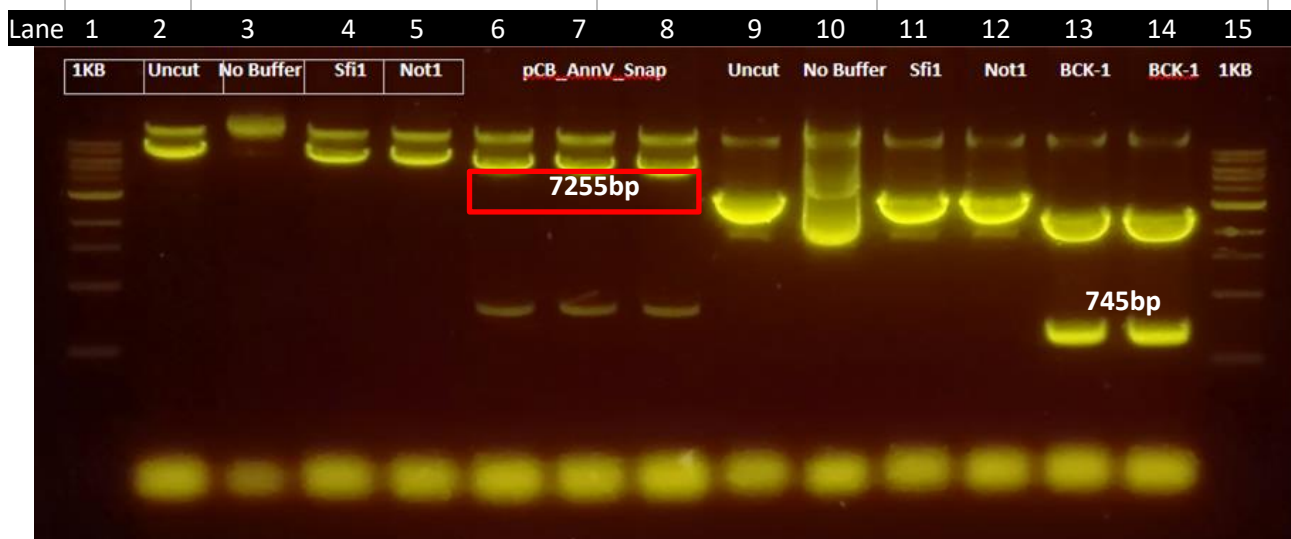


Figure 3.3.3.2.1: Agarose gel depicting DNA fragments after *sfi*1 and *Not*1 double digest: 7269bp pCB backbone and 745bp αBCK1 insert was extracted from gel for ligation.

The *Sfi*1 and *Not*1 double digest of pCB-AnnV-SNAP and puc57-BCK1 were both successful resulting in the generation of a 7255bp pCB-SNAP backbone plasmid and a 745bp BCK1 scFv

insert. Single enzyme controls were also included to make sure that both *Sfi1* and *Not1* were enzymatically active thus resulting in linearization of the plasmids. The appropriate bands (7255bp band in lanes 6-8 and 745bp band in lanes 13-14) were excised from the agarose gel. Bands containing the same DNA were pooled and the DNA was purified using the Qiagen quick gel extraction kit. The DNA was then quantified and assessed by spectrophotometry in preparation for ligation.

3.3.3.3 Ligation of BCK1(scFv) into the pCB-SNAP backbone

T4 DNA ligation was used to join the BCK1 scFv insert and the pCB-SNAP vector. This was done using T4 DNA ligase, an enzyme that forms phosphodiester bonds on compatible DNA termini by adding 3' hydroxyl groups and 5' phosphate groups to the cohesive ends of the DNA formed by restriction enzyme digest. The ligation reaction was set up in 3 different vector to insert molar ratios; 1:1, 1:3 and 1:5. The ligated products were transformed into chemically competent DH5 α *E.coli* cells and plated on LB agar plates containing 200 μ g/ml ampicillin. The plates were incubated overnight to allow individually colonies to grow as shown in figure 3.1.3.3.1.

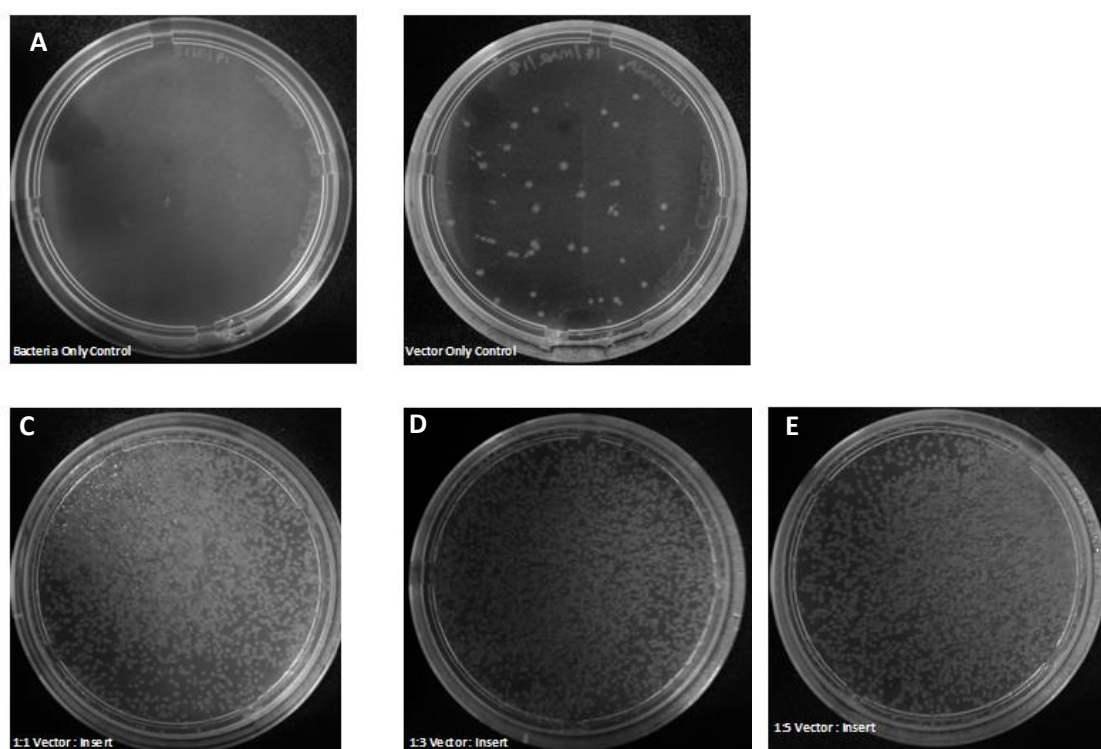


Figure 3.3.3.3.1: Growth of *E. coli* cells transformed with potential recombinant pCB-BCK1-snap plasmid DNA. Following transformation, 100 μ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 μ g/ml. (A) Contamination control (B) Vector-only plate (C) 1:1 ratio of vector to insert (D) 1:3 ratio of vector to insert (E) 1:5 ratio of vector to insert.

No bacterial growth was observed on the bacteria only contamination control plate. A contamination plate is important in detecting the presence of bacteria or fungus from the environment that may grow on the plates and compromise the reliability of the experiment. Ampicillin on the plates is used as a selective marker preventing the growth of untransformed

bacteria. Bacterial clones containing the pCB-BCK1-SNAP construct were able to grow as the plasmid contains a gene permitting the expression of β -lactamase, an enzyme which catalyzes the hydrolysis of the β -lactam ring of ampicillin. Hence, the colonies observed on the other LB plates are true transformants. The 'vector only' (digested vector backbone) plate showed presence of a few colonies, indicating that a small percentage of vector religation occurred (Figure 2B). Successful transformants were observed in all vector to insert ratios. The transformation efficiency was very high but could not be accurately calculated as many of the colonies were not individually identifiable. Multiple colonies were picked at random from the different plates, mostly from the periphery of the plate where individual colonies could be seen. The picked colonies were cultured in 5ml LB agar broth in preparation for restriction mapping.

3.3.3.4 Restriction mapping of potential recombinant pCB-BCK1-SNAP clones

Restriction mapping was used to verify successful ligation of BCK1 (scFV) into the pCB-SNAP backbone. A virtual simulation was initially carried out using SnapGene software, to determine which enzymes would result in differential fragmentation of the recombinant and parental plasmids. The recombinant pCB-BCK1-SNAP clones and parental pCB-AnnV-SNAP clones were both singly digested with PVU II and PSTI. The gel image obtained (Figure 3.3.3.4.1) showed distinct fragment patterns for the recombinant and parental plasmids that were in accordance with the theoretical gel simulated from SnapGene® (Figure 3.3.3.4.2).

Table: Description of snapgene agarose gel simulation

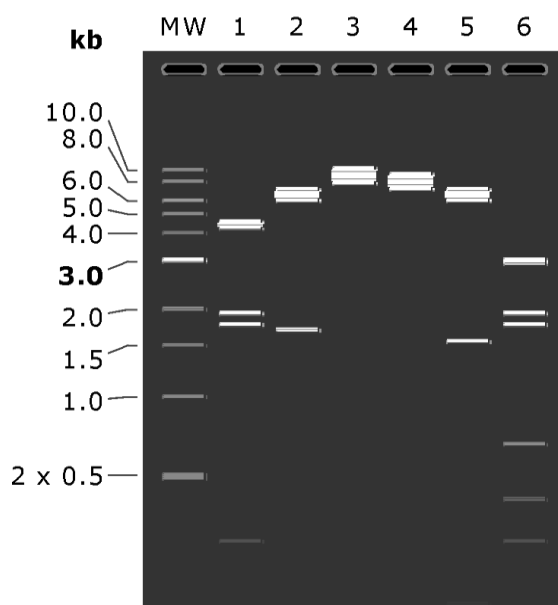


Figure 3.3.3.4.1: Theoretical gel simulation using Snapgene and corresponding band sizes for each enzyme.

Lane	Construct	Enzyme	Expected fragments (bp)
MW	DNA Ladder	NA	Multiple
1	pCB_AnnV_sna p	PVUII	4345, 1895, 1748, 245
2	pCB_AnnV_sna p	PSTI	6561, 1672
3	pCB_AnnV_sna p	Undigested	8233
4	pCB_αBCK1_sn ap	Undigested	8014
5	pCB_αBCK1_sn ap	PSTI	6419, 1542, 53
6	pCB_αBCK1_sn ap	PVUII	3034, 1895, 1748, 684, 408, 245

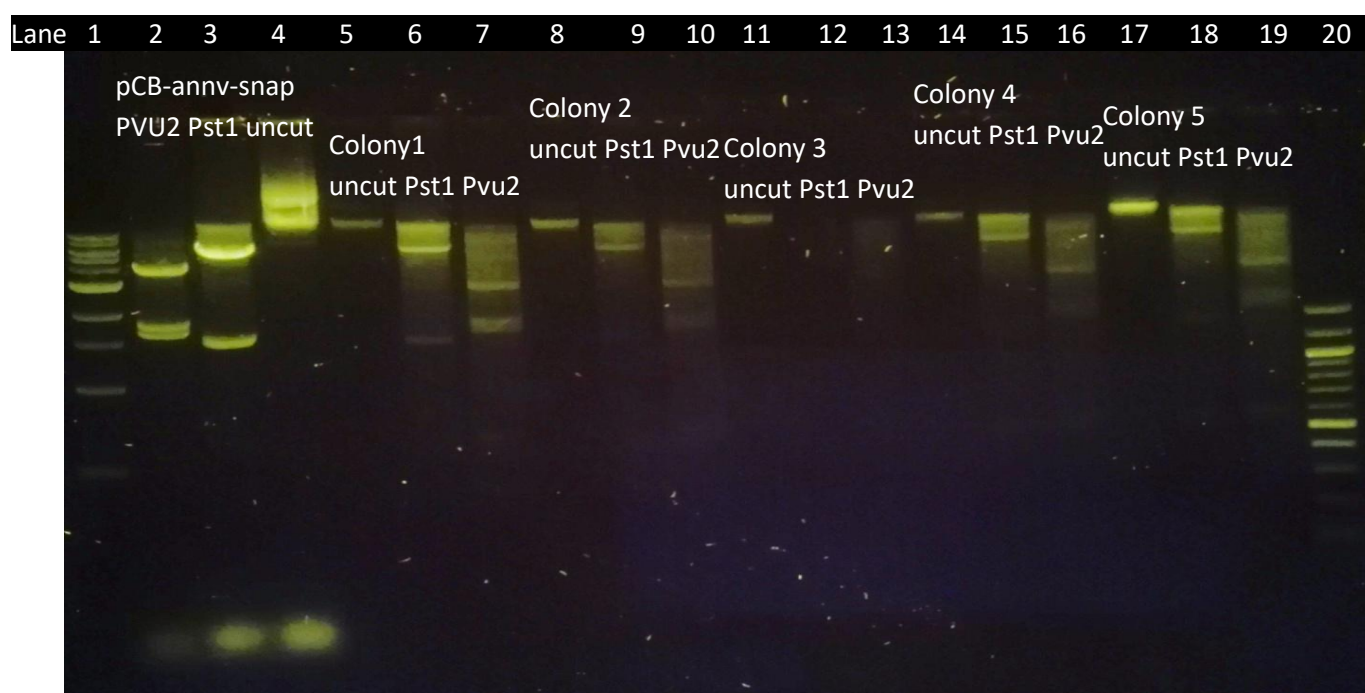


Figure 3.3.3.4.2 . Analysis of restriction mapping of potential pCB-BCK-1-Snap clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs. Digests of pCB-AnnV-SNAP was completed for comparative purposes.

Table: Description of agarose gel in figure 3.3.3.4.2

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	pCB_AnnV_snap	PVUII	4345, 1895, 1748, 245
3	pCB_AnnV_snap	PSTI	6561, 1672
4	pCB_AnnV_snap	Undigested	8233
5	Colony 1	Undigested	8014
6	Colony 1	PSTI	6419, 1542, 53
7	Colony 1	PVUII	3034, 1895, 1748, 684, 408, 245
8	Colony 2	Undigested	8014
9	Colony 2	PSTI	6419, 1542, 53
10	Colony 2	PVUII	3034, 1895, 1748, 684, 408, 245
11	Colony 3	Undigested	8014
12	Colony 3	PSTI	6419, 1542, 53
13	Colony 3	PVUII	3034, 1895, 1748, 684, 408, 245
14	Colony 4	Undigested	8014
15	Colony 4	PSTI	6419, 1542, 53
16	Colony 4	PVUII	3034, 1895, 1748, 684, 408, 245
17	Colony 5	Undigested	8014
18	Colony 5	PSTI	6419, 1542, 53
19	Colony 5	PVUII	3034, 1895, 1748, 684, 408, 245
20	100 KB DNA Ladder	NA	Mulitple

Colonies 1,4 and 5 was chosen, however due to the similarities between the pCB_AnnV_snap and pCB-BCK1-snap constructs we were unable to say with 100% certainty that the colonies have been cloned successfully and contain the correct insert. The chosen colonies were thus bulk prepped and sent to Inqaba Biotechnical Industries (South Africa) for sanger sequencing to confirm cloning.

3.3.4 Sequencing results of selected pCB-BCK1-SNAP clones

Sequence results received from Inqaba were aligned to the original in-silico designed sequence in SnapGene to confirm the integrity of each clone. Figure 3.3.3.5.1 below shows that the aligned clones share 100% homology with the theoretical in-silico designed pCB-BCK1-SNAP sequence confirming the successful incorporation of BCK1(scFv) into the pCB-SNAP backbone. The plasmid was also checked for mutations, insertions, deletions, and frameshifts in preparation for transfection.

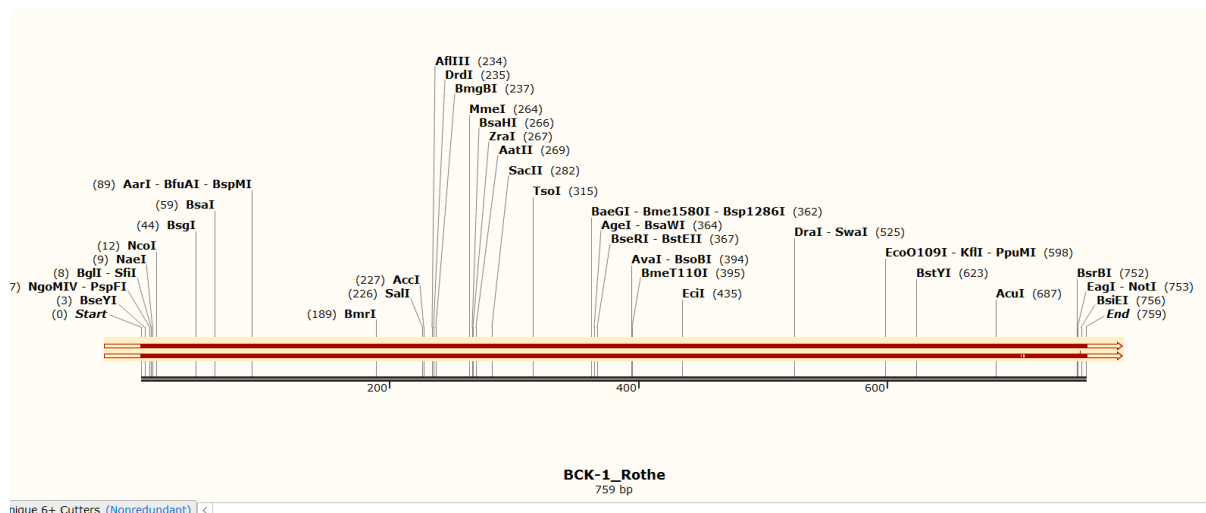


Figure 3.3.3.5.1: Map generated following the alignment of trace sequences with the corresponding in-silico sequence for pCB-BCK1-SNAP. Both clones showed 100% homology with the in-silico ORF sequence as indicated by the dark red sections of the arrows. Effectively transfected cells, potentially expressing the pCB-BCK1-Snap protein, were identified by microscopic visualization of eGFP.

3.3.3 Transfection and expression of pCB-BCK1-SNAP in HEK293T mammalian expression cells.

Sequence verified pCB-BCK1-SNAP clones were introduced into HEK293T cells for protein expression. Positively transfected cells, potentially expressing pCB-BCK1-SNAP proteins, were identified by fluorescent cell imaging of the reporter protein GFP. The transfection efficiency of the culture was determined using flow cytometry and found to be very low at 30%. Transfection was repeated two more times and found to always result between 10% and 40% transfection efficiencies. As the antigen of BCK1 is not currently known, it is possible that the protein interacts with the host cells in a manner which is harmful to the HEK cells making it difficult for transfected cells to express the protein. As such a high dose of Zeocin (300 µg/ml) was applied to the cultures to enrich for the growth of positively transfected cells housing a bleomycin resistance gene. Cell cultures were maintained under Zeocin pressure for several

weeks until a 70% eGFP efficiency was obtained. Due to the low transfection efficiencies in the initial starting cultures higher efficiencies were not obtained with pCB-BCK1-SNAP cultures. These cultures were maintained at a 70% eGFP efficiency and cell culture supernatant was harvested twice a week until 1 L of CCSN was obtained. Figure 3.1.3.1 is a snapshot of the cell culture at two-months post-transfection. As evident in the figure the cultures had very few cells which is probably attributed to the low transfection efficiency resulting in the death of majority of the cell population when Zeocin selection was applied. As a result of the low number of cells, cell cultures had to be maintained for extended periods of time prior to harvesting the CCSN. The high concentration of Zeocin also impacted the morphology of the cells, as evident by the unhealthy rounded shape seen in figure 3.1.3.1 below. The eGFP fluorescent intensity was also very dim suggesting that very small quantities of our putative protein might have been expressed in the CCSN.

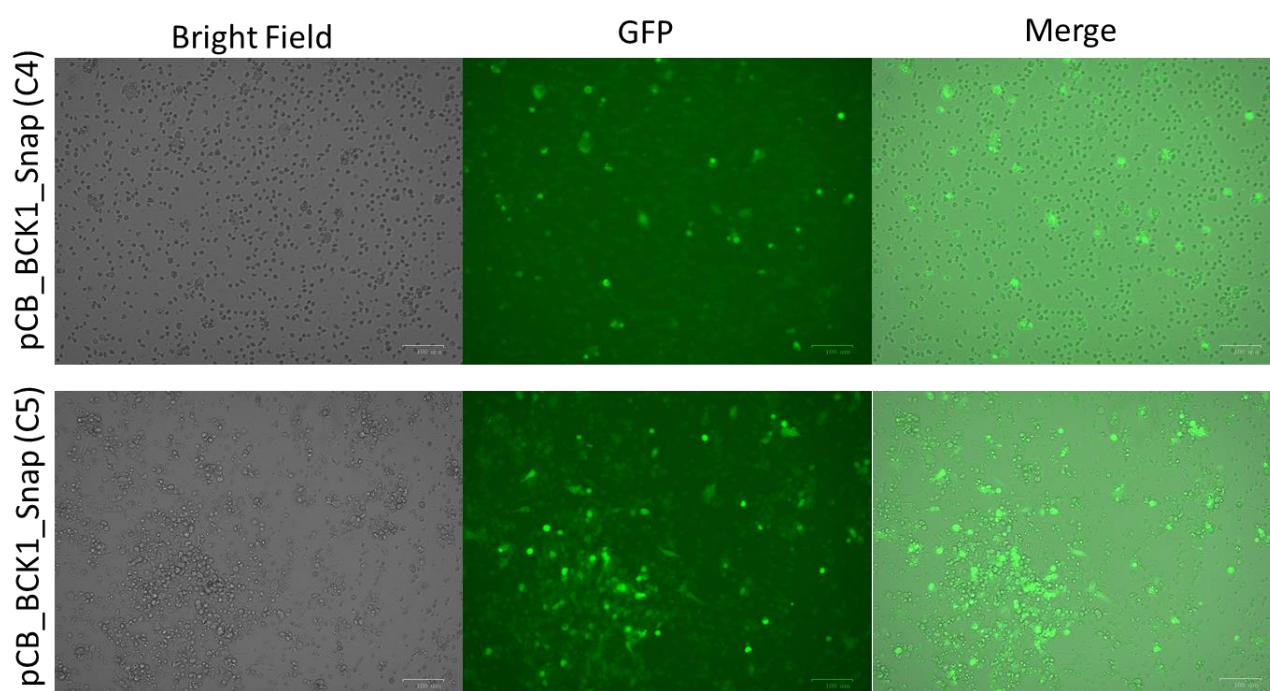


Figure 3.3.3.1: Green fluorescent protein expression in pCB_BCK1_Snap transfected HEK293T cells one week post transfection.

3.3.4 SDS-PAGE analysis of purified pCB-BCK1-SNAP protein fractions

One litre of cell culture supernatant was collected and then purified using IMAC, through Ni^{2+} binding to the His-tag of the protein. This is followed by elution of the protein using an Imidazole gradient. The elution profiles are depicted in figure 3.1.4.1 below. The eluted protein is collected in different fractions.

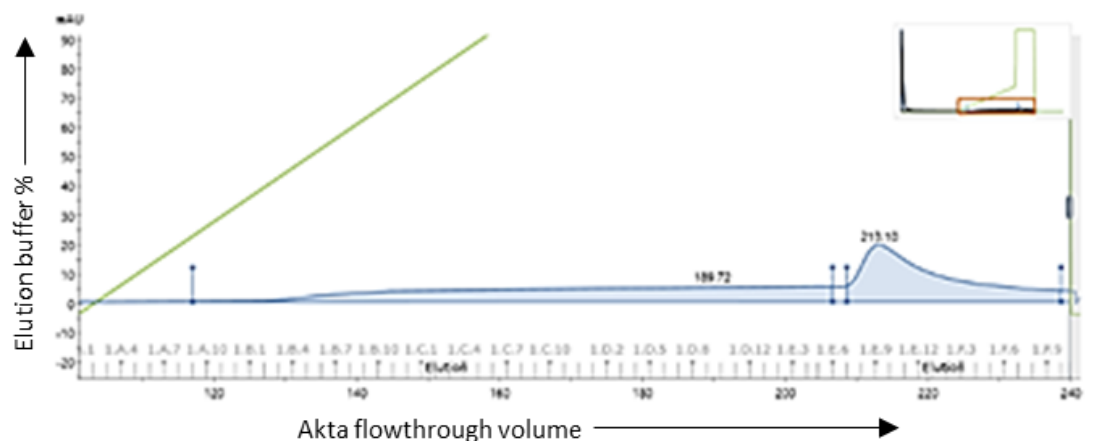


Figure 3.3.4.1: Chromatogram of BCK1-SNAP's purification profile during IMAC. The x-axis represents the AKTA flow-through volume over time while the Y-axis represents the percentage of the elution buffer. The green line indicates the imidazole concentration and the blue line shows the elution profile of the protein.

Each protein fraction is run on a 10% SDS-PAGE gel to determine which fractions contain the putative fusion protein.

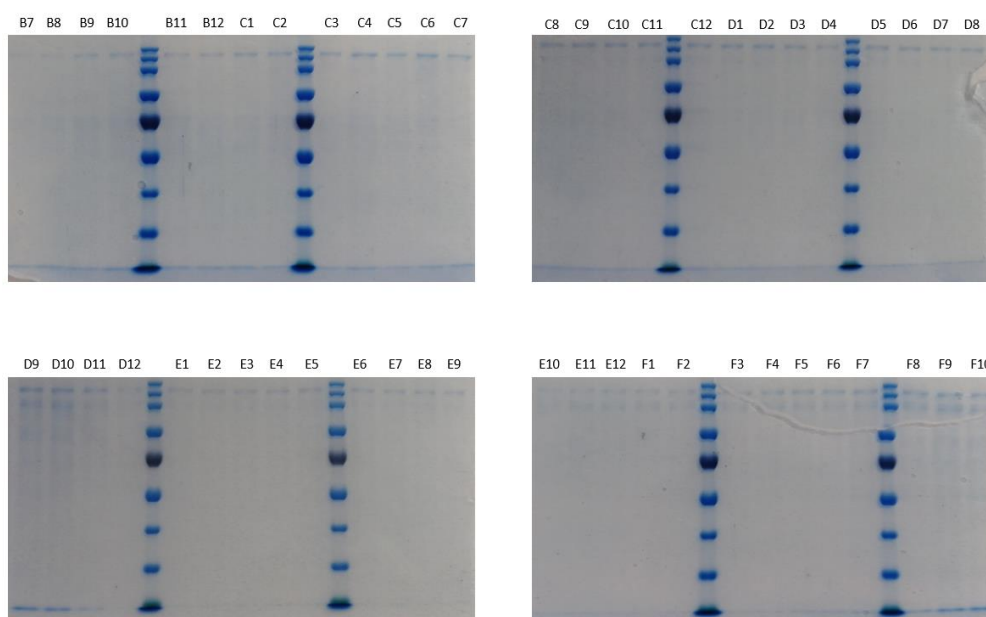


Figure 3.3.4.2: SDS-PAGE analysis of pre-concentrated IMAC eluate BCK1-SNAP protein fractions. 10% SDS-PAGE gel loaded with 15 μ l of sample and run at 100 volts for 95 minutes were used to confirm expression of BCK1-SNAP proteins. The protein bands were visualized with Aqua staining solution. Page Ruler pre-stain protein ladder was used to determine the protein molecular weights.

No protein fractions were observed on the SDS-PAGE gels in the individual fractions indicating that there was insufficient concentrations of the recombinant protein for visualization. As

such pooling was done based on the protein elution peaks derived from the AKTA purification. The concentrated samples were then run on a 10% SDS-PAGE gel.

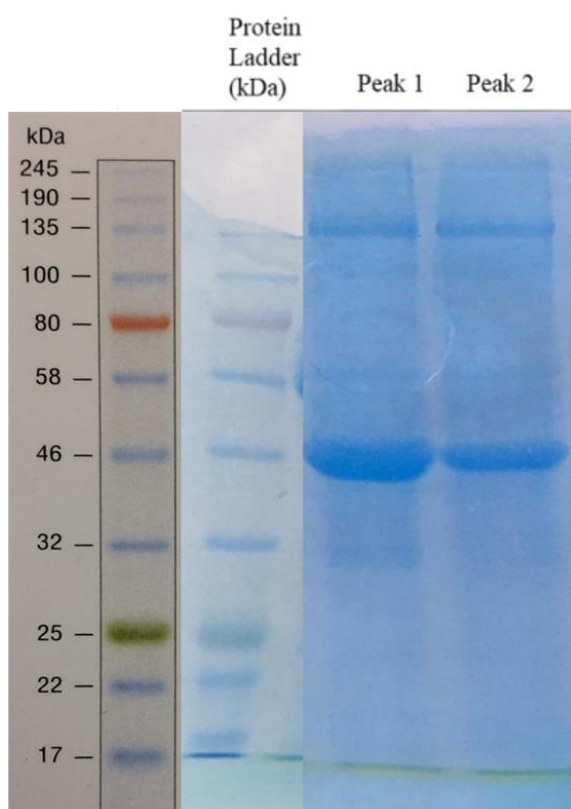


Figure 3.3.4.3: SDS-PAGE analysis of concentrated IMAC eluate BCK1-SNAP protein fractions. Amicon concentrated protein fractions were run on an SDS-PAGE gel at 100 volts for 95 minutes. 20 µg of total sample from each peak was loaded on lands 2 and 3 respectively. Page Ruler prestained protein ladder was loaded in lane 1 for identification of the proteins molecular weight. The SDS-PAGE gels were stained with Aqua staining solution for visualization of the protein bands.

The SDS-PAGE gel confirmed the presence of a 46KDa protein band which is the expected size of BCK1-snap suggesting that the Amicon concentration concentrated the protein from a previously invisible concentration to a high enough concentration to visualize.

3.3.5 Western blot analysis of purified BCK-SNAP protein.

Once the size of the detected protein was confirmed to be correct the protein was further analysed by western blot to confirm the proper folding of a full-length recombinant protein housing a functional his-tag. The western blot was carried out by transfer of protein bands from unstained SDS-PAGE gels to a nitrocellulose membrane. The proteins are then incubated with an anti-His rabbit primary antibody which specifically binds to the C-terminal His-tag of the recombinant protein. An anti-rabbit HRP-conjugated secondary antibody was then used for visualization. The HRP on the secondary antibody interacts with chemiluminescent substrate to produce a colour reaction which is visible under UV light. Intact full-length protein was observed on the western blot indicating that expression of BCK-1 was successful and that the expressed protein is functional.

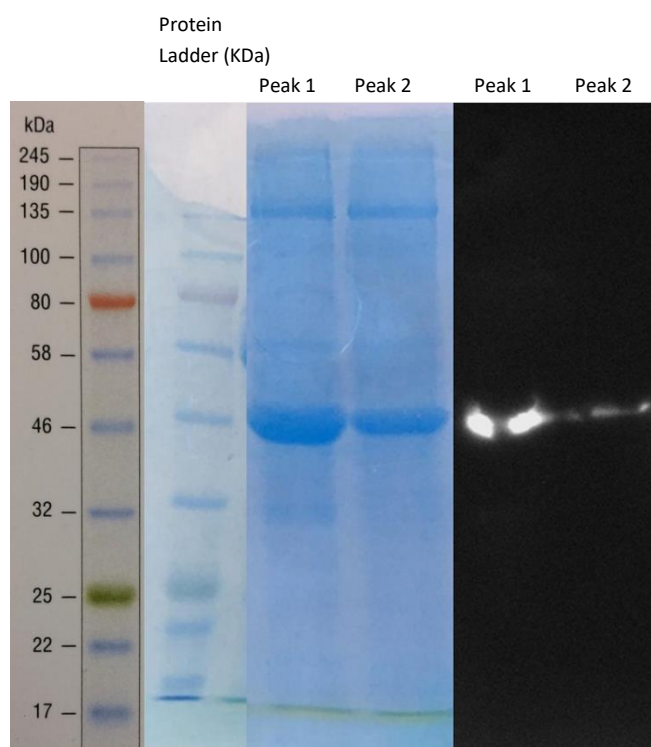


Figure 3.3.5.1 Western blot analysis of enriched BCK1-SNAP protein fractions.

The right panel of this figure depicts the 10% SDS-PAGE gel for comparison with the immunoblot (right panel) of the transferred proteins from the nitrocellulose membrane of a duplicate SDS-PAGE gel. Membranes were visualized on a Gel Doc™ XR Gel Documentation System.

3.3.6 Summary of section 3.3 results: the construction of BCK1-SNAP protein.

The initial aim of this section was to generate a BCK-SNAP fusion protein for the potential diagnostic and therapeutic application in TNBC. The major aim of this section was achieved as shown in the western blot analysis of the purified protein suggesting that full length protein was expressed by the HEK293T mammalian expression system. Further steps to confirm the binding of the construct on breast cancer specific cells would be the necessary next step to determine how valuable this construct would be as a diagnostic and/or therapeutic application. Unfortunately, the 2019 COVID pandemic restricted acquisitions of appropriate cell lines required to test the BCK1-SNAP construct for binding and cytotoxic efficacies. However, the construction of this fusion protein demonstrated that the workflow utilized in this section can successfully generate functional snap-tag based fusion proteins. This forms an important first step in the development of new TNBC targeting therapies.

Section 3.4: Applying the tools and technologies used to develop cancer specific snap-fusion proteins to infectious diseases using Zika virus as an example.

According to the world health organization (WHO) emerging infectious diseases (EID) are defined as any diseases caused by a recently discovered and previously unidentified pathogen, that has major health repercussions either locally or internationally (WHO definition). As EIDs are new to the human species many of them lack currently available treatment options and identifying new drugs for treatment could take years. Thus, EIDs could benefit from the application of immunotherapies, more specifically snap-tag based fusion proteins. As such, we sought to use Zika virus as an example of this and demonstrate both the versatility of snap-tag technology and its potential benefits in the treatment of infectious diseases.

Zika virus is a mosquito-borne flavivirus which is primarily transmitted to humans by the *Aedes mosquito*. The symptoms of this infection are generally mild and include fever, rash, conjunctivitis, headaches and joint/muscle pains. However, infection during pregnancy can have more severe effects on the infant, these include microcephaly and other congenital defects. There is currently no vaccine or treatment available for zika virus infections necessitating the need for research and development of new treatment. Like other flaviviruses, ZIKV is a positive-sense RNA virus with an 11 kb ORF [136]. The genome encodes a single polyprotein that is post-translationally cleaved by the host and viral proteases into three structural proteins, the capsid[C], pre-membrane [prM], and envelope [E], as well as seven non-structural proteins. The envelope protein is further subdivided into three domains which include, : a central β -barrel domain (domain I [DI]), an extended dimerization domain (DII), and an immunoglobulin-like segment (DIII). Haiyan et al, previous demonstrated that antibodies against the third domain of the envelope protein induces neutralizing antibodies [136].

Development of the pCB-Zika-DIII-Snap fusion protein was done in collaboration with Prof. Paolo Marinho from the Institute of Biomedical Sciences at the University of Sao Paulo. As part of the collaboration agreement, Prof Marinho's team provided us with the full-length α -DIII-LR ZV-67 monoclonal antibody sequence. The epitope recognized by the ZV-67 mAb is created by four discrete secondary structure elements: the A-strand, B-C loop, D-E loop, and F-G loop. A total of 21 residues are contacted by ZV-67. This epitope region is termed lateral ridge (LR). The LR epitope for ZV-67 is accessible on the mature virion. ZV-67 is ZIKV specific and neutralizes infection of African, Asian, and American strains, possibly inhibiting infection at post-attachment stage blocking viral fusion from the endosome [136]. The intended SNAP-tag construct derived from MAb ZV-67 was then chosen as the sequence to be used for a proof-of-concept experiment demonstrating the applicability of snap-tag technology in infectious diseases.

3.4.1 Validation and assembly of pCB- α ZIKA-DIII-Snap scFv sequence

The original sequence obtained from ZV-67 mAb was provided in an amino acid format. The first step in processing was to obtain the nucleotide sequence of ZV-67 by reverse translation using the geneinfinity online tool (http://www.geneinfinity.org/sms/sms_backtranslation.html) . The sequence was then codon optimized for mammalian expression. This was followed by an in-silico quality control analysis using IgBLAST, to confirm the presence of intact full length CDRs and FR regions in the scFv sequence as illustrated in figure 3.4.1.1 below.

A)			<-----FR1-IMGT-----><-----CDR1-IMGT----->
	Query_1	4	Q L Q Q S G T G L A R P G A S V K L S C K A S G Y T F T S Y
V 66.9% (194/290)	IGHV1-18*04	7	CAGCTGCAGCAGAGCGGCACCGCTGGCCAGACCGGCCAGCGTGAAGCTGAGCTGCAAGGCCAGCGGTACACCTTACCAGCTAC 93
			GT....TCT..AG..T.AGG..AAG..AG..T..G...TCA.....G..CTC.....TTCT..T.....T..... 96
V 66.9% (194/290)	IGHV1-46*01	7	Q L V Q S G A E V K K P G A S V K V S C K A S G Y T F T S Y 96
V 66.9% (194/290)	IGHV1-46*03	7	GT....TCT..GG..T.AGG..AAG..AG..T..G...TCA.....G..TTC.....ATCT..A..... 96
			--><-----FR2-IMGT-----><-----CDR2-IMGT-----><-----
	Query_1	94	G I S W V T Q R A G Q G L E W I G V I Y P R S G N T Y Y N E
V 66.9% (194/290)	IGHV1-18*04	97	GGCATCAGCTGGGTGACCCAGAGAGCGGCCAGGGCCTGGAGTGGATCGGCGTGATCTACCCAGAAAGCGGCAACCACTACTACAACGAG 183
			..T.....CGA...GCC..T..A..A..G..T.....G..ATG...AG..G..TTAC..AT..T.....AA...TGAC.. 186
V 66.9% (194/290)	IGHV1-46*01	97	G I S W V R Q A P G Q G L E W M G W I S A Y N G N T N Y A Q 186
V 66.9% (194/290)	IGHV1-46*03	97	TAT..GCA.....CGA...GCC..T..A..A..G..T.....G..AA..A..A...T..TG..T..T..G...AAG...GCAC.. 186
			TAT..GCA.....CGA...GCC..T..A..A..G..T.....G..AA..A..A...T..TG..T..T..G...AAG...GCAC.. 186
			-----FR3-IMGT-----
	Query_1	184	K F R G K A T L T A D K S S S S A Y M E L R G L T A E D S A
V 66.9% (194/290)	IGHV1-18*04	187	AAGTTTCAGAGGCAAGGCACCCCTGACGCGCCAGCAAGAGCAGCAGCGCTACATGGAGCTGAGAGGCCGTGACCGCCGAGGACAGCGCC 273
			...C..CAG...GA..T...A...A...A...CATC..CG...CA.....GA...T...C...CG... 276
V 66.9% (194/290)	IGHV1-46*01	187	K L Q G R V T M T T D T S T S T A Y M E L R S L R S D D T A 276
V 66.9% (194/290)	IGHV1-46*03	187	...CAG...GA..T...A...A...AGG...C..TC..CG...CA..T.....CA.....GAT..T.....CG... 276
			...CAG...GA..T...A...A...AGG...C..TC..CG...CA..T.....CA.....GAT..T.....CG... 276
			-----><-----CDR3-IMGT-----><-----FR4-IMGT----->
	Query_1	274	V Y F C A R E N Y G S V Y W G Q G T
V 66.9% (194/290)	IGHV1-18*04	277	GTGTACTTCTGCGCCAGAGAGAATACGGCAGCGTGTACTGGGCCAGGGCACC 327
			T..A...T..G..... 296
V 66.9% (194/290)	IGHV1-46*01	277	V Y Y C A R 296
V 66.9% (194/290)	IGHV1-46*03	277	T..A...T..T..... 296
D 100.0% (7/7)	IGHD1-7*01	11	----- 17
D 100.0% (7/7)	IGHD4-17*01	3	----- 9
D 100.0% (7/7)	IGHD4-23*01	3	----- 9
J 100.0% (17/17)	IGHJ1*01	17	----- 33
J 94.4% (17/18)	IGHJ4*02	12	----- 29
J 93.8% (15/16)	IGHJ5*02	17	----- 32

B)			<-----FR1-IMGT-----><-----CDR1-IMGT-----><-----
	Query_1	396	T Q S Q K F M S T S V G D R V S I T C K A S Q N V G T A V
V 67.2% (184/274)	IGKV1-13*02	11	TGACTCAATCTCAAAAGTTTATGCTACTTCTGTTGGTGATCGTGTCTATTACTTGAAGGCTAGTCAAAAGCTTGGTACTGCTGTTG 485
			...C..G...C..TCC..CCC...G..A...A...A...CA..A..CA..C...CCG...A...GGG..A...A..C..G...T..A.. 100
V 67.2% (184/274)	IGKV1D-13*02	11	T Q S P S S L S A S V G D R V T I T C R A S Q G I S S A L 100
V 66.8% (183/274)	IGKV1D-13*01	11	...C..G...C..TCC..CCC...G..A...A...A...CA..A..CA..C...CCG...A...GGG..A...A..C..G...T..A.. 100
			-----FR2-IMGT-----><-----CDR2-IM-----><-----
	Query_1	486	A W Y Q Q K P G Q S P K L L I Y S A S N R Y T G V P D R F T
V 67.2% (184/274)	IGKV1-13*02	101	CTTGATACCAACAAAAGCAGGTCAATCTCCAAAGCTTCTTATTACTTGCTTCAACAGATACACTGGTGTCCAGATAGATTACTG 575
			.C....T..G..G..A...GA...G...T...C..G...C..TGA...C..C..GTTTGG..A..G...C...TCA..G...C..GC.. 190
V 67.2% (184/274)	IGKV1D-13*02	101	A W Y Q Q K P G K A P K L L I Y D A S S L E S G V P S R F S 190
V 66.8% (183/274)	IGKV1D-13*01	101	.C....T..G..G..A...GA...G...T...C..G...C..TGA...C..C..GTTTGG..A..G...C...TCA..G...C..GC.. 190
			.C....T..G..G..A...GA...G...T...C..G...C..TGA...C..C..GTTTGG..A..G...C...TCA..G...C..GC.. 190
			-----FR3-IMGT-----><-----CDR3-IM-----><-----
	Query_1	576	G S G S G T D F T L T I S N M Q S E D L A D Y F C Q Q F S S
V 67.2% (184/274)	IGKV1-13*02	191	GTAGTGGTTCTGGTACTGATTTTACTTTGACTATTAGTAACATCTGAAGATTGGCTGATTACTTTTGTCAACAATTTAGTAGTT 665
			.C....A...G..A...C...C..C..C..C..G..C...GC.....T..AAC...T..AC.....G...A... 280
V 67.2% (184/274)	IGKV1D-13*02	191	G S G S G T D F T L T I S S L Q P E D F A T Y C Q Q F N S 280
V 66.8% (183/274)	IGKV1D-13*01	191	.C....A...G..A...C...C..C..C..C..G..C...GC.....T..AAC...T..AC.....G...A... 280
			.C....A...G..A...C...C..C..C..C..G..C...GC.....T..AAC...T..AC.....G...A... 280
			GT-----><FR4>
	Query_1	666	Y P Y T F G
V 67.2% (184/274)	IGKV1-13*02	281	ACCCATACACTTTTGG 681
			 284
V 67.2% (184/274)	IGKV1D-13*02	281	Y P 284
V 66.8% (183/274)	IGKV1D-13*01	281	 284
J 100.0% (11/11)	IGKJ2*01	3	----- 13
J 100.0% (9/9)	IGKJ2*02	4	----- 12

Figure 3.4.1.1: IgBLAST analysis of α BCK-1 (scFV) extracted (A) heavy-chain and (B) light chain nucleotide sequences compared to previously identified immunoglobulin germline variable region gene sequences. Variations from the germline sequences are depicted in pink.

Next CLC genomics workbench was used to align and compare the reliability of the V-gene sequence. Once verified the sequence was genetically modified with the insertion of a Gyl(4) Ser linker and *SfiI* and *NotI* restriction enzyme recognition sites as demonstrated in figure 3.4.1.2.

GGGCCAGCCGGCCgcccagctgcagcagagcggcaccggcctggccagaccggcgccagcgtgaagctgagttgcaaggccagcggctac
 accttcaccagctacggcatcagctgggtgacccagagagccggccagggcctggagtggatcggcgtgatctacccagaagcggcaacacctac
 tacaacgagaagttcagaggcaaggccaccctgaccgccgacaagagcagcagcagcgctacatggagctgagaggcctgaccgccgaggaca
 gcgccgtgtacttctgcgcagagagaactacggcagcgtgtactggggccagggcaccGGAGGTGGCGGCTCCGGAGGTGGAGGC
 AGCGGAGGGGGCGGATCCgacatcgtgatgaccagagccagaagttcatgagcaccagcgtgggacagagtgagcatcacctgcaa
 ggccagccagaacgtgggcaccgctggcctggtaccagcagaagcccggccagagccccaagctgctgatctacagcgcagcaacagataca
 ccggcgtgcccagagattcaccggcagcggcagcggcaccgacttcaccctgaccatcagcaacatgcagagcgaggacctggccgacttct
 gccagcagttcagcagctaccctacacctcggcggcgccaccaagctggagatcGCGGCCGC

Figure 3.4.1.2: Assembled ZIKA-DIII (scFv) sequence for cloning. The different components of the sequence are highlighted in different colours; dark orange is the Sfi1 recognition site, green is the heavy chain CDRs, light orange is the Gly(4)Ser linker, purple are the light chain CDRs and red is the Not1 recognition site.

3.4.2 In-silico cloning of pCB-ZIKA_DIII-snap

An ORF for the pCB-ZIKA-DIII-snap construct was generated through in-silico insertion of the assembled scFv sequence into the pCB-SNAP backbone as shown in figure 3.4.2.1. As with all other pCB-constructs designed in this project the construct included an enterokinase recognition sequence essential for generating the final active product by removing the N-terminal pre-peptide. The construct also had an Ig-Kappa leader sequence and poly HIS-tag located at the N-terminus of the SNAP-tag while the scFv sequence was located at the 3'.

In-silico cloning in snapgene software confirmed that the assembled scFv genes would be compatible for ligation into the pCB-snap backbone. After the final assembly the construct was checked for the presence of inappropriate stop codons and translated into its amino-acid sequence which was aligned to its original parental (rothe article) sequence. This was to validate the integrity of the original CDRs and FRs prior to having the sequence synthesised at Genescript (USA).

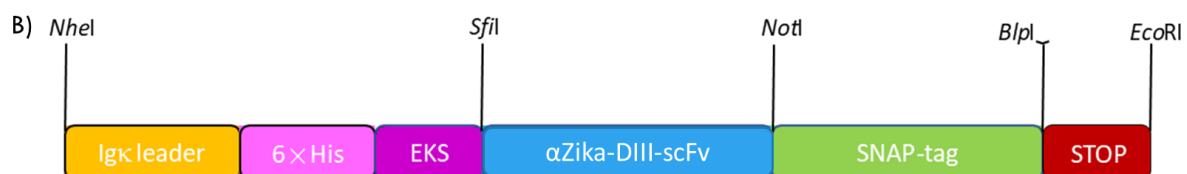


Figure 3.4.2.1: pCB-αZIKA-DIII-Snap construct. (A) In-silico restriction cloning of αZIKA-DIII (scFv) into pCB-AnnV-SNAP using Sfi1 and Not1. (b) ORF of pCB-αZIKA-DIII-SNAP construct including Sfi1 and Not1 recognition sites; IgK leader: Ig-*Kappa* leader sequence important for secretion of the fusion protein during host expression, 6xHis : 6 histidine tags used for protein purification during IMAC and detection in western blot analysis, EKS: enterokinase cleavage site important for enzymatic removal of the N-terminal elements and STOP: stop codon for termination of protein synthesis.

3.4.3 Molecular cloning of pCB-αZIKA-DIII-snap

A diverse set of experiments were performed collectively to clone the pCB-αZIKA-DIII-SNAP construct, these included DNA amplification, restriction enzyme digest and T4 DNA ligation.

3.4.3.1 Bulk preparation of DNA plasmids by transformation

Genescript synthesized αZIKA-DIII scFv plasmids were bulk prepared together with pCB-AnnV-SNAP by transformation in DH5α cells. Both cultures were grown at 37 °C overnight followed by plasmid DNA isolation from each culture individually. DNA isolation was performed with the phenol/chloroform DNA isolation method, followed by quantification using spectrophotometry.

Table 3.4.3.1.1: Purity of DNA extracted from bulk preparations.

Plasmid	A260/230	A260/280	Concentration (ng/ul)
Puc57_αZIKA-DIII	2.011	1.820	37564
pCB_AnnV_Snap	2.115	1.734	41041

The A260/230 ratios of both DNA samples were between 2.0 and 2.2 indicating that they were free of contaminating organic compounds. The A260/280 ratio was also within the acceptable range of 1.7-2 for DNA suggesting the sample was pure and could be used for cloning.

3.4.3.2 Restriction enzyme digest and agarose gel electrophoresis

The puc57 plasmid housing the αZIKA-DII scFv was double digested with *Sfi1* and *Not1* in preparation for ligation. The digested products were run on a 1.2% agarose gel for visualization of the digested fragments as seen in figure 3.4.3.2.1 below.

Table: Description of agarose gel lanes depicted in figure 3.4.3.2.1

Lane	Construct	Enzyme/control	Expected fragment size/s (bp)
1	1KB molecular weight marker	NA	Multiple
2	Puc57-αZIKA-DIII	Undigested	3421
3	Puc57-αZIKA-DIII	<i>Sfi1</i> + <i>Not1</i>	2710+ 711
4	Puc57-αZIKA-DIII	<i>Sfi1</i>	3421
5	Puc57-αZIKA-DIII	<i>Not1</i>	3421
6	Puc57-αZIKA-DIII	<i>Sfi1</i> + <i>Not1</i>	2710+ 711
7	Puc57-αZIKA-DIII	<i>Sfi1</i> + <i>Not1</i>	2710+ 711
8	Puc57-αZIKA-DIII	<i>Sfi1</i> + <i>Not1</i>	2710+ 711
9	Puc57-αZIKA-DIII	<i>Sfi1</i> + <i>Not1</i>	2710+ 711

The *Sfi*I and *Not*I double digest of puc57- α ZIKA-DIII was successful resulting in the generation of a 711bp insert, highlighted in red on the gel in figure 4.3.3.2.1. Single enzyme controls were also included to make sure that both *Sfi*I and *Not*I were enzymatically active thus resulting in

linearization of the plasmids. The 711bp fragments from lanes 3, 6,7,8 and 9 were excised from the gel and pooled into a single tube. The DNA was then purified using the Qiagen quick gel extraction kit. The resulting DNA was then quantified by spectrophotometry.

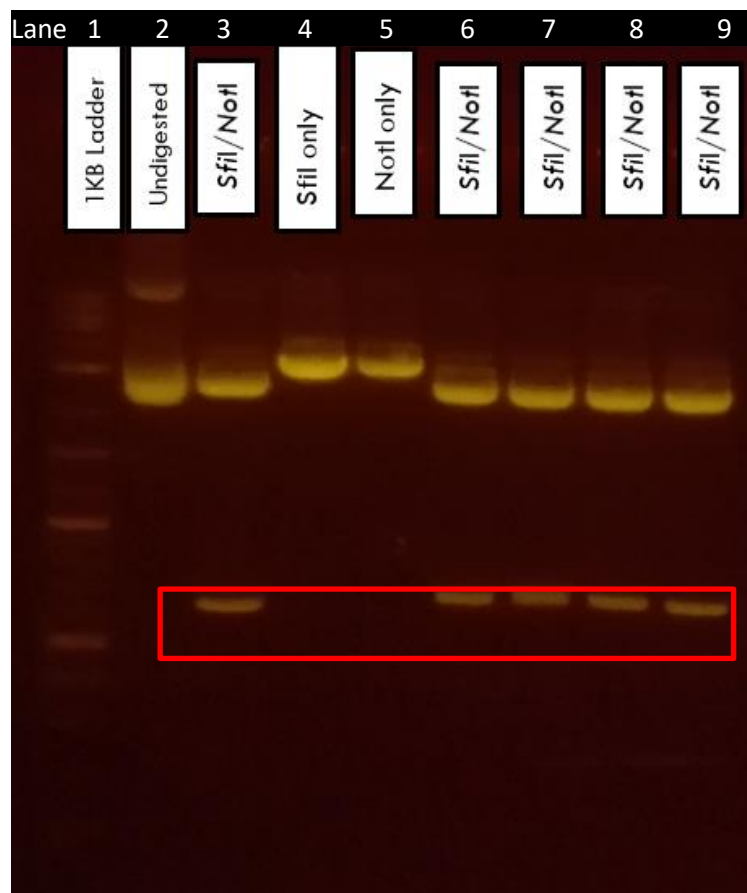


Figure 3.4.3.2.1: Agarose gel depicting DNA fragments after *sfi*I and *Not*I double digest: 711bp α ZIKA-DIII insert was extracted from gel for ligation

3.4.3.3 Ligation of α ZIKA-DIII (scFv) into the pCB-SNAP backbone

T4 DNA ligation was used to join the α ZIKA-DIII scFv insert and the pCB-SNAP vector which had been previously digested and isolated in section 3.4.3.2 above. T4 DNA ligase was used to add 3' hydroxyl groups and 5' phosphate groups to the cohesive ends of the DNA allowing for the formation of phosphodiester bonds necessary in ligation. The ligation reaction was set up in 3 different vector to insert molar ratios; 1:1, 1:3 and 1:5. The ligated products were transformed into chemically competent DH5 α *E.coli* cells and plated on LB agar plates containing 200 μ g/ml ampicillin. The plates were incubated overnight to allow individually colonies to grow as shown in figure 3.4.3.3.1.

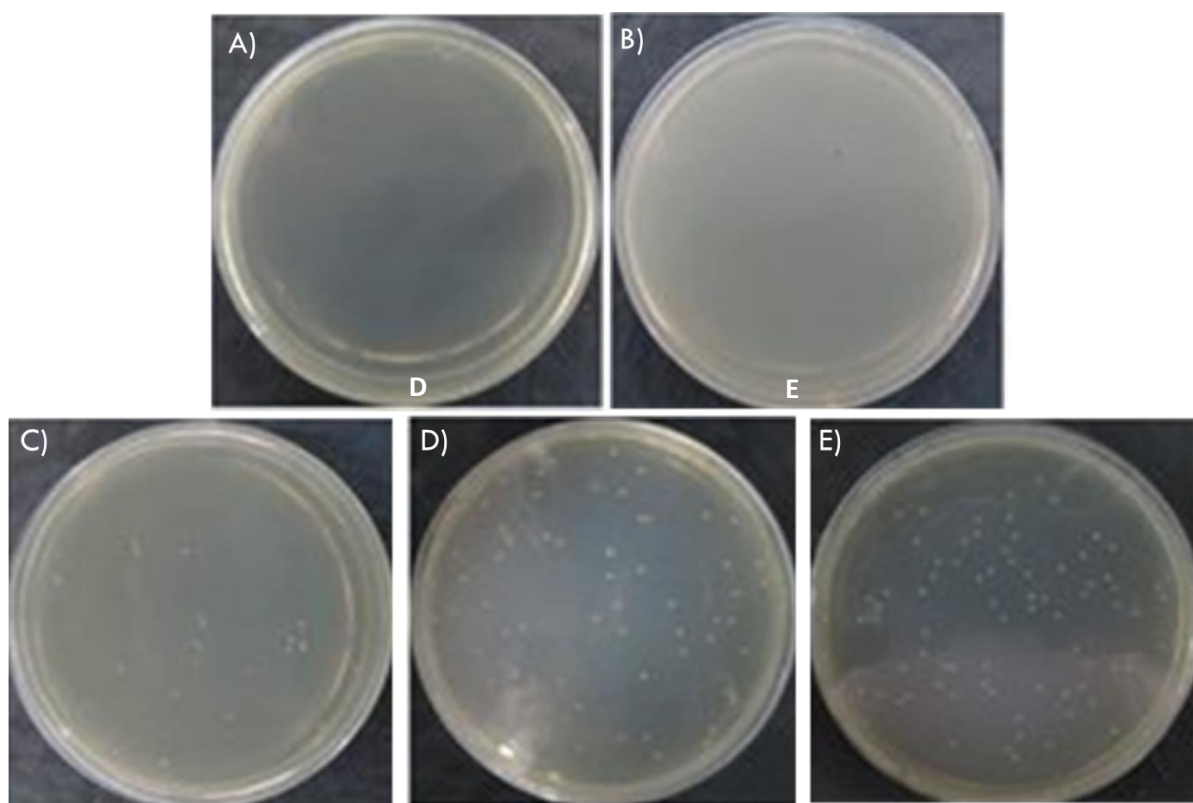


Figure 3.4.3.3.1: Growth of *E. coli* cells transformed with potential recombinant pCB- α ZIKA-DIII-snap plasmid DNA. Following transformation, 100 μ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C. The concentration of ampicillin used was 200 μ g/ml. **(A)** Contamination control **(B)** Vector-only plate **(C)** 1:1 ratio of vector to insert **(D)** 1:3 ratio of vector to insert **(E)** 1:5 ratio of vector to insert.

As shown in plate A of figure 3.4.3.3.1 no bacterial or fungal colonies were observed on the contamination control plate confirming the reliability of the experiment. Plate B also showed no colonies indicating that no vector relegation had occurred, this was expected as the plates were supplemented with ampicillin which would restrict the growth of any untransformed bacterial clones. Contrastingly clones containing the pCB- α ZIKA-DII plasmid would be expressing β -lactamase, an enzyme which catalyzes the hydrolysis of β -lactam ring of ampicillin allowing for growth of successfully transformed clones, such as those observed on plates C, D and E. While successful transformants were observed for all ligation reactions, the 1:3 and 1:5 vector to insert reactions had better transformation efficiencies. Twelve

different colonies were picked at random from the three plates and cultured in 5 ml LB agar broth in preparation for restriction mapping.

3.4.3.4 Restriction mapping of potential recombinant pCB-αZIKA-DII-SNAP clones

Successfully ligated pCB-αZIKA-SNAP clones were verified with restriction mapping. To determine which enzymes would differentially fragment the recombinant clones from its parental counterparts, a virtual simulation was carried out using SnapGene software. Five different enzymes were run on the simulation, from which BamH1 and PVUII were chosen for the restriction mapping.

The gel image obtained (Figure 3.4.3.4.1) showed distinct fragment patterns for the recombinant and parental plasmids that were in accordance with the theoretical gel simulated from SnapGene® (Figure 3.4.3.4.2).

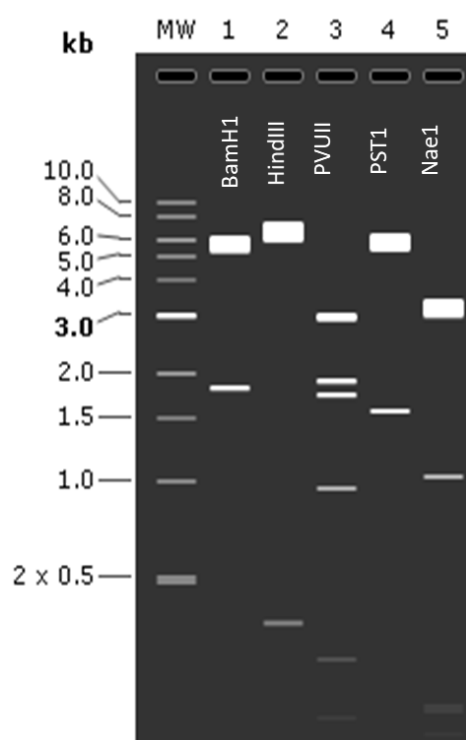


Table: Description of snapgene agarose gel simulation

Lane	Construct	Enzyme	Expected fragments (bp)
MW	DNA Ladder	NA	Multiple
1	pCB-ZIKA-DIII-snap	BamH1	6088, 1879
2	pCB-ZIKA-DIII-snap	HindIII	7657, 310
3	pCB-ZIKA-DIII-snap	PVUII	3001, 1924, 1776, 900, 235, 131
4	pCB-ZIKA-DIII-snap	PSTI	6192, 1775
5	pCB-ZIKA-DIII-snap	NaeI	3212, 3103, 1264, 180, 179

Figure 3.4.3.4.1: Theoretical gel simulation using Snapgene and corresponding band sizes for each enzyme.

Table: Description of agarose gel in figure 3.4.3.4.2

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	Colony 1	Undigested	7967
3	Colony 1	BamH1	6088, 1879
4	Colony 1	PvuII	3001, 1924, 1776, 900, 235 ,131
5	Colony 2	Undigested	7967
6	Colony 2	BamH1	6088, 1879
7	Colony 2	PvuII	3001, 1924, 1776, 900, 235 ,131
8	Colony 3	Undigested	7967
9	Colony 3	BamH1	6088, 1879
10	Colony 3	PvuII	3001, 1924, 1776, 900, 235 ,131
11	Colony 4	Undigested	7967
12	Colony 4	BamH1	6088, 1879
13	Colony 4	PvuII	3001, 1924, 1776, 900, 235 ,131
14	Colony 5	Undigested	7967
15	Colony 5	BamH1	6088, 1879
16	Colony 5	PvuII	3001, 1924, 1776, 900, 235 ,131
17	Colony 6	Undigested	7967
18	Colony 6	BamH1	6088, 1879
19	Colony 6	PvuII	3001, 1924, 1776, 900, 235 ,131
20	100 KB DNA Ladder	NA	Multiple

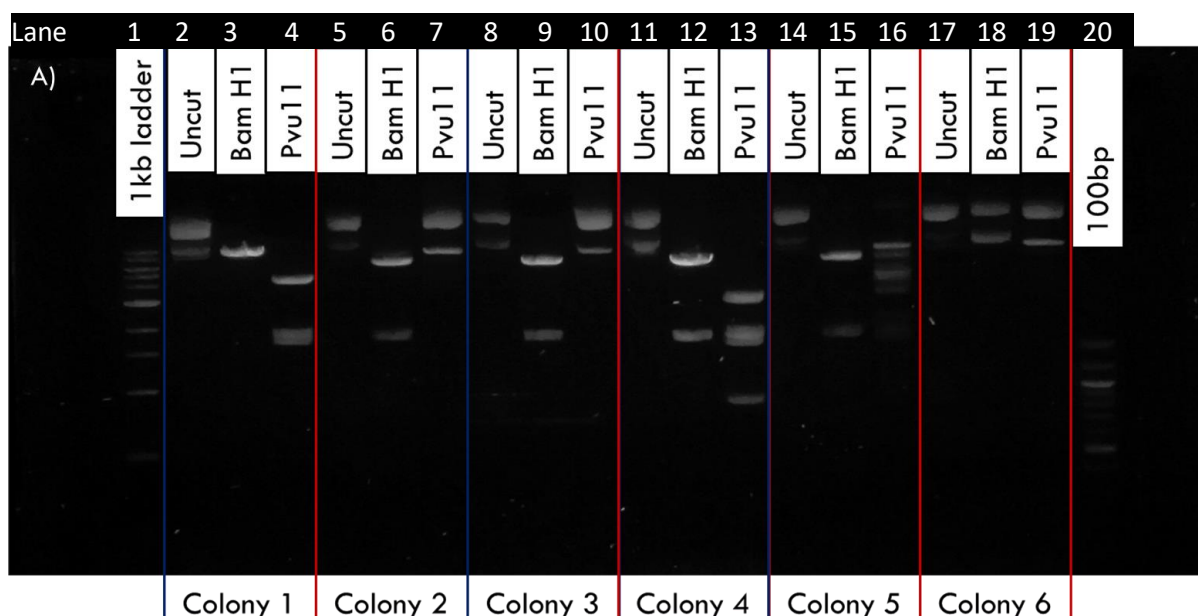


Figure 3.4.3.4.2 . Analysis of restriction mapping of potential pCB-ZIKA-DIII-Snap clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.

Table: Description of agarose gel in figure 3.4.3.4.3

Lane	Construct	Enzyme	Expected fragments (bp)
1	1KB DNA Ladder	NA	Multiple
2	Colony 7	Undigested	7967
3	Colony 7	BamH1	6088, 1879
4	Colony 7	PvuII	3001, 1924, 1776, 900, 235 ,131
5	Colony 8	Undigested	7967
6	Colony 8	BamH1	6088, 1879
7	Colony 8	PvuII	3001, 1924, 1776, 900, 235 ,131
8	Colony 9	Undigested	7967
9	Colony 9	BamH1	6088, 1879
10	Colony 9	PvuII	3001, 1924, 1776, 900, 235 ,131
11	Colony 10	Undigested	7967
12	Colony 10	BamH1	6088, 1879
13	Colony 10	PvuII	3001, 1924, 1776, 900, 235 ,131
14	Colony 11	Undigested	7967
15	Colony 11	BamH1	6088, 1879
16	Colony 11	PvuII	3001, 1924, 1776, 900, 235 ,131
17	Colony 12	Undigested	7967
18	Colony 12	BamH1	6088, 1879
19	Colony 12	PvuII	3001, 1924, 1776, 900, 235 ,131
20	100 KB DNA Ladder	NA	Multiple

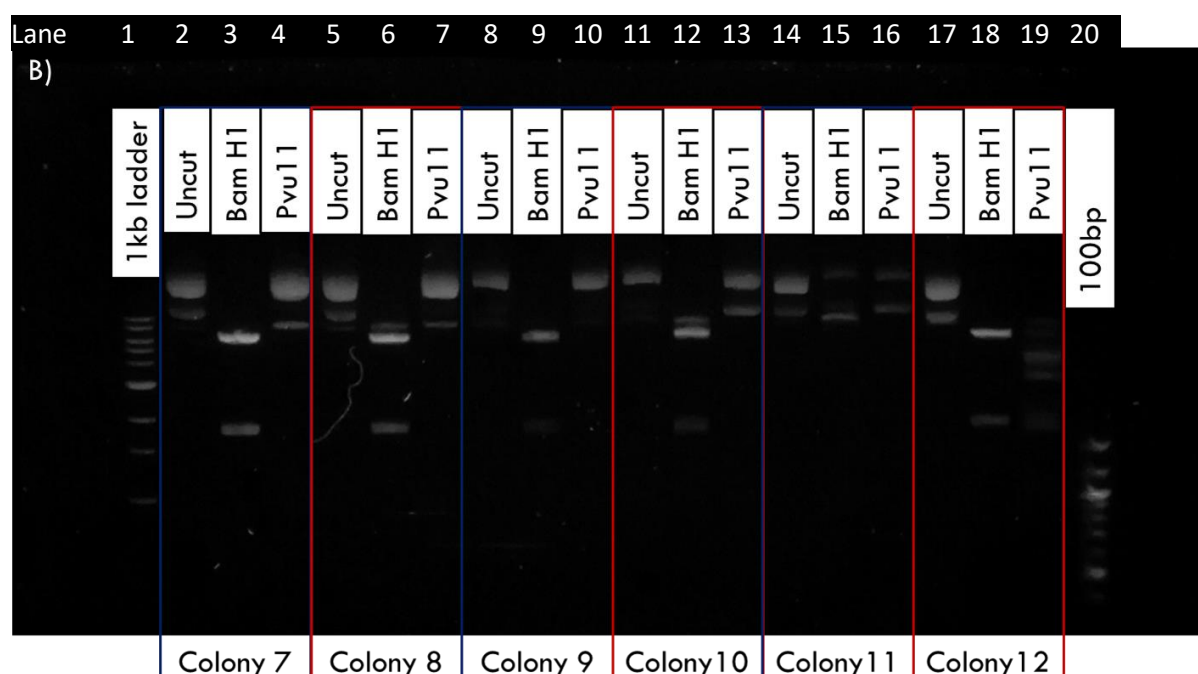


Figure 3.4.3.4.3 . Analysis of restriction mapping of potential pCB-ZIKA-DIII-Snap clones using agarose gel electrophoresis. All digestions were run overnight at 37°C. DNA fragments were separated using a 1.2% (w/v) agarose gel run for 1 hour at 100 volts. Units of the DNA ladder (NEB, UK) are in kilobase pairs.

Colonies 4 and 12 were chosen based on their similar banding pattern to that observed in the gel simulation. The chosen colonies were bulk prepped and sent to Inqaba Biotechnical Industries (South Africa) for sanger sequencing to confirm that the cloning was successful and had not resulted in any mutations.

3.4.3.5 Sequencing results of selected pCB-ZIKA-DIII-SNAP clones

Sequence results received from Inqaba were aligned to the original in-silico designed sequence in SnapGene to confirm the integrity of each clone. Figure 3.4.3.5.1 below shows that both aligned clones share 100% homology with the theoretical in-silico designed pCB-ZIKA-DIII-SNAP sequence confirming the successful incorporation of ZIKA_DIII(scFv) into the pCB-SNAP backbone. The plasmid was also checked for mutations, insertions, deletions, and frameshifts in preparation for transfection.

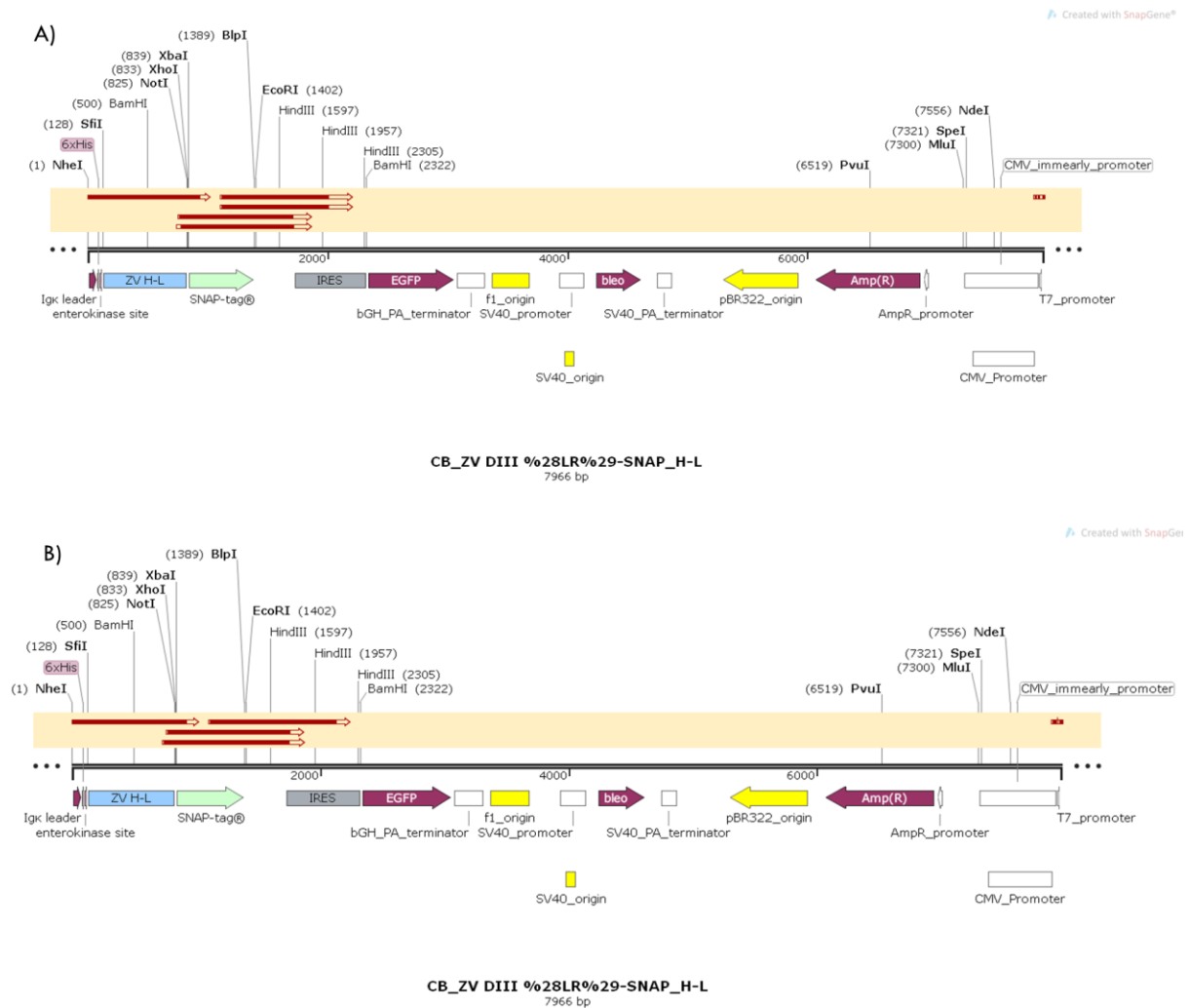


Figure 3.4.3.5.1: Map generated following the alignment of trace sequences with the corresponding in-silico sequence for pCB-ZIKA_DIII-SNAP. Both clones showed 100% homology with the in-silico ORF sequence as indicated by the dark red sections of the arrows.

3.4.4 Transfection and expression of pCB-ZIKA-DIII-SNAP in HEK293T mammalian expression cells.

Sequence verified pCB-ZIKA-DIII-SNAP clones were introduced into HEK293T cells for protein expression. Positively transfected cells, potentially expressing pCB-ZIKA-DIII-SNAP proteins, were identified by fluorescent cell imaging of the reporter protein GFP. At week one post-transfection the cultures looked healthy and had approximately 65-77-% eGFP expressing cells. However, after week one the cells morphology began to change and the GFP expression decreased. By week four very few cells were expressing GFP and many of the cells began to die (figure 3.4.4.1). The cultures were treated with 100ug/ml of Zeocin to enrich the GFP population but Zeocin was unable to rescue the culture. Transfection was re-attempted several times, always resulting in the death of all the cells within 3 months. Upon further investigation, we found that being of mammalian origin, HEK293 cells are susceptible to infection by Zika virus [152]. This indicated that our putative protein was likely being expressed and was actively infecting the expression cells resulting in cell death. We thus re-attempted the transfection process, starting several cultures at once, and harvested all of them from one week post-transfection. As the culture conditions were not optimal, we did not complete a purification of the culture as we were unsure if sufficient protein would be eluted from the IMAC column for downstream assessment and opted to test the unpurified protein for binding directing after harvesting.

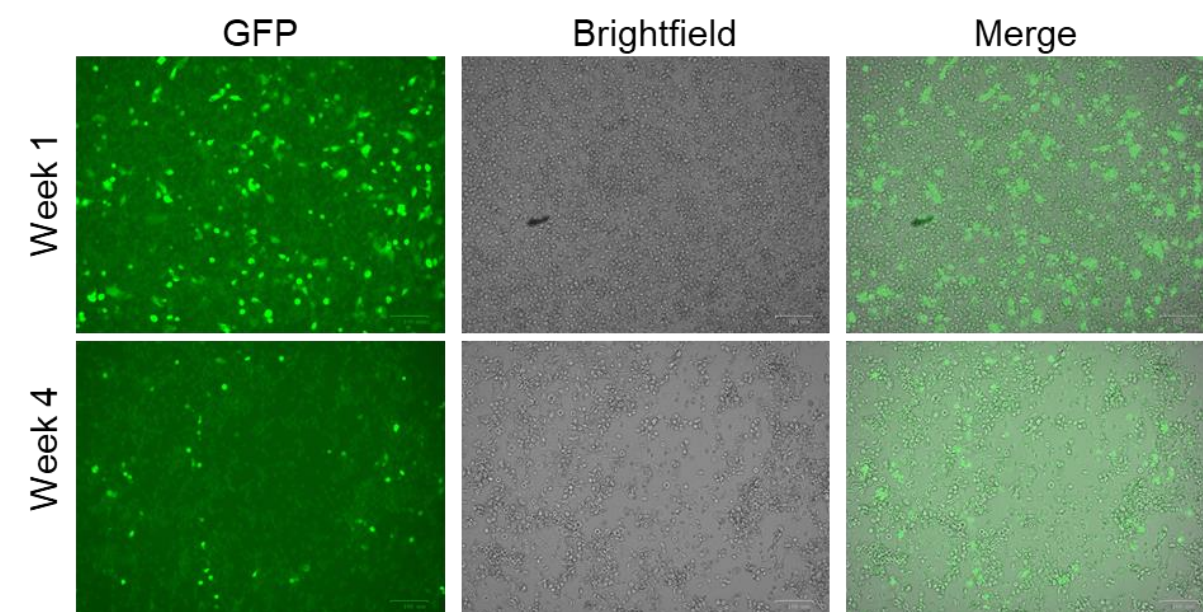


Figure 3.4.4.1: Green fluorescent protein expression in pCB-ZIKA-DII-Snap transfected HEK293T cells one week post transfection.

3.4.5 SDS-PAGE analysis of pCB-ZIKA-DIII-SNAP cell culture supernatant

Cell culture supernatant harvested from HEK293t culture were run on an SDS gel to determine if there was pCB-ZIKA-DIII-Snap proteins present in the supernatant. The supernatant was first concentrated with an 10KDa amicon filter to make sure that any protein present in the supernatant would be sufficiently concentrated to visualize on an SDS gel. Fifteen microliters of the supernatant was then loaded on a 10% SDS gel

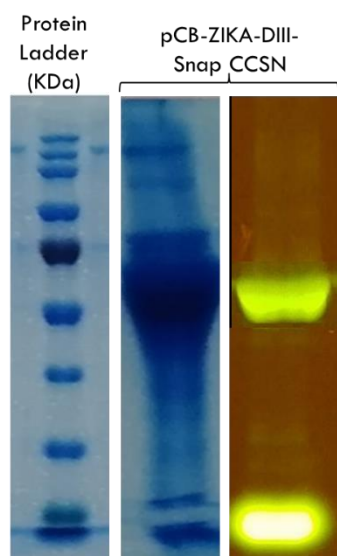


Figure 3.4.5.1: Assessing the presences and binding activity of pCB- α ZIKA-DIII-snap by SDS and conjugation to BG-Alexa Fluor 488. A ratio of 1:2 of supernatant to BG-Alexa Fluor 488 was used in the conjugation reaction to α ZIKA-DIII(scFv)-SNAP (51 kDa) . Left panel: the protein pre-stained ladder. The middle panel: Alexa488-conjugated CCSN ran on a 10% SDS-PAGE gel stained with Aqua staining solution. Right Results panel: The same SDS-PAGE gel visualized under blue light for potential fluorescence. The SDS-PAGE gel was run at 100 volts for 95 minutes and 15 μ l of conjugated supernatant was loaded per well. Page Ruler prestained protein ladder (Thermo Fisher Scientific, South Africa) was used to confirm the protein sizes. A Dark Reader Transilluminator was used for visualization of the fluorescent signal.

As depicted in figure 3.4.5.1 pCB-ZIKA-DIII protein was indeed expressed by the HEK293t cell cultures as evident by the 51 KDa bands. The panel on the right of the figure also shows that the SNAP-tag within the protein was active and bound specifically to the BG-modified Alexa flour 488 fluorophore. These results were an important first step in demonstrating the versatility of SNAP-tag technology in non-cancer specific targets.

4.3.6 Investigating the binding activity of pCB-ZIKA-Snap to target antigens

In addition to providing the antibody sequences used to design the pCB-ZIKA-DII protein, Prof. Marinho's team also developed an ELISA based plate assay which could be used to confirm specific binding of our snap-tag constructed protein. The ELISA plates were made using two different Zika virus strains, a low passage MR766 strain and Bahia strain. The gene encoding for the envelope protein of both strains were cloned into a mammalian vector system, amplified, purified and then precipitated following a PEG-8000 protocol. The envelope proteins were then fixed on an ELISA plate using a carbonate-bicarbonate buffer protocol. The plates were tested in Brazil with three primary antibodies that were, polyclonal antibody against ZIKV obtained from the ascitic fluid of ZIKV infected mouse, serum from a pregnant woman that was infected with ZIKV, and serum from a male patient infected with ZIKV. Once binding of the antibodies to the ELISA plates were confirmed the plates were shipped to Cape Town and used as a tool to verify binding of our pCB-SNAP-ZIKA-DIII fusion protein. Two different sets of ELISA plates were received from Brazil, set one was coated with 150 ng of envelope protein while set two was coated with 200 ng of envelope protein.

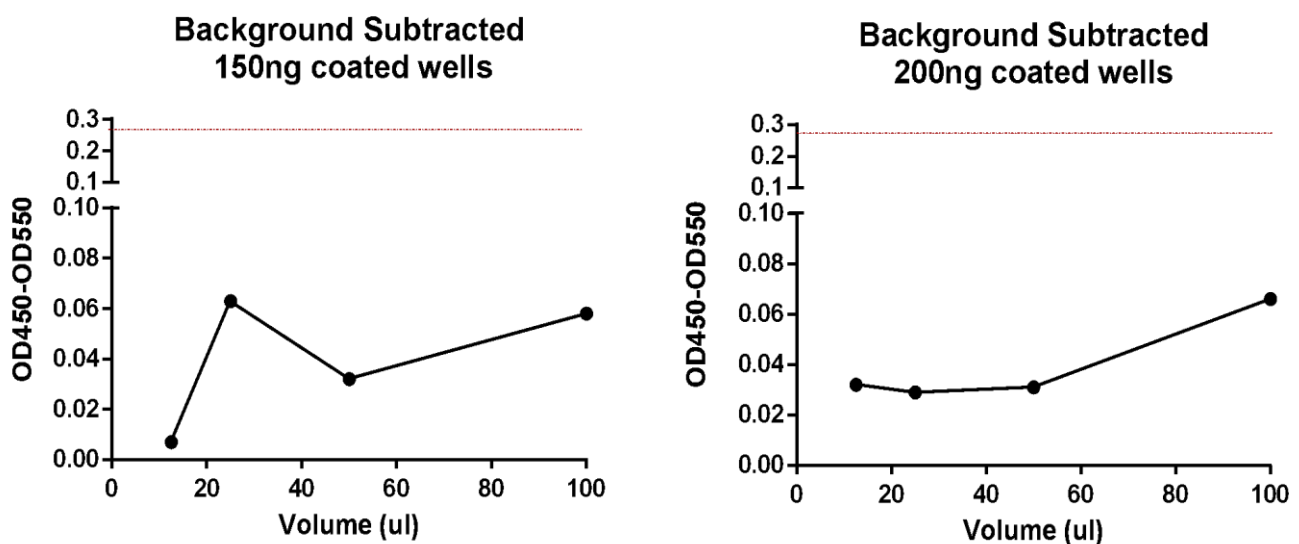


Figure 4.3.6.1: Enzyme-linked immunosorbent assay (ELISA) results of pCB-ZIKA-DIII-Snap concentrated cell culture supernatant (CCSN). 100 μ l of concentrated CCSN was coated onto wells of (A)150 ng or (B) 200ng envelope-protein pre-coated plates and incubated for 1 hour at room temperature. Optical density was measured at OD 450 nm. The red dotted line indicates the OD reading of a positive control plated at 100 μ g/ μ l.

Both sets of ELISA experiments demonstrated a dose dependent binding confirming both the presence and binding activity of the pCB- α ZIKA-DIII-SNAP protein.

Section 3.4.6: Summary of Results in Section 3.4

In this section we were able to apply the methodology set up for the developing TNBC relevant recombinant proteins to successfully generate a functional Zika virus targeting SNAP-tag based fusion protein. These results support the use of SNAP-tag technology for infectious disease applications. While there are still many pitfalls in the overall reproducibility and stability of the mammalian expression system, it seems it is still very possible to generate full length, functional SNAP-tag based fusion proteins using this methodology. This is an important first step in the development of immunotherapeutic applications for infectious diseases. Furthermore, the development of the pCB-ZIKA-DIII-SNAP protein will be used to study the infectious cycle of Zika virus within host cells. By conjugating the protein to a BG-modified fluorescent molecule and then infecting a host cell, the migration of viruses within the cell can be tracked with microscopic visualization providing a very valuable and innovative tool to understanding the infection cycle of zika virus.

Chapter 4: Discussion

4.1 Triple Negative Breast Cancer: A need for medical intervention

Breast cancer remains a major health concern globally, with approximately one million cases being diagnosed yearly. Of these, an estimated 170,000 cases are designated as triple negative breast cancer (TNBC) [153]. TNBC classically lack expression of ER, PR and HER2 and represent one of the most aggressive and clinically challenging breast cancer subtypes [154, 155]. In addition, TNBCs often present with a lung and brain involvement, have poorer prognosis and higher mortality rates [9, 153, 156]. The highest prevalence of this disease is found in premenopausal women of African origin, almost 3 times more likely than women of other racial genetics [154, 157]. Due to the complex nature of the disease, there is a severe lack of specific guidelines on the management of afflicted patients [158]. To make the situation worse, there is currently no effective, proven single agent therapy for TNBC. Even well-established targeted therapies such as tamoxifen, aromatase inhibitor and trastuzumab which help other breast cancer patients, show a lack of benefit in TNBC patients [158-160]. Consequently, the only option of these patients is traditional treatment options which have major limitations [129, 161]. For example, chemotherapeutic regimens using taxanes and anthracyclines persist as mainstream treatment for many cancers due to their preliminary sensitivity however their short diseases-free survival rates do not justify its medical merit [162, 163].

Furthermore, the identification of 6 subtypes of TNBC cancers, each with its own ontologies and unique gene expression profiles demonstrate the heterogeneity and acute diversity of this disease. Thus, illustrating the urgency to direct global research efforts towards stratification of patients based on prognostic factors or diseases specific molecular markers that can differentially predict responsiveness to targeted therapeutics. Based on these premises the Medical Biotechnology and Immunotherapy (MB&I) Research unit aim to establish an array of companion immunodiagnostics which when used in combination should be able to characterise TNBC tumours into various subpopulations based on their potential sensitivity to a specific immunotherapy. As such one of the major aims of this study was to develop one of the companion diagnostic panels, particularly for the classification of TNBCs based on their TIL and checkpoint inhibitor populations. This panel consisted of scFv's targeting T-cell receptors CD3, CD4 and CD8, B-cell receptors CD19 and CD20 and checkpoint inhibitory receptors PDL-1 and B7.1. All these markers play a significant role in predicting disease progression and assessing the sensitivity of a tumour to certain immunotherapies. Checkpoint inhibitor therapy for example, initially demonstrated very promising results, showing more durable responses than chemo or other targeted therapies. However, with the accumulation of clinical data over the last few years many of the drawbacks associated with this therapy have become evident. The main issue being a lack of efficacy, with only 10-30% of patients responding to therapy. Various mechanisms of non-responsiveness have been studied, and the factors found to contribute to non-responsiveness were tumour mutational burden, expression levels of PD1/PDL1, loss of MHC and IFN signalling. To address some of these issues understanding the tumour before deciding on an appropriate therapeutic approach is becoming more important. By utilizing a differential diagnostic

panel like the one described in this thesis, we would be able to identify the number of cells within a tumour expressing PDL1 and determine the level of expression of PDL1 on each expressing cell. This information is key to determining which patients would most likely benefit from PDL1 targeting immunotherapies, only patients with high enough levels of PDL1 would be treated with such a therapy thus preventing the unnecessary application of these drugs in patients who would likely not benefit from such a treatment. A similar scenario can be envisioned for the other targets in this panel. Patients with low number of T-cell and B-cells would not be treated with T-cell and B-cell targeting therapies respectively. Of course, this panel only contains a few of the biomarkers necessary for predicting efficacy of immunotherapies and several other biomarkers would be needed to comprehensively predict efficacy of all the currently available immunotherapies.

Moreover, most biomarker research currently available pertain mainly to a Caucasian population. There is very limited research identifying biomarkers that are applicable to people with African ancestry, who display the highest degree of genetic variations. Therefore, medicines developed today could be based on research where this population of people are underrepresented. The outcomes of such research would require further proof-of-concept to validate it in the African population. Thus, there is tremendous need to lessen the knowledge and treatment gap within this population to ensure proper management of the escalating disease burden and ensure effective treatment is administered. Consequently, understanding African precision medicine is of extreme importance, allowing for the correlation of African patient profiles to targeted-tailor-made therapeutics.

In summary the differential diagnostic panel proposed in this thesis, would provide the tools needed to link a diagnosis based on differential expression of disease markers with the most appropriate immunotherapy available. Once available, the patient profile data resulting from these kinds of precision medicine applications, might be able to identify cell surface disease markers with differential expression on diseased and healthy cells allowing for the discrimination between these cell types. This type of biomarker would be the ideal target for development of future antibody based immunotherapies.

4.2 Identifying potential therapeutic antibody sequences through data mining and patent searching.

In addition to assessing tumours based on immune-markers, characterising tumours based on diseases-specific antigens within a tumour is equally as important. Not only can tumour specific antigens be used for diagnostic purposes, but many disease-specific markers can be used as the target for novel immunotherapies. Thus, the more disease-specific antigens that can be identified the better the chances of finding a combination of antigens/ligands that when targeted simultaneously could eradicate all the cancer cells within a tumour. A good way to identify disease specific antigens with the potential to generate targeted therapeutics is through rigorous database and patent searches.

While the features of therapeutic antibodies can be found in various naturally sourced sequences , effective biotherapeutic sequences require tailored engineering for clinical safety and developability [95]. This scientific knowledge can be found in patents containing antibody sequences. Within this thesis, seven out of the nine constructs generated utilized antibody sequences from patents. During our analysis of various patents, we found that majority of them were explicitly developed for therapeutic antibodies many of which were already approved for clinical application. Furthermore, other antibody sequences identified were antibody sequences in late-stage clinical trials or closely resembled those in clinical trials. Many of these patents provide information on the developmental choices made during production of a successful therapeutic antibody. By utilizing this information, it is possible to streamline the development of new therapeutics and bypass the lengthy experimental route usually taken to identify possible new targets for immunotherapeutic.

However, this does not come without its caveats, with no standardization of deposited sequences and legal protection afforded by patents might only allow the use of such information for research purposes [95]. Furthermore, in the context of a heterogenous disease such as TNBC, the discovery of new antibody sequences allowing for a granular stratification of various subtypes would take an extensive amount of time unless focused on specifically. Thus, it is still necessary and useful to develop tools and techniques that allow us to identify novel antibodies for specific applications. To this end, one of the most widely used methods of identifying and isolating novel antibodies is phage display.

Despite its limitations, patents currently offer a rich source of information to those who can extract it, as exhibited by the extraction of 7 antibody sequences utilized for the generation of fusion proteins within this research. The number of antibody sequences available in patents continues to rise and provide a wealth of knowledge on antibody engineering, accelerating delivery of biotherapeutics for clinical application.

4.3 Phage Display as a tool for novel antibody identification

Bacteriophages often referred to as phages, are stable, ubiquitous, and extraordinarily robust viruses that are harmless to humans but have the capacity to infect bacteria. In the early 1900s one of the first applications exploiting the phages ability to infect bacterial cells was a therapeutic tool against bacterial infections [164, 165]. With the development of antibiotics, this specific use was no longer necessary. A few centuries later, in the late 1980s and 1990s came the invention of phage display for peptides and proteins which lead to the use of phage display for various applications including the production of antibodies [166]. One of the key driving factors supporting the use of phage display was that it rarely required the inoculation of animals making the production of phages both simple and economical.

In all phages the RNA or DNA genome is encapsulated by a coat protein. This coat protein can be readily conjugated or genetically engineered to display proteins, peptides, and antibodies. There are various types of phages but the *Escherichia coli* phages, which include the M13 filamentous phages and their close relatives are the most widely used due to their quick amplification and ease of culture, making their propagation fast and economically feasible. Phage display has contributed to the advancement of immunotherapy and vaccine development and has revolutionized the development of therapeutic antibodies. About 40% of therapeutic antibodies in the 20 top-selling biotechnology drugs were obtained using phage display.

There are two phases to the phage display technology: construction of antibody libraries and screening of antibody libraries. For library construction, large repertoires of human antibody genes, usually single-chain variable region antibody fragment (scFv) or Fab genes, are fused to the amino terminus of the phage coat protein (pIII) in an engineered phage or phagemid vector. *E. coli* is infected to generate a library containing approximately 10^{10} cells by inoculating the library with an additional helper phage. Subsequently, scFv-presenting or Fab-presenting phage particles are secreted from the host bacterial cells [137, 138]. The established library is then screened, to isolate the phage antibodies of interest. This involves the capture of the target antigen, either by immobilization on a plastic surface (solid phase) or coupled to magnetic beads (solution phase). Incubation of the captured antigen with the purified virions in solution allows binding of specific antibody variants. The nonbinding phage particles are removed with multiple washes, while attached phage are typically eluted under either low- or high-pH conditions to disrupt the antibody-antigen association [139]. These phage particles are amplified by infecting an *E. coli* culture. This initial selection event of purified, amplified phage antibodies is used for multiple rounds of panning to reduce nonspecific or “background” binding events and to enrich for positive binders [138].

The design of an appropriate selection strategy can vastly impact the outcome of phage display technology. The most common strategies include biopanning on immobilized antigen, coated onto solid supports, columns, or sensor chips, selection using biotinylated antigen, panning on fixed mammalian cells, enrichment on tissue sections, selection on paramagnetic liposomes, and, selection using living animals, as reported for peptide phage libraries [140,

141]. For tumor-specific antibodies, standard phage display panning makes use of purified proteins, antigen-transfected cells or tumor cell lines. However, expression and purification of recombinant proteins, in their native conformation, can be challenging; suitable cell lines are not always available and antigen expression can change after several passages and thus differ from malignant cells, *in vivo* [142]. Thus, selection of tumor-specific antibodies directly on tissue biopsies allows for the selection of clinically relevant epitopes. Importantly, immobilization of the tissue biopsy, by formalin-fixation, limits this selection strategy to the isolation of only surface-binding antibodies [143]. Alternatively, the whole-cell, or intact cell, selection strategy allows for the selection of internalizing scFv antibodies, which are beneficial for the development of novel immunotherapeutic agents, as receptor-mediated endocytosis plays a crucial role in drug delivery to target cells. Furthermore, panning on intact cells allows the inclusion of native cell surface proteins and non-protein antigens, thereby minimizing the selection of non-specific scFv antibody fragments or denatured epitopes that would not be recognized by the native protein *in vivo* [144].

The identification of novel internalizing scFv antibodies has immense implications on drug delivery and tumor modulation applications. Specific molecules targeted to TNBC cells, such as immunotoxins (cytotoxic fusion proteins, CFPs) have shown promising results [145]. Previously, the genetic fusion of an EpCAM-selective scFv to a mutated granzyme B (Gb), resulted in a fully human targeted CFP that inhibited the TNBC tumor growth, *in vivo* [130]. More recently, a novel targeted approach for the elimination of CSPG4-positive TNBC, *in vivo*, was confirmed by the activity of a high affinity CSPG4-specific scFv genetically fused to a functionally enhanced form of the human microtubule-associated protein (MAP) tau [146].

Despite its ease of use and its versatility, there are very few scientists and experts in South Africa with the skills to successfully utilize this technology. As such, one of the aims within this thesis was to learn phage display and set up the technology at UCT. As part of a bigger MB&I aim, once available this technology would be used by various students to screen for and potentially isolate novel antibodies targeting different cancer antigens. For the purposes of training, the Tomlinson I + J libraries containing human single fold scFv's were chosen. Since all our constructs use the scFv format and our collaborators had access to this library, we thought it was well suited for our initial needs. Using this library, we were able to screen for and select Ki4 binding phages on Ki4 antigen coated ELISA plates, demonstrating the functionality of the library as well as the propagation and selection protocol.

The next steps involved setting up a dual target library containing phages with either an scFv specific for CSPG4 or EGFR. These antibodies were previously cloned into our pCB-SNAP-tag based vectors and showed specific binding to their target antigens on cell lines. Thus, we sought to use these validated scFv's to optimize a bio-panning process on living cells recently developed by the Barth group in Germany. To this end I was able to successfully clone α CSPG4-PIT2 and α EGFR-PIT2 constructs for propagation. However due to a lack of necessary infrastructure we were unable to proceed with propagation and optimization of the bio-panning process. Phages are notoriously difficult to contain and must be handled in a contained sterilized area. Initially, the MB&I unit had plans to set up such an area in the lab which would be dedicated to phage work. However, with the COVID19 lockdown some of these plans were derailed and delayed so the phage lab was never set up. Given that the research

in the MB&I lab is majorly focused on bacterial work, handling phage in the same area was not an option due to the high risk of contamination.

Once the lab can be set up for phage work, the optimization of the bio-panning can continue. While we were not able to achieve the aim of optimizing the technique, we were able to learn and understand how to propagate phage libraries. We were also able to demonstrate specific binding and selection of an antibody from the library and we were able to clone new antibody sequences into the PIT2 phagemid backbone. All of these are important initial steps in establishing the technology at UCT and lay the foundation for development of phage-based applications. Given the exponential growth in phage display applications, training of scientist with the skills to execute phage display is becoming more important and is a necessary requirement to bring the scientific and technological standards of South Africa on par with first world countries.

4.3 SNAP-tag Technology as an important tool in development of immunodiagnostic and immunotherapeutic fusion proteins.

Cancer immunotherapy has revolutionized the fields of oncology and immunology ever since the discovery that stimulating an immune response can produce an anti-tumour response [155]. This discovery refocused research efforts into understanding cancer biology [156], which resulted in the initial observation of antigen expression on tumour cells and the subsequent exponential expansion of antibody-based therapeutics to incomprehensible heights. Full-length mAbs targeting molecular markers that are enriched for on cancerous cells initially demonstrated substantial benefit in management of haematological malignancies and solid tumours like pancreatic cancer and HER2+ breast cancer [160, 167, 168]. However, most of the FDA approved mAb's were not curative, signifying the need to improve their therapeutic efficacy [169]. Theoretically, there are 3 main pathways in which mAb exert their therapeutic effect: 1) by reducing proliferation and apoptosis via binding of cell surface receptors expressed on tumour cells which inhibits downstream signalling, 2) by initiating ADCC, the complement cascade and activation of T-cell through Fc receptor signalling resulting in immune mediated killing of tumour cells and 3) by inhibition of tumour stroma and vasculature function [170, 171]. As many of the therapeutic effects exerted by an antibody are through stimulation of other immune components, a major prerequisite for clinical efficacy is obviously a functional host immune system [172]. Unfortunately, most cancer patients receive immunotherapies in conjunction with chemo and radiation therapy which likely results in a compromised immune system. Moreover, like many other diseases under immune pressure, cancer has evolved numerous escape mechanisms to avoid the effects of antibodies [171]. High dosages and repetitive administration have the potential ensure clinical success with the caveat of a heightened risk of developing adverse reactions and resistance [171, 173]. Thus, there has been a major effort on developing antibodies for the specific delivery of toxic agents to tumour cells rather than utilizing there naturally occurring therapeutic efficacy.

Such targeted drug delivery was originally achieved by the development of antibody-drug conjugates (ADC)s, bifunctional particles utilizing the specificity of mAbs for differentially expressed tumour associated antigens to deliver chemically conjugated chemotherapeutics to tumour cells [174]. Without the specificity achieved through conjugation, these chemotherapeutics would be to toxic for systemic application [175]. On the other hand, early ADC designs exposed that the choice of mAb carrier, the method of drug attachment and the toxic payload chosen need to be carefully designed for adequate clinical efficacy and that current designs required improvements [176]. Some of the common issues encountered with early ADC designs were as follows: 1) antibody binding was restricted or inhibited when more than a certain number of drug molecules were loaded on them, 2) very little drug molecules accumulated in solid tumours, 3) the therapeutic efficacy of the ADC was reduced with low levels of antigen expression on tumour cells, 4) the molecules had short half-lives due to serum instability and most pertinently 5) there were challenges in generating homogenised ADC populations suitable for clinical application [177].

The mainstream conjugation strategy employed to load ADCs with toxic payloads was through reduced disulfide residues or lysine side chains giving rise of heterogenous ADC species with drug to antibody ratios (DAR) ranging anywhere from zero to eight [100, 178, 179]. This resulted in varying

pharmacokinetics and safety profiles which were difficult to fine tune [100, 178, 179]. Since then, major research focus has gone into understanding how loading drugs on antibodies affected the therapeutic potential and stability of the antibody [180-182]. In majority of cases, ADC species with entirely drug-loaded hinges (>DAR6) have been linked with decreased stability in stress conditions resulting in the formation of aggregates [182, 183]. This inclination for aggregation occurs from protein interactions, agitation, ionic strength and temperature shifts [183, 184]. Aggregation affects a drugs pharmacokinetics and shelf life and thus represents a major failing in the ADC production strategy [185]. To prevent the formation of aggregates, DAR reduction from 8 to 2 was implemented resulting a 2-fold increase in the ADCs therapeutic index (maximum tolerated dose/curative dose) but at the cost of decreasing conjugation yield and amplified heterogeneity [180]. Consequently, there is a need for improved ADC design that would allow for the generation of homogenized ADC species while retaining the required traits of an ADC including: structural integrity of the antibody moiety needed for optimal binding to target antigens, serum stability, effective therapeutic index, and cleavability within cells necessary for the release of toxic payloads [183, 186]. To this end, various site-specific conjugation methods have been developed through the introduction of unnatural amino acids into antibodies enabling for controlled chemical conjugation of cytotoxic molecules. However, self-labelling tags like Halo-tag, Clip-tag and SNAP-tag genetically fused to antibodies have demonstrated superiority over chemical conjugation in generating homogenous ADC mixtures and are fast gaining popularity as an efficient one-step conjugation method [100, 187]. SNAP-tag stands out, as its human origin allows for human application without the immunogenicity driven limitations faced with Halo-tag and Clip-tag.

Thus, SNAP-tag technology was chosen to generate a unique SNAP-tag based antibody fusion protein format that could possibly be utilized in an ADC application. SNAP-tag, is an engineered mutant of the human O(6)-alkylguanine-DNA alkyltransferase enzyme. It forms a highly specific, irreversible covalent bond with benzylguanine (BG) derivatives in a 1:1 stoichiometry giving rise to stable, homogenised conjugated molecules [188-190]. The rapidness, simplicity and predictable stoichiometry of the conjugation reaction makes SNAP-tag a superior application in ADC design [100, 190]. Furthermore, the applicability of SNAP-tag has already been authenticated in the generation of scFv-Snap based fusion proteins with specificity for tumour-associated antigens [100, 131]. Fusion proteins could be generated using any antibody-based format, but we chose the scFv for its ease of use in early-stage proof of concept applications. This class of therapeutic fusion proteins demonstrates several cutting-edge traits: 1) they can be expressed in different expression systems (mammalian, bacterial, yeast) with high yield, 2) they do not interfere with the binding capacity of the conjugated antibody, 3) they are small in size allowing for improved tumour penetration and rapid renal clearance and 4) they are capable of delivering a range of effector molecules to target cells, these include; cytotoxic drugs and fluorescent dyes [131, 191].

In summary the aforementioned claims support the specified aims and objectives of this thesis. The novelty of this study lies in the genetic fusion of SNAP-tag to 9 different scFvs (targeting CD3, CD4, CD8, CD19, CD20, PDL1, B7.1, BCK1 and ZIKA-DIII-E) for the development of a differential diagnostic tool while simultaneously offering an inventive conjugation method for the targeted delivery of therapeutic agents to diseased cells.

4.4 Expression and first step purification of recombinant SNAP-tag based fusion proteins

For this study, 9 SNAP-tag based fusion proteins; α CD3(scFv)-snap, α CD4(scFv)-snap, α CD8(scFv)-snap, α CD19(scFv)-snap, α CD20(scFv)-snap, α PDL1(scFv)-snap, α B7.1(scFv)-snap, α BCK1(scFv)-snap and α Zika(scFv)-DIII-snap were successfully cloned. However, due various unforeseen challenges only α CD4(scFv)-snap, α CD20(scFv)-snap, α BCK1(scFv)-snap and α Zika(scFv)-DIII-snap were successfully expressed in HEK293T mammalian cell culture in sufficient quantities for downstream analysis. In terms of cloning, the first challenge was vector re-ligation, which is noticeable in the vector only control plate. Due to this several colonies had to be selected for restriction screening to increase the chances of finding a clone with the correct insert. Vector re-ligation can be caused by incomplete digestion or by the presence of 5'-phosphate groups from the vector. Thus, as a preventative measure, digested vectors should be treated with phosphatase before ligation. Despite the high levels of vector re-ligation we were able to select for colonies with the appropriate insert.

The more significant obstacle in this study was protein expression in mammalian cell cultures. In this study HEK293T cells were chosen for transient protein expression of all the cloned SNAP-tag based constructs. HEK293T cells are one of two of the most widely used mammalian expression cell lines used in research and industry, mainly due to its capacity for post-translational modification and ability to assemble proteins in a human like molecular structure. According to literature this cell line has several more advantageous characteristics including; a rapid replication rate, can tolerate low serum levels in media, and is easily transformed [192]. In contradiction to this publication and others, we found that the main problem with our protein expression was low transfection efficiencies. There are various methods of transfection which may be broadly categorised into 3 groups, chemical, biological, and physical [193]. Chemical methods, like the lipid-based method employed in this study all work on the same basic principle. Positively charged chemicals are complexed with negatively charged nucleic acids. The overall chemical/nucleic acid complex has a resulting positive charge which are attracted to the negatively charged cell membranes. While the exact method of cell entry is unknown, it is speculated that endocytosis and phagocytosis maybe involved. Various factors influence this process, such as the ratio of nucleic acid:chemical, the pH of the reaction solution and the conditions of the cell membrane. As such, chemical methods commonly result in low transfection efficiencies compared to biological and physical methods [193]. As the transfection efficiencies in this research were far lower than required (<10% instead of >70%) it would be worthwhile assessing other transfection methods as an alternative. Biological methods for example have been shown to have high transfection efficiencies and easily achieved sustainable transgene expression. This is largely because biological transfection methods utilize viruses (eg: retrovirus murine leukemia virus) which are able to integrate into the host genome and replicate when the host cell replicates [193]. Consequently, every time a transfected host cell replicates, it also replicated the viral genome resulting in rapid transgene expression. Despite its high transfection efficiencies, virus-mediated transfection methods have many drawbacks such as immunogenicity, cytotoxicity, high tendency for insertional mutation and limited package space for foreign genes. These drawbacks were some of the reasons for choosing a lipid-based chemical transfection method in this study. The lipid-based method has low cytotoxicity, no mutagenesis, and no package size limit. During the time of this research no physical methods of transfection was available due to a lack of necessary available equipment. However, a significant improvement in transfection efficiencies, without the drawbacks associated with viral transfections

maybe achieved through electroporation. Electroporation is a physical method of transfection which is thought to work by emitting a short electrical pulse capable of temporarily disturbing cell membranes allowing for the formation of holes through which nucleic acid may enter the cell [193]. This method is easy and rapid and may transfect large quantities of cells in a few seconds once optimal conditions for electroporation are determined.

The low transfection efficiencies faced in this study prompted prolonged zeocin application for enrichment of eGFP positive cells. This prolonged antibiotic selection pressure caused a change in the morphology of the expression cells, changing them from a healthy spindle like structure to a rounded unhealthy appearance. In addition, these transient expression cells have a limited number of passages, but low transfection efficiencies require longer cell culturing time which inevitably requires more passaging. These unfavourable culture conditions resulted in the expression of very minimal amounts of proteins being expressed. Consequently, very large volumes of cell culture supernatant was harvested for purification, in the hopes that sufficient protein enrichment could be achieved during IMAC purification. All recombinant proteins expressed in this study contained an N-terminal polyhistidine tag (HIS x 6) fused to the peptide sequence for the purposes of purification. The small size of this tag fortifies its utility, enabling effective purification without interfering with the folding, structure, and function of the expressed protein [194, 195]. Pre-concentrated SDS-PAGE analysis of recovered fusion proteins showed no visible bands of interest. There were however visible high-molecular weight contaminating bands suggesting that IMAC purification alone is insufficient to achieve high purity proteins. The implementation of a second purification step such as size exclusion chromatography may circumvent such problems in the future. Since no bands of interest were seen on the pre-concentrated SDS-PAGE gels, an amicon protein concentration was done, combining eluted fractions based on their elution peak generated during IMAC purification. The post-concentration samples were then re-run on an SDS-PAGE gel and revealed the presence of a pCB- α CD4-snap, pCB- α CD20-snap and pCB- α BCK-1-snap protein bands at the correct size suggesting that sufficient protein concentration had occurred for visualization.

In summary, despite the lack of industrial production and purification systems this research demonstrated the successful development of SNAP-tag based fusion proteins. While the generation of SNAP-tag based proteins has been demonstrated before this is the first instance specifically showing successful production of snap fusion proteins targeting CD4, CD20 and BCK1 [146, 196]. This early-stage proof-of-function provides a feasible university level method, albeit with lots of room for improvement, for the production of innovative research and lays the foundations for the development of industrial biopharmaceuticals.

4.5 SNAP-tag based immunodiagnostics and therapeutics for classification and treatment of TNBC.

It is a well-established fact that early detection is a key component of successful management of breast cancer [197, 198]. The aggressive nature of TNBC further exasperates the need for early, accurate, diagnosis to ensure that patients receive the appropriate treatment as soon as possible [199]. The diagnosis of TNBCs in clinical settings traditionally involves a two-step process, combining morphological imaging and immunohistochemistry (IHC) [153]. However, owing to their lack of cancerous mammographic characteristics, such as the presences of an irregularly shaped mass, spiculated tumour margins and related calcification, up to 18% of TNBC tumours go undetected on their initial mammogram [200] emphasising the sub-standard diagnostic capacity of a mammogram alone. Ultrasound represents a diagnostic alternative with higher sensitivity for detection of smooth boards on some TNBC tumours but similarly to mammograms they still fall short at effectively imagining intratumorally characteristic like fibrosis and necrosis [199, 201].

Moreover, gene-expression profiling which has improved our understanding of breast cancers and its molecular subtypes still offers no practical benefit in clinical applications [202]. This lack of a high-throughput routine molecular diagnostic tools prodded the use of breast cancer molecular makers as a substitute for classification of TNBC tumours based on their prognosis and sensitivity to specific treatments [158]. The heterogenous clinical behaviour of TNBC's undoubtably requires a combination of various targets to achieve the required degree of stratification needed to provide meaningful improvement in patient outcomes. Consequently, the development of SNAP-tag based fusion proteins established in this research has the potential to contribute to the development of a differential diagnostic panel for TNBC patient stratification.

Most of the constructs chosen for development within this research was for the specific purpose of characterizing TNBC tumours based on tumour infiltrating lymphocytes (TILs). TILs together with cancer cells and tumour vasculature make up a complex tumour microenvironment (TME) and there is a growing body of evidence indicating that disease progression is dependent on the various interactions between cells of the TME [203]. Increased frequencies of TILs are frequently associated with better prognosis and sensitivity to immunotherapies however low levels of TILs does not necessarily equate to disease progression. Moreover, in metastatic TNBC, the application of anti-PD1 checkpoint therapy in combination with chemotherapy have demonstrated only modest response rates. One reason for this could be that the role of PD1 in effector cells, regulatory cells and memory cells differ from each other and are multifactorial. Certainly, PD1 can lead to Treg overaction resulting in immune suppression rather than immune activation as evident in hyper progressive disease statue in some melanoma patients undergoing PD1 therapy (immune checkpoint signalling and cancer immunotherapy).

Thus, there is urgent need to better understand why patients have ineffective and unintended response to such therapies providing a biological rational to further characterize the TME. To this end, various high-end technologies have been utilized to elucidate the inner workings of TME, these include flow cytometry, single cell mass cytometry and single cell RNA sequencing among others. The major drawback of these technologies is that they do not capture the spatial relationships between cells within the TME.

The application of multiplex immunohistochemistry has the capacity to overcome this and in combination with SNAP-tag technology provide a rapid one step diagnostic tool for TME characterization [204]. As SNAP-tag has the inherent ability to conjugate to any BG-modified molecule, various recombinant proteins targeting different antigens of interest could be conjugated to different coloured BG-modified fluorophores and applied to tumour biopsies simultaneously for imaging analysis. Multiplex imaging with a panel of targets would provide a far more dynamic understanding of cell-to-cell interactions within the TME by allowing for visualization in a 3-dimensional space. This information would be greatly beneficial in informing treatment strategies. For example, if PD1 expression is co-localized with CD8+ cells, it would suggest that anti-PD1 therapy would increase CD8+ T-cell responses and hence allow for better tumour clearance. On the other hand, if PD1 expression is co-localized with a T-reg cell then application of PD1 therapy would likely block of inhibitory signals currently suppressing T-reg cells thus resulting in more suppression due to increased T-reg activity. Within this study we were able to show the specific conjugation of BG-modified Alexa fluoro488 with two SNAP-tag based fusion proteins, pCB- α CD4-snap and pCB- α CD20-snap. This data, together with other studies at the MB&I research unit provide supportive evidence for the usage of SNAP-tag recombinant proteins in combination with IHC as a diagnostic and characterization tool for various tumours.

In addition to diagnosis, conjugation of recombinant SNAP-tag fusion protein with a cytotoxic drug allows for the generation of SNAP-tag based antibody-drug conjugates (ADC) capable of killing target specific cells [151]. Of particular interest is the pCB- α BCK1-snap protein. As previously mentioned BCK1 was found to be expressed on the breast carcinoma cell line 8701BC but not on 184A1 healthy mammary epithelial cells. This differential expression on a breast cancer cell makes BCK1 a very promising target. In this thesis we were able to successfully clone the BCK1 targeting scFv into the pCB vector backbone and express the pCB- α BCK1-Snap fusion protein in HEK293T mammalian expression cells. Western blot analysis further demonstrated functionality of the SNAP-tag domain within the construct suggesting that it is possible to conjugate pCB- α BCK1-snap to any BG-modified molecule including a cytotoxic drug.

While we were unable to obtain enough pCB-BCK1-snap protein for conjugation studies, there are several other studies demonstrating conjugation of SNAP-tag fusion proteins to various BG-modified payloads such as single auristatin derivatives (AURIF) [205]. AURIF has high hydrophobicity, low membrane permeability and potent cytotoxic capacity making it a very suitable option for ADCs. AURIF functionality is driven by its ability to inhibit tubulin dependent GTP hydrolysis and microtubule assembly, culminating in cell cycle arrest and eventually apoptosis [206]. In addition to the payload choice, several other factors contribute to the success or failure of ADC application in a clinical setting, these include but are not limited to; the antibody binding affinity, levels of antigen expression and the rate of internalization and trafficking [207, 208].

For the specific treatment of solid tumours, the ADC must further be able to travel through the blood stream and penetrate the tumour mass. Of note, unstable linkage of the payload to the antibody can result in premature release of the payload causing of target effects and reduced tumour killing. In a study by Woitok et al (2016) a scFv-snap-AURIF demonstrated activity after incubation in human serum for 48 hours, exemplifying the stability of a scFv-snap-AURIF ADCs targeting various antigens [191]. Furthermore, the efficient conjugation of SNAP-tag to BG-AURIF in a 1:1 stoichiometry

circumvents the need for integration of unnatural amino acids in the antibody binding domain allowing for full bind efficiency and more stable, controlled pharmacological characterises.

Taken together, the studies discussed in this section and the research within this thesis provide important first steps for the institution of methodologies tailored toward generation of effective ADCs.

4.6 Zika Virus : a proxy for the use of SNAP-tag technology in Infectious Diseases.

Over the past two decades, advances in cancer biology and pathogenesis have paved the way for development of novel immunotherapeutic strategies for the treatment of various malignancies. As such, majority of immunotherapies developed to date are almost exclusively associated with cancer therapy. However, there are various other diseases which could derive clinical benefit from the advances made in cancer immunotherapy. Of note infectious diseases, have remained a leading cause of morbidity and mortality worldwide, emphasizing the need for development of novel therapies. There are various similarities between cancer pathogenesis and infectious diseases, namely, pathogens must fashion a hospitable environment within the host for survival, they alter host metabolic function to meet their nutritional requirements, and they employ various mechanisms of host immune suppression, frequently via alteration of host regulatory mechanisms. These parallels, inform the rational development of immunotherapies for infectious diseases derived from information obtained from cancer immunotherapy. On this basis, we aimed to generate a zika virus specific SNAP-tag fusion protein using the same tools and technologies employed for the generation of TNBC targeted therapies.

Zika virus is classified as an emerging infectious disease that belongs to a group of mosquito-borne flaviviruses which include yellow fever, West Nile virus and Dengue virus. Zika virus infections were found to have a wide distribution in human populations throughout South-East Asia, India and Africa [209]. However, until major outbreaks in Brazil and other South American countries, Zika virus was considered an obscure tropical pathogen. Zika virus poses a significant arising threat to human health due to it's association with neurological complication such as Guillain-Barre syndrome and fetal development defects [210]. Birth defects found in infants with zika virus infected mums are collectively defined as congenital Zika syndrome (CZS) and encompass various symptoms including, microcephaly, complications with hearing and vision, seizures, and motor disabilities [210] In Rio de Janeiro, foetal abnormalities are seen in 42% of foetuses of zika virus infected women [210]. The rapid spread of Zika virus resulting in its global presence in tropical regions together with the high human cost associated with infection led the WHO to declare Zika a public health emergency in 2016 [209].

The virus persists through enzootic cycles between the carrier mosquito (*Aedes* genus) and primate hosts, which include humans. Unlike other flaviviruses, Zika has the unique capacity to persevere for months in immune-privileged sites like the eyes and testis, enabling for sexual transmission [209]. Within mammalian hosts, viral infection occurs through sequential engagement of various cell types. The initial viral inoculum originating from a mosquito bite, infects skin cells like fibroblasts and keratocytes which are in close proximity [209]. Sooner or later, the virus is picked up by resident Langerhans cells which transport the virus to draining lymph nodes. Viremia is subsequently established through the infection of lymph node infiltrating monocytes and macrophages. Circulating

infected monocytes and macrophages leaving the lymph nodes, carry the virus to other sites including those associated with Zika pathologies such as the testes and placenta [209]. Furthermore, circulating virus infected cells are able to infect new mosquitoes during blood feeding continuing the cycle of infection.

Despite the server health risks associated with Zika virus infections there is still no approved therapeutic option for treatment of the disease. All current treatment options are palliative and do not prevent CZS. However, targeted therapeutics has the potential to combat many of the clinical complications associated with Zika virus infections. Various animal models demonstrate the effective prophylactic and therapeutic outcomes of mAb based therapy in flavivirus infections. Sapparapu et al, studied the therapeutic capacity of a collection of human mAb isolated from patients who had recovered from zika virus infections. Many of the antibodies had specific binding affinity for different epitopes on the envelop (E) protein and demonstrated potent neutralizing activity. One of the antibodies exhibiting the highest neutralizing capacity showed marked reduction in tissue pathology, mortality and placental and foetal infections when administered in zika infected mice [211]. Zhao H et al further demonstrated the protective effective of mAb therapy in mice infected with the Asian, African or American strain of zika virus [211]. In another study, mice were administered with single does DNA vaccines containing zika pre-membrane and envelop proteins. Adoptive transfer of the resulting IgG antibodies conferred passive protection from zika infection[211]. These studies as well as others [211] support the use of zikavirus envelop protein as a target for the development of immunotherapeutics.

In a follow up study conducted by Zhao et al, the authors developed a panel of zika-virus specific mAb's that could neutralize the Asian, American and African strains of zika virus. Assessment of these antibodies revealed that 2 mAb (ZV-54 and ZV-67) targeting the lateral ridge of the DIII domain of the envelop (E) protein could confer in vivo protection in animal models [136]. Similar observations were seen in other flavivirus studies, where protection against lethal infections by DENV, WNV were conferred by mAb targeting the lateral domain of DIII E protein. Within this thesis, the scFv sequence derived from the ZV-67 antibody sequence was used for the construction of a recombinant SNAP-tag fusion protein specifically targeting the DIII domain of zika virus E protein. The pCB- α ZIKA DIII-snap protein demonstrated functionality of the SNAP-tag domain by successful conjugation to BG-alexa flouro488. Furthermore, functional binding of the protein was demonstrated through a dose dependent ELISA response on plates coated with ZIKA virus membrane components. The dose dependant binding supports the use of pCB-ZIKA DII-snap for targeted therapy of zika virus infection.

Combining the ZIKA-DII scFv to SNAP-tag, confers dual functionality to the protein. Conjugation with fluorophores will enable microscopic visualization of zika virus infected cells and may assist in elucidating details of the viral infection process, which to date is poorly understood. Conjugation of the pCB-ZIKA-DIII-snap protein to a cytotoxic drug, like AURIF would likely allow for specific killing of zika virus infected cells. Given that widespread killing of monocytes and macrophages would likely result in sever immune compromise within a patient, a targeted therapeutic approach killing only viral infected cells provides a promising alternative.

In summary this research demonstrates the versatility and utility of SNAP-tag technology in infectious disease applications. Although this research is only an early phage proof of concept it provides a basis for the development of SNAP-tag based zika virus biotherapeutics.

4.7 Conclusions and Future Developments.

Vast progress has been made in cancer genomic research over the past few decades, highlighting the belief that numerous genetic alterations are involved in encouraging tumorigenesis [212]. The resilience, plasticity and complexity of cancer cells is additional intensified by the presence of stem-cell like tumour cell subpopulations that largely contribute to treatment resistance and relapse [213, 214]. Precision medicine approaches account for the molecular and biological disparities among patients, classifying individuals based on their susceptibility to specific diseases and their probably sensitivity to particular treatments [215-217]. As such, development of individualized therapies for TNBC, requires an exhaustive understanding of oncolytic pathways and microenvironment alterations and the effects of the immune system and appropriate therapies on these pathways [75]. Current mainstream approaches rely heavily on conventional clinical and pathological features for diagnosis and treatment, making personalizing therapies a mammoth task. Furthermore, logistical and scientific challenges such as ; low expression of target antigens, affordability and practicability of high-throughput machineries in clinical settings, additional expenses of companion treatments and lack of specific medications targeting mutations have all become more apparent [218].

This thesis reports on the development of a methodology for the generation of SNAP-tag based fusion protein with potential diagnostic and therapeutic application in TNBC patients. In brief the methodology relies on identification or isolation of antibody sequences either through patent searching or phage display. Utilizing phage display we were able to specifically isolate Ki-4 targeting antibodies showing that the phage propagation and screening protocol can be successfully utilized for the purposes of antibody isolation. We were further successful in identifying and isolating 9 different antibody sequences from patents, all of which were relevant to the classification of TNBC tumours showing the relevance of patents as a source of sequencing information. Once extracted the sequences were analysed with IgBlast and the heavy chain and light chain sequences were used to design various scFv's with the same specific binding as the originally identified antibody sequences. The finalized scFv sequence was then synthesized and cloned into the pCB vector, which was specifically designed for the generation of SNAP-tag fusion proteins. All 9 constructs of interest which included scFv's targeting CD3, CD4, CD8, CD19, CD20, PDL1, B7.1, and BCK1 were successfully cloned into the SNAP-tag containing pCB vector. This was followed by transfection of the resulting plasmids into HEK293T mammalian expression cells. Undoubtedly, the major pitfall of this study was the transfection of HEK293T cells and protein expression, which would need significant improvement and optimization for future studies aiming at generating SNAP-tag based fusion proteins. Despite this however, we were able to harvest and purify sufficient α CD4-snap, α CD20-snap and α BCK1-snap fusion proteins to demonstrated functionality of the snap domain through successful conjugation to BG-modified Alexafluoro488 and visualization on SDS-PAGE and western blots.

In summary, this research showcases the development of novel SNAP-tag based human fusion proteins with the high translational potential for targeted diagnosis and killing of TNBC cells. By harnessing the beneficial aspects of SNAP-tag together with other antibody manipulation tools, this study provides a working methodology for the economical production of homogenous recombinant fusion proteins. Application of this method in infectious diseases resulted in the successful generation of a fully functional zika virus specific recombinant SNAP-tag fusion protein, further substantiating the utility of this methodology. The efficient conjugation of Zika(scFv)-Snap to BG-Alexafluoro488,

followed by visualization of the correct sized fluorescent band on an SDS-PAGE gel was used to confirm functionality of the snap domain. The specific binding of the scFv within the Zika(scFv)-snap protein was confirmed on Zika antigen coated ELISA plates. On a wider scale this method provides a blueprint for the development of a variety of companion immunodiagnostic tools and therapies providing a necessary motivation for a shift toward personalized medicine. While there is still a long way to go before clinical implementation of personalized medicines, the findings of thesis emphasize the versatility of SNAP-tag technology in the generation of pharmaceutically relevant diagnostic and therapeutic agents needed for implementation of individualized treatment strategies.

,

References

1. Abdulrahman, G.O. and G.A. Rahman, *Epidemiology of breast cancer in Europe and Africa*. Journal of cancer epidemiology, 2012. **2012**.
2. Tran, B. and M. Rosenthal, *Survival comparison between glioblastoma multiforme and other incurable cancers*. Journal of Clinical Neuroscience, 2010. **17**(4): p. 417-421.
3. Tian, T., et al., *The origins of cancer robustness and evolvability*. Integrative Biology, 2011. **3**(1): p. 17-30.
4. Herbst, C., *Fact sheet on breast cancer in women*. Cancer Association of South Africa, 2017.
5. Turkoz, F.P., et al., *Association between common risk factors and molecular subtypes in breast cancer patients*. The breast, 2013. **22**(3): p. 344-350.
6. Prat, A., et al., *Clinical implications of the intrinsic molecular subtypes of breast cancer*. The Breast, 2015. **24**: p. S26-S35.
7. Prat, A., et al., *Prognostic significance of progesterone receptor–positive tumor cells within immunohistochemically defined luminal A breast cancer*. Journal of clinical oncology, 2013. **31**(2): p. 203.
8. Prat, A., et al., *Molecular features of the basal-like breast cancer subtype based on BRCA1 mutation status*. Breast cancer research and treatment, 2014. **147**(1): p. 185-191.
9. Podo, F., et al., *Triple-negative breast cancer: present challenges and new perspectives*. Molecular oncology, 2010. **4**(3): p. 209-229.
10. Lee, A. and M.B. Djamgoz, *Triple negative breast cancer: emerging therapeutic modalities and novel combination therapies*. Cancer treatment reviews, 2018. **62**: p. 110-122.
11. Rapiti, E., et al., *Opportunities for improving triple-negative breast cancer outcomes: results of a population-based study*. Cancer medicine, 2017. **6**(3): p. 526-536.
12. Cleator, S., W. Heller, and R.C. Coombes, *Triple-negative breast cancer: therapeutic options*. The lancet oncology, 2007. **8**(3): p. 235-244.
13. Garrido-Castro, A.C., N.U. Lin, and K. Polyak, *Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment*. Cancer discovery, 2019. **9**(2): p. 176-198.
14. Aaltomaa, S., et al., *Lymphocyte infiltrates as a prognostic variable in female breast cancer*. European journal of cancer, 1992. **28**(4-5): p. 859-864.
15. Matsumoto, H., et al., *Role of inflammatory infiltrates in triple negative breast cancer*. Journal of clinical pathology, 2015. **68**(7): p. 506-510.
16. Fridman, W.H., et al., *Prognostic and predictive impact of intra-and peritumoral immune infiltrates*. Cancer research, 2011. **71**(17): p. 5601-5605.
17. Gruosso, T., et al., *Distinct immune microenvironments stratify triple-negative breast cancer and predict outcome*. Annals of Oncology, 2017. **28**: p. i16.
18. Karn, T., et al., *Abstract S1-07: Immune sculpting of the triple negative breast cancer genome*. Cancer Research, 2017. **77**(4_Supplement): p. S1-07-S1-07.
19. Li, Z., et al., *Immunotherapeutic interventions of triple negative breast cancer*. Journal of translational medicine, 2018. **16**(1): p. 1-19.
20. Ashley, E.A., *Towards precision medicine*. Nature Reviews Genetics, 2016. **17**(9): p. 507-522.
21. Ahr, A., et al., *Molecular classification of breast cancer patients by gene expression profiling*. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 2001. **195**(3): p. 312-320.
22. Orentas, R.J., et al., *Identification of cell surface proteins as potential immunotherapy targets in 12 pediatric cancers*. Frontiers in oncology, 2012. **2**: p. 194.
23. Rothe, A., et al., *Construction of phage display libraries from reactive lymph nodes of breast carcinoma patients and selection for specifically binding human single chain Fv on cell lines*. International journal of molecular medicine, 2004. **14**(4): p. 729-764.

24. Varadé, J., S. Magadán, and Á. González-Fernández, *Human immunology and immunotherapy: main achievements and challenges*. Cellular & Molecular Immunology, 2021. **18**(4): p. 805-828.
25. Ramamurthy, D., et al., *Recent advances in immunotherapies against infectious diseases*. Immunotherapy Advances, 2021. **1**(1): p. Itaa007.
26. Reichert, J.M., et al., *Monoclonal antibody successes in the clinic*. Nature biotechnology, 2005. **23**(9): p. 1073-1078.
27. Carter, P.J., *Potent antibody therapeutics by design*. Nature reviews immunology, 2006. **6**(5): p. 343-357.
28. Scott, C.T., *The problem with potency*. Nature biotechnology, 2005. **23**(9): p. 1037-1040.
29. Workman, P., et al., *How much longer will we put up with \$100,000 cancer drugs?* Cell, 2017. **168**(4): p. 579-583.
30. Hoogenboom, H.R., *Selecting and screening recombinant antibody libraries*. Nature biotechnology, 2005. **23**(9): p. 1105-1116.
31. Chiu, M.L., et al., *Antibody structure and function: the basis for engineering therapeutics*. Antibodies, 2019. **8**(4): p. 55.
32. Fernandes, J.C., *Therapeutic application of antibody fragments in autoimmune diseases: current state and prospects*. Drug discovery today, 2018. **23**(12): p. 1996-2002.
33. Manoutcharian, K., R. Perez-Garmendia, and G. Gevorkian, *Recombinant antibody fragments for neurodegenerative diseases*. Current neuropharmacology, 2017. **15**(5): p. 779-788.
34. Brinkmann, U., *Disulfide-Stabilized Fv Fragments*, in *Antibody Engineering*, R. Kontermann and S. Dübel, Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 181-189.
35. Dobosz, P. and T. Dzieciatkowski, *The intriguing history of cancer immunotherapy*. Frontiers in immunology, 2019. **10**: p. 2965.
36. Weiner, L.M., J.C. Murray, and C.W. Shuptrine, *Antibody-based immunotherapy of cancer*. Cell, 2012. **148**(6): p. 1081-1084.
37. Naran, K., et al., *Principles of immunotherapy: implications for treatment strategies in cancer and infectious diseases*. Frontiers in microbiology, 2018. **9**: p. 3158.
38. Allen, B.J., *Can α -radioimmunotherapy increase efficacy for the systemic control of cancer?* Immunotherapy, 2011. **3**(4): p. 455-458.
39. Trail, P.A., G.M. Dubowchik, and T.B. Lowinger, *Antibody drug conjugates for treatment of breast cancer: novel targets and diverse approaches in ADC design*. Pharmacology & therapeutics, 2018. **181**: p. 126-142.
40. Herrera, A.F., et al., *Interim results of brentuximab vedotin in combination with nivolumab in patients with relapsed or refractory Hodgkin lymphoma*. Blood, The Journal of the American Society of Hematology, 2018. **131**(11): p. 1183-1194.
41. Lehar, S.M., et al., *Novel antibody-antibiotic conjugate eliminates intracellular *S. aureus**. Nature, 2015. **527**(7578): p. 323-328.
42. Sharpe, A.H. and K.E. Pauken, *The diverse functions of the PD1 inhibitory pathway*. Nature Reviews Immunology, 2018. **18**(3): p. 153-167.
43. Brunet, J.-F., et al., *A new member of the immunoglobulin superfamily—CTLA-4*. Nature, 1987. **328**(6127): p. 267-270.
44. Dariavach, P., et al., *Human Ig superfamily CTLA-4 gene: chromosomal localization and identity of protein sequence between murine and human CTLA-4 cytoplasmic domains*. European journal of immunology, 1988. **18**(12): p. 1901-1905.
45. Walunas, T.L., et al., *CTLA-4 can function as a negative regulator of T cell activation*. Immunity, 1994. **1**(5): p. 405-413.
46. Lo, B., et al., *Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy*. Science, 2015. **349**(6246): p. 436-440.
47. Linsley, P.S., et al., *CTLA-4 is a second receptor for the B cell activation antigen B7*. The Journal of experimental medicine, 1991. **174**(3): p. 561-569.

48. Pentcheva-Hoang, T., et al., *B7-1 and B7-2 selectively recruit CTLA-4 and CD28 to the immunological synapse*. Immunity, 2004. **21**(3): p. 401-413.
49. Greenwald, R.J., et al., *CTLA-4 regulates induction of anergy in vivo*. Immunity, 2001. **14**(2): p. 145-155.
50. Hou, T.Z., et al., *A transendocytosis model of CTLA-4 function predicts its suppressive behavior on regulatory T cells*. The Journal of Immunology, 2015. **194**(5): p. 2148-2159.
51. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice*. Nature Reviews Immunology, 2020. **20**(11): p. 651-668.
52. Sun, C., R. Mezzadra, and T.N. Schumacher, *Regulation and function of the PD-L1 checkpoint*. Immunity, 2018. **48**(3): p. 434-452.
53. Li, C.-W., et al., *Eradication of triple-negative breast cancer cells by targeting glycosylated PD-L1*. Cancer cell, 2018. **33**(2): p. 187-201. e10.
54. Foxp, G.H., *tregs and B7-H1+/PD-1+ T lymphocytes co-infiltrate the tumor tissues of high-risk breast cancer patients: implication for immunotherapy*. BMC Cancer, 2008. **8**: p. 57.
55. Li, Z., et al., *PD-L1 expression is associated with tumor FOXP3+ regulatory T-cell infiltration of breast cancer and poor prognosis of patient*. Journal of Cancer, 2016. **7**(7): p. 784.
56. Aguiar Jr, P.N., et al., *The role of PD-L1 expression as a predictive biomarker in advanced non-small-cell lung cancer: a network meta-analysis*. Immunotherapy, 2016. **8**(4): p. 479-488.
57. Hartman, Z., et al. *CTLA-4 and PD-1 checkpoint inhibitors enhance individually tailored adaptive antitumor immune responses to overcome tumor immunosuppression and effectively treat triple-negative breast cancer*. in *CANCER RESEARCH*. 2017. AMER ASSOC CANCER RESEARCH 615 CHESTNUT ST, 17TH FLOOR, PHILADELPHIA, PA
58. Liu, L., et al., *Combination immunotherapy of MUC1 mRNA nano-vaccine and CTLA-4 blockade effectively inhibits growth of triple negative breast cancer*. Molecular Therapy, 2018. **26**(1): p. 45-55.
59. Stanton, S.E., S. Adams, and M.L. Disis, *Variation in the incidence and magnitude of tumor-infiltrating lymphocytes in breast cancer subtypes: a systematic review*. JAMA oncology, 2016. **2**(10): p. 1354-1360.
60. Taylor, N.A., et al., *Treg depletion potentiates checkpoint inhibition in claudin-low breast cancer*. The Journal of clinical investigation, 2017. **127**(9): p. 3472-3483.
61. Brockwell, N.K., et al., *Neoadjuvant interferons: Critical for effective PD-1-based Immunotherapy in TNBC*. Cancer Immunology Research, 2017. **5**(10): p. 871-884.
62. Köhler, G. and C. Milstein, *Continuous cultures of fused cells secreting antibody of predefined specificity*. nature, 1975. **256**(5517): p. 495-497.
63. Ober, R.J., et al., *Differences in promiscuity for antibody–FcRn interactions across species: implications for therapeutic antibodies*. International immunology, 2001. **13**(12): p. 1551-1559.
64. Ayyar, B.V., S. Arora, and R. O’Kennedy, *Coming-of-age of antibodies in cancer therapeutics*. Trends in pharmacological sciences, 2016. **37**(12): p. 1009-1028.
65. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface*. Science, 1985. **228**(4705): p. 1315-1317.
66. O'Brien, P.M. and R. Aitken, *Antibody phage display: methods and protocols*. Vol. 178. 2004: Springer Science & Business Media.
67. Huang, J.X., S.L. Bishop-Hurley, and M.A. Cooper, *Development of anti-infectives using phage display: biological agents against bacteria, viruses, and parasites*. Antimicrobial agents and chemotherapy, 2012. **56**(9): p. 4569-4582.
68. Kay, B.K., J. Winter, and J. McCafferty, *Phage display of peptides and proteins: a laboratory manual*. 1996: Elsevier.
69. McCafferty, J., et al., *Phage antibodies: filamentous phage displaying antibody variable domains*. nature, 1990. **348**(6301): p. 552-554.

70. Zacher III, A.N., et al., *A new filamentous phage cloning vector: fd-tet*. Gene, 1980. **9**(1-2): p. 127-140.
71. Knaul, J.K., et al., *Lung-residing myeloid-derived suppressors display dual functionality in murine pulmonary tuberculosis*. American journal of respiratory and critical care medicine, 2014. **190**(9): p. 1053-1066.
72. Skerra, A. and A. Plückthun, *Assembly of a functional immunoglobulin Fv fragment in Escherichia coli*. Science, 1988. **240**(4855): p. 1038-1041.
73. Clackson, T. and H.R. Hoogenboom, *Making antibody fragments using phage display libraries*. Nature, 1991. **352**(6336): p. 624.
74. Qi, H., et al., *Phagemid vectors for phage display: properties, characteristics and construction*. Journal of molecular biology, 2012. **417**(3): p. 129-143.
75. DU, D.-X. and R. Zhang, *Application and progress of helper phage in phage display*. Microbiology, 1992.
76. Chowdhury, P.S., et al., *Isolation of a high-affinity stable single-chain Fv specific for mesothelin from DNA-immunized mice by phage display and construction of a recombinant immunotoxin with anti-tumor activity*. Proceedings of the National Academy of Sciences, 1998. **95**(2): p. 669-674.
77. Barth, S., et al., *Combining phage display and screening of cDNA expression libraries: a new approach for identifying the target antigen of an scFv preselected by phage display*. Journal of molecular biology, 2000. **301**(4): p. 751-757.
78. Clementi, N., et al., *Phage display-based strategies for cloning and optimization of monoclonal antibodies directed against human pathogens*. International Journal of Molecular Sciences, 2012. **13**(7): p. 8273-8292.
79. Mancini, N., et al., *Phage display for the production of human monoclonal antibodies against human pathogens*. New Microbiologica, 2004. **27**(4): p. 315-328.
80. Williamson, R.A., et al., *Human monoclonal antibodies against a plethora of viral pathogens from single combinatorial libraries*. Proceedings of the National Academy of Sciences, 1993. **90**(9): p. 4141-4145.
81. Burton, D.R., et al., *A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals*. Proceedings of the National Academy of Sciences, 1991. **88**(22): p. 10134-10137.
82. Burioni, R., et al., *Dissection of human humoral immune response against hepatitis C virus E2 glycoprotein by repertoire cloning and generation of recombinant Fab fragments*. Hepatology, 1998. **28**(3): p. 810-814.
83. Zhang, M.-Y., et al., *Identification and characterization of a new cross-reactive human immunodeficiency virus type 1-neutralizing human monoclonal antibody*. Journal of virology, 2004. **78**(17): p. 9233-9242.
84. Mao, S., et al., *Phage-display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl Lewisx and Lewisx*. Proceedings of the National Academy of Sciences, 1999. **96**(12): p. 6953-6958.
85. Reiche, N., et al., *Generation and characterization of human monoclonal scFv antibodies against Helicobacter pylori antigens*. Infection and immunity, 2002. **70**(8): p. 4158-4164.
86. Kim, S.J., et al., *Neutralizing human monoclonal antibodies to hepatitis A virus recovered by phage display*. Virology, 2004. **318**(2): p. 598-607.
87. Chang, T.Y. and D.L. Siegel, *Isolation of an IgG anti-B from a human Fab-phage display library*. Transfusion, 2001. **41**(1): p. 6-12.
88. Roark, J.H., et al., *Genetic analysis of autoantibodies in idiopathic thrombocytopenic purpura reveals evidence of clonal expansion and somatic mutation*. Blood, The Journal of the American Society of Hematology, 2002. **100**(4): p. 1388-1398.
89. Xu, M.-Y., et al., *Production of a human single-chain variable fragment antibody against esophageal carcinoma*. World journal of gastroenterology: WJG, 2004. **10**(18): p. 2619.

90. Dooley, H. and M. Flajnik, *Antibody repertoire development in cartilaginous fish*. Developmental & Comparative Immunology, 2006. **30**(1-2): p. 43-56.
91. De Genst, E., et al., *Antibody repertoire development in camelids*. Developmental & Comparative Immunology, 2006. **30**(1-2): p. 187-198.
92. Sixholo, J., et al., *Improving the characteristics of a mycobacterial 16 kDa-specific chicken scFv*. Biologicals, 2011. **39**(2): p. 110-116.
93. Wang, Y., et al., *Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics*. Nucleic acids research, 2020. **48**(D1): p. D1031-D1041.
94. Li, X., J.A. Van Deventer, and S. Hassoun, *ASAP-SML: An antibody sequence analysis pipeline using statistical testing and machine learning*. PLoS computational biology, 2020. **16**(4): p. e1007779.
95. Krawczyk, K., A. Buchanan, and P. Marcatili. *Data mining patented antibody sequences*. in MAbs. 2021. Taylor & Francis.
96. Ye, J., et al., *IgBLAST: an immunoglobulin variable domain sequence analysis tool*. Nucleic acids research, 2013. **41**(W1): p. W34-W40.
97. Yu, K., et al., *Synthetic fusion protein design and applications*. Biotechnology advances, 2015. **33**(1): p. 155-164.
98. Padayachee, E.R., et al., *Applications of SNAP-tag technology in skin cancer therapy*. Health science reports, 2019. **2**(2): p. e103.
99. Fawzi Hussain, A., M. Amoury, and S. Barth, *SNAP-tag technology: a powerful tool for site specific conjugation of therapeutic and imaging agents*. Current pharmaceutical design, 2013. **19**(30): p. 5437-5442.
100. Hussain, A.F., et al., *One-step site-specific antibody fragment auto-conjugation using SNAP-tag technology*. Nature Protocols, 2019. **14**(11): p. 3101-3125.
101. Graham, B.S. and D.M. Ambrosino, *History of passive antibody administration for prevention and treatment of infectious diseases*. Current Opinion in HIV and AIDS, 2015. **10**(3): p. 129.
102. Parren, P.W. and D.R. Burton, *The antiviral activity of antibodies in vitro and in vivo*. 2001.
103. Hu, J. and J.L. Robinson, *Treatment of respiratory syncytial virus with palivizumab: a systematic review*. World Journal of Pediatrics, 2010. **6**(4): p. 296-300.
104. Da Silva, S., D. Oliveira Silva Martins, and A.C.G. Jardim, *A review of the ongoing research on Zika virus treatment*. Viruses, 2018. **10**(5): p. 255.
105. Dick, G.W., S.F. Kitchen, and A.J. Haddow, *Zika virus (I). Isolations and serological specificity*. Transactions of the royal society of tropical medicine and hygiene, 1952. **46**(5): p. 509-520.
106. Vinet, L. and A. Zhedanov, *A 'missing' family of classical orthogonal polynomials*. Journal of Physics A: Mathematical and Theoretical, 2011. **44**(8): p. 085201.
107. Dick, G.W., S. Kitchen, and A. Haddow, *Zika virus (II). Pathogenicity and physical properties*. Transactions of the royal society of tropical medicine and hygiene, 1952. **46**(5).
108. Simpson, D.I., *Zika virus infection in man*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 1964. **58**(4): p. 335-8.
109. Musso, D., et al., *Inactivation of dengue virus in plasma with amotosalen and ultraviolet A illumination*. Transfusion, 2014. **54**(11): p. 2924-2930.
110. Duffy, M.R., et al., *Zika virus outbreak on Yap Island, federated states of Micronesia*. New England Journal of Medicine, 2009. **360**(24): p. 2536-2543.
111. Haddow, A.D., et al., *Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage*. PLoS neglected tropical diseases, 2012. **6**(2): p. e1477.
112. Olson, J. and T. Ksiazek, *Zika virus, a cause of fever in Central Java, Indonesia*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 1981. **75**(3): p. 389-393.
113. Cao-Lormeau, V.-M., et al., *Zika virus, French polynesia, South pacific, 2013*. Emerging infectious diseases, 2014. **20**(6): p. 1085.

114. Chan, J.F., et al., *Zika fever and congenital Zika syndrome: an unexpected emerging arboviral disease*. Journal of Infection, 2016. **72**(5): p. 507-524.
115. Mwaliko, C., et al., *Zika virus pathogenesis and current therapeutic advances*. Pathogens and Global Health, 2021. **115**(1): p. 21-39.
116. Kuno, G. and G.-J. Chang, *Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses*. Archives of virology, 2007. **152**(4): p. 687-696.
117. Rasmussen, S.A., et al., *Zika virus and birth defects—reviewing the evidence for causality*. New England Journal of Medicine, 2016. **374**(20): p. 1981-1987.
118. Gallichotte, E.N., et al., *Role of Zika virus envelope protein domain III as a target of human neutralizing antibodies*. MBio, 2019. **10**(5): p. e01485-19.
119. Stettler, K., et al., *Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection*. Science, 2016. **353**(6301): p. 823-826.
120. Robbiani, D.F., et al., *Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico*. Cell, 2017. **169**(4): p. 597-609. e11.
121. Rogers, T.F., et al., *Zika virus activates de novo and cross-reactive memory B cell responses in dengue-experienced donors*. Science immunology, 2017. **2**(14): p. eaan6809.
122. Sapparapu, G., et al., *Neutralizing human antibodies prevent Zika virus replication and fetal disease in mice*. Nature, 2016. **540**(7633): p. 443-447.
123. Yu, L., et al., *Delineating antibody recognition against Zika virus during natural infection*. JCI insight, 2017. **2**(12).
124. Liang, H., et al., *Recombinant Zika virus envelope protein elicited protective immunity against Zika virus in immunocompetent mice*. PloS one, 2018. **13**(3): p. e0194860.
125. Klausner, R.D., *The fabric of cancer cell biology—weaving together the strands*. Cancer Cell, 2002. **1**(1): p. 3-10.
126. Mungra, N., et al., *Targeted human cytolytic fusion proteins at the cutting edge: harnessing the apoptosis-inducing properties of human enzymes for the selective elimination of tumor cells*. Oncotarget, 2019. **10**(8): p. 897.
127. Hassan, M., et al., *Chemotherapy for breast cancer*. Oncology reports, 2010. **24**(5): p. 1121-1131.
128. Jin, J., et al., *Phototheranostics of CD44-positive cell populations in triple negative breast cancer*. Scientific reports, 2016. **6**(1): p. 1-12.
129. Liedtke, C., et al., *Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer*. Journal of clinical oncology, 2008. **26**(8): p. 1275-1281.
130. Amoury, M., et al., *Granzyme B-based cytolytic fusion protein targeting EpCAM specifically kills triple negative breast cancer cells in vitro and inhibits tumor growth in a subcutaneous mouse tumor model*. Cancer letters, 2016. **372**(2): p. 201-209.
131. Kampmeier, F., et al., *Site-specific, covalent labeling of recombinant antibody fragments via fusion to an engineered version of 6-O-alkylguanine DNA alkyltransferase*. Bioconjugate chemistry, 2009. **20**(5): p. 1010-1015.
132. Kampmeier, F., et al., *Rapid optical imaging of EGF receptor expression with a single-chain antibody SNAP-tag fusion protein*. European journal of nuclear medicine and molecular imaging, 2010. **37**(10): p. 1926-1934.
133. Recker, T., et al., *Directed covalent immobilization of fluorescently labeled cytokines*. Bioconjugate chemistry, 2011. **22**(6): p. 1210-1220.
134. Pardo, A., et al., *In vivo imaging of immunotoxin treatment using Katushka-transfected A-431 cells in a murine xenograft tumour model*. Cancer Immunology, Immunotherapy, 2012. **61**(10): p. 1617-1626.
135. Rami, A., et al., *An overview on application of phage display technique in immunological studies*. Asian Pacific Journal of Tropical Biomedicine, 2017. **7**(7): p. 599-602.
136. Zhao, H., et al., *Structural basis of Zika virus-specific antibody protection*. Cell, 2016. **166**(4): p. 1016-1027.

137. Liu, J.K., *The history of monoclonal antibody development—progress, remaining challenges and future innovations*. Annals of medicine and surgery, 2014. **3**(4): p. 113-116.
138. Sheehan, J. and W.A. Marasco, *Phage and yeast display*. Microbiology spectrum, 2015. **3**(1): p. 3.1. 11.
139. Tur, M., et al., *Selection of scFv phages on intact cells under low pH conditions leads to a significant loss of insert-free phages*. Biotechniques, 2001. **30**(2): p. 404-413.
140. Hoogenboom, H.R., *Overview of antibody phage-display technology and its applications*. Antibody Phage Display, 2002: p. 1-37.
141. Larsen, S.A., et al., *Selection of a breast cancer subpopulation-specific antibody using phage display on tissue sections*. Immunologic Research, 2015. **62**(3): p. 263-272.
142. Ruan, W., et al., *Identification of Clinically Significant Tumor Antigens by Selecting Phage Antibody Library on Tumor Cells in Situ Using Laser Capture Microdissection** S. Molecular & Cellular Proteomics, 2006. **5**(12): p. 2364-2373.
143. Ten Haaf, A., et al., *Phage display-based on-slide selection of tumor-specific antibodies on formalin-fixed paraffin-embedded human tissue biopsies*. Immunol Lett, 2015. **166**(2): p. 65-78.
144. Fitting, J., et al., *Phage display-based generation of novel internalizing antibody fragments for immunotoxin-based treatment of acute myeloid leukemia*. MAbs, 2015. **7**(2): p. 390-402.
145. Brehm, H., et al., *Targeted killing of rhabdomyosarcoma cells by a MAP-based human cytolytic fusion protein*. Cancer letters, 2015. **365**(2): p. 149-155.
146. Amoury, M., et al., *A novel approach for targeted elimination of CSPG4-positive triple-negative breast cancer cells using a MAP tau-based fusion protein*. International journal of cancer, 2016. **139**(4): p. 916-927.
147. Nicolosi, P.A., A. Dallatomasina, and R. Perris, *Theranostic impact of NG2/CSPG4 proteoglycan in cancer*. Theranostics, 2015. **5**(5): p. 530.
148. Wang, X., et al., *CSPG4 in cancer: multiple roles*. Current molecular medicine, 2010. **10**(4): p. 419-429.
149. Yuan, P., et al., *Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for Clostridium difficile toxin B*. Cell research, 2015. **25**(2): p. 157-168.
150. Wang, X., et al., *CSPG4 protein as a new target for the antibody-based immunotherapy of triple-negative breast cancer*. Journal of the National Cancer Institute, 2010. **102**(19): p. 1496-1512.
151. Woitok, M., et al., *Comparison of a mouse and a novel human scFv-SNAP-auristatin F drug conjugate with potent activity against EGFR-overexpressing human solid tumor cells*. OncoTargets and therapy, 2017. **10**: p. 3313.
152. Liu, H., et al., *Comparative genomics, infectivity and cytopathogenicity of American isolates of Zika virus that developed persistent infections in human embryonic kidney (HEK293) cells*. International Journal of Molecular Sciences, 2019. **20**(12): p. 3035.
153. Anders, C.K. and L.A. Carey, *Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer*. Clinical breast cancer, 2009. **9**: p. S73-S81.
154. Ismail-Khan, R. and M.M. Bui, *A review of triple-negative breast cancer*. Cancer control, 2010. **17**(3): p. 173-176.
155. Mehanna, J., et al., *Triple-negative breast cancer: current perspective on the evolving therapeutic landscape*. International journal of women's health, 2019. **11**: p. 431.
156. Lin, N.U., et al., *Clinicopathologic features, patterns of recurrence, and survival among women with triple-negative breast cancer in the National Comprehensive Cancer Network*. Cancer, 2012. **118**(22): p. 5463-5472.
157. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study*. Jama, 2006. **295**(21): p. 2492-2502.
158. Hudis, C.A. and L. Gianni, *Triple-negative breast cancer: an unmet medical need*. The oncologist, 2011. **16**(S1): p. 1-11.

159. Cuzick, J., et al., *Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial*. The lancet oncology, 2010. **11**(12): p. 1135-1141.
160. Gianni, L., et al., *Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial*. The lancet oncology, 2011. **12**(3): p. 236-244.
161. Guarneri, V., et al., *Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors*. Journal of clinical oncology, 2006. **24**(7): p. 1037-1044.
162. Kaplan, H.G., J.A. Malmgren, and M. Atwood, *T1N0 triple negative breast cancer: risk of recurrence and adjuvant chemotherapy*. The breast journal, 2009. **15**(5): p. 454-460.
163. Verma, S., L. Provencher, and R. Dent, *Emerging trends in the treatment of triple-negative breast cancer in Canada: a survey*. Current Oncology, 2011. **18**(4): p. 180-190.
164. Mann, N.H., *The third age of phage*. PLoS biology, 2005. **3**(5): p. e182.
165. Matsuzaki, S., et al., *Perspective: The age of the phage*. Nature, 2014. **509**(7498): p. S9-S9.
166. Barderas, R. and E. Benito-Peña, *The 2018 Nobel Prize in Chemistry: phage display of peptides and antibodies*. Analytical and bioanalytical chemistry, 2019. **411**(12): p. 2475-2479.
167. Luedke, E., et al., *Monoclonal antibody therapy of pancreatic cancer with cetuximab: potential for immune modulation*. Journal of Immunotherapy (Hagerstown, Md.: 1997), 2012. **35**(5): p. 367.
168. Gisselbrecht, C., et al., *Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era*. Journal of Clinical Oncology, 2010. **28**(27): p. 4184.
169. Senter, P.D., *Potent antibody drug conjugates for cancer therapy*. Current opinion in chemical biology, 2009. **13**(3): p. 235-244.
170. Green, M., J. Murray, and G.N. Hortobagyi, *Monoclonal antibody therapy for solid tumors*. Cancer treatment reviews, 2000. **26**(4): p. 269-286.
171. Scott, A.M., J.D. Wolchok, and L.J. Old, *Antibody therapy of cancer*. Nature reviews cancer, 2012. **12**(4): p. 278-287.
172. Kohrt, H.E., et al., *Combination strategies to enhance antitumor ADCC*. Immunotherapy, 2012. **4**(5): p. 511-527.
173. Mohit, E., A. Hashemi, and M. Allahyari, *Breast cancer immunotherapy: monoclonal antibodies and peptide-based vaccines*. Expert review of clinical immunology, 2014. **10**(7): p. 927-961.
174. Alley, S.C., N.M. Okeley, and P.D. Senter, *Antibody–drug conjugates: targeted drug delivery for cancer*. Current opinion in chemical biology, 2010. **14**(4): p. 529-537.
175. Polakis, P., *Antibody drug conjugates for cancer therapy*. Pharmacological reviews, 2016. **68**(1): p. 3-19.
176. Pietersz, G.A., et al., *Chemoimmunoconjugates for the treatment of cancer*. Advances in immunology, 1994. **56**: p. 301-387.
177. Sievers, E.L. and P.D. Senter, *Antibody-drug conjugates in cancer therapy*. Annual review of medicine, 2013. **64**.
178. Shefet-Carasso, L. and I. Benhar, *Antibody-targeted drugs and drug resistance—challenges and solutions*. Drug Resistance Updates, 2015. **18**: p. 36-46.
179. Beerli, R.R., et al., *Sortase enzyme-mediated generation of site-specifically conjugated antibody drug conjugates with high in vitro and in vivo potency*. PloS one, 2015. **10**(7): p. e0131177.
180. Hamblett, K.J., et al., *Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate*. Clinical cancer research, 2004. **10**(20): p. 7063-7070.
181. Busby, S. and R.H. Ebright, *Transcription activation by catabolite activator protein (CAP)*. Journal of molecular biology, 1999. **293**(2): p. 199-213.

182. Beckley, N.S., et al., *Investigation into temperature-induced aggregation of an antibody drug conjugate*. Bioconjugate chemistry, 2013. **24**(10): p. 1674-1683.
183. Adem, Y.T., et al., *Auristatin antibody drug conjugate physical instability and the role of drug payload*. Bioconjugate chemistry, 2014. **25**(4): p. 656-664.
184. Cromwell, M.E., E. Hilario, and F. Jacobson, *Protein aggregation and bioprocessing*. The AAPS journal, 2006. **8**(3): p. E572-E579.
185. Moussa, E.M., et al., *Immunogenicity of therapeutic protein aggregates*. Journal of pharmaceutical sciences, 2016. **105**(2): p. 417-430.
186. McDonagh, C.F., et al., *Engineered antibody–drug conjugates with defined sites and stoichiometries of drug attachment*. Protein Engineering, Design and Selection, 2006. **19**(7): p. 299-307.
187. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo*. Nature biotechnology, 2003. **21**(1): p. 86-89.
188. Gautier, A., et al., *An engineered protein tag for multiprotein labeling in living cells*. Chemistry & biology, 2008. **15**(2): p. 128-136.
189. Cole, N.B., *Site-Specific Protein Labeling with SNAP-Tags*. Current protocols in protein science, 2013. **73**(1): p. 30.1. 1-30.1. 16.
190. Keppler, A., et al., *Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro*. Methods, 2004. **32**(4): p. 437-444.
191. Woitok, M., et al., *The efficient elimination of solid tumor cells by EGFR-specific and HER2-specific scFv-SNAP fusion proteins conjugated to benzylguanine-modified auristatin F*. Cancer letters, 2016. **381**(2): p. 323-330.
192. Hunter, M., et al., *Optimization of protein expression in mammalian cells*. Current protocols in protein science, 2019. **95**(1): p. e77.
193. Kim, T.K. and J.H. Eberwine, *Mammalian cell transfection: the present and the future*. Analytical and bioanalytical chemistry, 2010. **397**(8): p. 3173-3178.
194. Carson, M., et al., *His-tag impact on structure*. Acta Crystallographica Section D: Biological Crystallography, 2007. **63**(3): p. 295-301.
195. Spriestersbach, A., et al., *Purification of His-tagged proteins*, in *Methods in enzymology*. 2015, Elsevier. p. 1-15.
196. Amoury, M., et al., *Photoimmunotheranostic agents for triple-negative breast cancer diagnosis and therapy that can be activated on demand*. Oncotarget, 2016. **7**(34): p. 54925.
197. Wu, P., et al., *Aptamer-guided silver–gold bimetallic nanostructures with highly active surface-enhanced raman scattering for specific detection and near-infrared photothermal therapy of human breast cancer cells*. Analytical chemistry, 2012. **84**(18): p. 7692-7699.
198. Nover, A.B., et al., *Modern breast cancer detection: a technological review*. International Journal of Biomedical Imaging, 2009. **2009**.
199. Penault-Llorca, F. and G. Viale, *Pathological and molecular diagnosis of triple-negative breast cancer: a clinical perspective*. Annals of oncology, 2012. **23**: p. vi19-vi22.
200. Dogan, B. and L. Turnbull, *Imaging of triple-negative breast cancer*. Annals of Oncology, 2012. **23**: p. vi23-vi29.
201. Uematsu, T., M. Kasami, and S. Yuen, *Triple-negative breast cancer: correlation between MR imaging and pathologic findings*. Radiology, 2009. **250**(3): p. 638-647.
202. Van't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer*. nature, 2002. **415**(6871): p. 530-536.
203. Nederlof, I., et al., *A high-dimensional window into the micro-environment of triple negative breast cancer*. Cancers, 2021. **13**(2): p. 316.
204. Hardcastle, N.S., *Using SNAP tag-based antibody fusion proteins for immunofluorescence imaging of triple-negative breast cancer*. 2022.

205. Mungra, N., *Development of SNAP-tag based fusion proteins as novel auristatin F-containing immunoconjugates and photoimmunotheranostics in the detection and treatment of triple-negative breast cancer*. 2022.
206. Bai, R., G.R. Petit, and E. Hamel, *Dolastatin 10, a powerful cytostatic peptide derived from a marine animal: inhibition of tubulin polymerization mediated through the vinca alkaloid binding domain*. *Biochemical pharmacology*, 1990. **39**(12): p. 1941-1949.
207. Doronina, S.O., et al., *Enhanced activity of monomethylauristatin F through monoclonal antibody delivery: effects of linker technology on efficacy and toxicity*. *Bioconjugate chemistry*, 2006. **17**(1): p. 114-124.
208. Carter, P.J. and P.D. Senter, *Antibody-drug conjugates for cancer therapy*. *The Cancer Journal*, 2008. **14**(3): p. 154-169.
209. Sager, G., et al., *Role of host cell secretory machinery in Zika virus life cycle*. *Viruses*, 2018. **10**(10): p. 559.
210. Magnani, D.M., et al., *Fetal demise and failed antibody therapy during Zika virus infection of pregnant macaques*. *Nature communications*, 2018. **9**(1): p. 1-8.
211. Wahid, B., et al., *Current status of therapeutic and vaccine approaches against Zika virus*. *European journal of internal medicine*, 2017. **44**: p. 12-18.
212. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types*. *Nature*, 2013. **502**(7471): p. 333-339.
213. Stratford, A.L., et al., *Targeting tumour-initiating cells to improve the cure rates for triple-negative breast cancer*. *Expert reviews in molecular medicine*, 2010. **12**.
214. Kai, K., et al., *Breast cancer stem cells*. *Breast Cancer*, 2010. **17**(2): p. 80-85.
215. Bettaieb, A., et al., *Precision medicine in breast cancer: reality or utopia?* *Journal of translational medicine*, 2017. **15**(1): p. 1-13.
216. Ginsburg, G.S. and K.A. Phillips, *Precision medicine: from science to value*. *Health Affairs*, 2018. **37**(5): p. 694-701.
217. Collins, F.S. and H. Varmus, *A new initiative on precision medicine*. *New England journal of medicine*, 2015. **372**(9): p. 793-795.
218. Sachdev, J.C., A.C. Sandoval, and M. Jahanzeb, *Update on precision medicine in breast cancer*, in *Precision Medicine in Cancer Therapy*. 2019, Springer. p. 45-80.