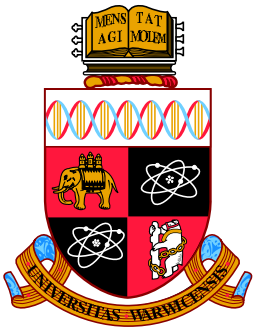


An Analysis of the Structure and Function of Malarial Duffy-Binding-Like Protein Domains using Recombinant Fusion Proteins

by

Shona Moore

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

June 2016

Liverpool School of Tropical Medicine

Warwick Systems Biology Doctoral Training Centre

EPSRC
Engineering and Physical Sciences Research Council

SYSTEMS BIOLOGY
DOCTORAL TRAINING CENTRE

LSTM
LIVERPOOL SCHOOL OF TROPICAL MEDICINE

Contents

List of Figures

List of Tables

Abbreviations

α_2M	α_2-macroglobulin
A1AT	α-1-antitrypsin
Ab	Antibody
ADCC	Antibody Dependent Cell-Mediated Cytotoxicity
ADCI	Antibody Dependent Cellular Inhibition
Amp	Ampicillin
AnAPN1	Anopheline midgut analyl aminopeptidase N 1
AP	Alkaline phosphatase
APC	Antigen presenting cell
BCR	B cell receptor
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
CD36	Cluster of Differentiation 36
CHO	Chinese hamster ovary
CSA	Chondroitin sulphate A
CSP	Circumsporozoite protein
DARC	Duffy antigen receptor for chemokines
DBL	Duffy-Binding-Like
DH5α	DH5α competent *E.coli* cells
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
DTT	Dithiothreitol
EBA	Erythrocyte Binding Antigen
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
FAB	Fragment antigen binding
FBS	Fetal bovine serum
Fc	Fragment crystallizable
FPLC	Fast protein liquid chromatography
HBSS	Hank's balanced salt solution
HeK	Human embryonic kidney
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase

ICAM-1	Intracellular adhesion molecule-1
IE	Infected erythrocyte
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgG1	Immunoglobulin G subclass 1
IgG2	Immunoglobulin G subclass 2
IgG3	Immunoglobulin G subclass 3
IgG4	Immunoglobulin G subclass 4
IgM	Immunoglobulin M
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVIG	Intravenous immunoglobulin
LB	Lysogeny broth
LDS	Lithium dodecyl sulphate
MBP	Mannose binding protein
MHC	Major histocompatibility complex
MOPS	3-(N-morpholino)propanesulfonic acid
MSP	Merozoite surface protein
MSP1	Merozoite surface protein-1
MSP3	Merozoite surface protein-3
MSPDBL	Merozoite surface protein duffy-binding-like
MWCO	Molecular weight cut-off
NEB	New England Biolabs
nIgM	Natural immunoglobulin M
NK cell	Natural killer cells
OD	Optical density
PAMPs	Pathogen-associated molecular patterns
PBMC	Peripheral-blood mononuclear cell
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline +0.05% Tween 20
PCR	Polymerase chain reaction
PES	Polyethersulfone
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PRRs	Pathogen recognition receptors
rSAP	Shrimp alkaline phosphatase

RTS,S	Malaria vaccine candidate
SDS	Sodium dodecyl sulphate
sIgM	Secreted immunoglobulin M
SOC	Super optimal broth
SPAM	Secreted polymorphic antigen associated with merozoites
TB	Terrific broth
TBE	Tris-borate EDTA
TBV	Transmission blocking vaccine
TCR	T cell receptor
TE	Tris-EDTA
TLR	Toll-like receptor
TNF	Tumour-necrosis Factor
TOP10	TOP10 competent *E.coli* cells
TRAP	Thrombospondin related anonymous protein
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultra violet
Var2CSA	Variant surface antigen 2 - CSA
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XL1-blue	XL1-blue competent *E.coli* cells
Zeo	Zeocin

Acknowledgements

Firstly, I would like to thank Prof. Richard Pleass and Dr. Hugo van den Berg without whose support this work would not have been possible. Particularly, thanks to Richard for his enthusiastic supervision! His contagious energy, optimism and immense knowledge have been inspiring. His patience and valuable advice throughout are greatly appreciated.

I am grateful to members of the Pleass group past and present for their support and guidance. In particular, a special thanks to Dr. Pat Blundell who has been a motivating and extremely encouraging mentor. Her technical support and guidance have been invaluable and her moral support truly appreciated. Thanks to Dr. Katy Lloyd for valuable discussions, support and reassurance throughout the darker times and for sharing the good times.

This project would not have been possible without the following collaborators to whom I am very grateful: Dr. Gavin Wright and Dr. Cecile Crosnier at the Sanger Institute for providing MSPDBL1 and MSPDBL2 recombinant proteins and αMSPDBL1 and αMSPDBL2 antibodies and Dr. Daniel Czajkowsky at Shanghai Jiao Tong University for providing molecular simulations.

I have been fortunate to have the support of both the Systems Biology Doctoral Training Centre and the Liverpool School of Tropical Medicine throughout this project. These welcoming and creative environments made the project all the more enjoyable. The encouragement and support of my colleagues, both in Warwick and Liverpool, is greatly appreciated. In particular, I am grateful to Dr. Brent Kiernan for the support and help provided. To my Systems/MOAC cohort, it has been a pleasure.

I could not have done this without my friends, family and loved ones. I genuinely cannot thank them enough for being there for me throughout the journey.

Finally, this work would not have been possible without the funding from EPSRC, to whom I express my gratitude.

Declaration of authorship

I declare that this thesis and the work presented in it is entirely my own original work, except where otherwise indicated. It has been composed by myself and has not been submitted in any previous application for any degree. Where I have consulted the work of others, this is always clearly attributed. Where work has been done in collaboration with others, I have made clear what was done by others and what I have contributed myself.

List of publications

i Cecile Crosnier, Zamin Iqbal, Ellen Kneupfer, Sorina Maciuca, Abigail J. Perrin, Gathoni Kamutu, David Goulding, Leyla Y. Bustamante, Alistair Miles, **Shona C. Moore**, Gordon Dougan, Anthony A. Hodder, Domonic P. Kwiatkowski, Julian C. Rayner, Richard J. Pleass, and Gavin J. Wright (2016) Binding of Plasmodium falciparum Merozoite Surface Proteins DBLMSP and DBLMSP2 to Human Immunoglobulin M is Conserved Amongst Broadly Diverged Sequence Variants. *The Journal of Biological Chemistry. doi* : 10.1074/*jbc.M*116.722074

ii Phyllis M. Quinn, David W Dunne, **Shona C. Moore**, and Richard J. Pleass (2016) IgE-tailpiece associates with α-1-antitrypsin (A1AT) to protect IgE from proteolysis without compromising its ability to interact with FcεRI. *Scientific Reports.* 6, 20509.

iii Richard J. Pleass, **Shona C. Moore**, Liz Stevenson, and Lars Hviid (2015) Immunoglobulin M: Restrainer of Inflammation and Mediator of Immune Evasion by *Plasmodium falciparum* Malaria. *Trends in Parasitology.* 1431, 1-12.

iv Daniel M. Czajkowsky, Jan T. Andersen, Anja Fuchs, Timothy J. Wilson, David Mekhaiel, Marco Colonna, Jianfeng He, Zhinfeng Shao, Daniel A. Mitchell, Gang Wu, Anne Dell, Stuart Haslam, Katy A. Lloyd, **Shona C. Moore**, Inger Sandlie, Patricia A. Blundell, and Richard J. Pleass (2015) Developing the IVIG biomimetic, Hexa-Fc, for drug and vaccine applications. *Scientific Reports.* 5, 9526.

Abstract

Duffy-binding-like domains are present in two potential malaria vaccine candidates. Located on the merozoite surface, MSPDBL1 and MSPDBL2 have been implicated in erythrocyte invasion and identified as targets of natural immunity. Merozoite DBL domains have been shown to bind the Fc region of natural IgM. This is characteristic of several PfEMP1s, and is also well documented in bacteria, viruses and other parasites, where it is thought to prevent specific binding of the more deadly IgG antibodies.

We have developed a mammalian expression system to produce merozoite DBL domains as Fc fusion proteins, facilitating investigation into their adhesive properties. Fc-fusion proteins are composed of the Fc region of IgG fused to a peptide and are a rapidly expanding field of bio-engineering. They have been successful in drug delivery due to their ability to increase serum half-life of the fused protein by the interaction of the IgG Fc with the neonatal Fc receptor (FcRn). Engineering of the Fc scaffold has shown improved receptor binding, allowing cross-linking of Fc receptors for improved vaccine design.

The expression of homodimeric DBL-Fc fusions is difficult, evidenced by incorrect folding and low protein yield. A flexible, extended hinge region was designed to increase the distance between the Fc and the fused DBL domain, and improved protein folding and IgM binding. Further work may optimise this hinge region for the development of malarial vaccines, or therapeutics for IgM-mediated diseases.

The structural analysis of all known IgM-binding DBL domains and residues on the merozoite DBL surface predict the involvement of helix 2a in IgM binding. This contradicts a recent homology model of the IgM-binding interaction, and suggests that the model needs revision. An improved DBL-Fc fusion could be used to identify critical binding residues located in this helix using the more focused approach of site-directed mutagenesis.

Chapter 1

Introduction

1.1 The Immune System

The human immune system has evolved as a result of powerful selection pressure imposed by pathogenic microbes (Cooper and Herrin, 2010; Medzhitov and Janeway, 1997). This complex and intricate system of cells, tissues and molecules is capable of detecting, neutralising and destroying infectious microorganisms, therefore protecting the host from infection (Charles A Janeway et al., 2001; Baron and Klimpel, 1996). The immune system comprises two interconnected and co-acting reactions: (i) innate and (ii) adaptive response (Alberts et al., 2002). In order to recognise and neutralise pathogens, innate and adaptive responses are both capable of distinguishing between self and non-self but they differ in their approaches.

The innate immune system provides a robust and immediate but non-specific response to pathogens, relying on conserved features essential to pathogen survival. This system is present in all multicellular organisms and employs two strategies, firstly by preventing the pathogen from gaining access, and secondly by the destruction of any invading pathogens, which is mediated by phagocytic cells and antimicrobial proteins.

The adaptive immune system, a distinct feature of vertebrate immune systems, is designed to recognise and remember specific pathogens (Alberts et al., 2002). Adaptive immunity involves the process of somatic cell gene rearrangement and relies on an enormous and diverse repertoire of antigen-specific recognition receptors which it builds up by exposure to pathogens. Although slower to respond, adaptive immunity is more effective than innate and is capable of initiating a stronger response every time a pathogen is encountered, offering long-lasting protection.

1.1.1 Innate immunity

The innate immune system plays a crucial role in early recognition and response to pathogens. The first line of defence is to use anatomical barriers to prevent it from entering the body. Epithelial surfaces provide an almost impermeable physical barrier while the acidic pH of the stomach and perspiration, antimicrobial secretions such as lysozyme, microbial antagonism by resident flora and surface-deposited antimicrobial lipids further deter invading microbes and prevent pathogens from colonising (Elias, 2007). However, should a pathogen pass these barriers and defences, the next strategy is to neutralise and destroy the invading microorganism (Alberts et al., 2002).

This line of defence is based on the detection of conserved products of microbial metabolism known as pathogen-associated molecular patterns (PAMPs) (Medzhitov, 2001). Crucially, these are readily distinguishable from "self" so that the host is not harmed. This means that pathogens of different biochemical composition and with entirely different life cycles can be recognised by overlapping mechanisms (Alberts, 2008; Mogensen, 2009). For example, any peptide of bacterial origin is identifiable by a formylmethionine at the N terminus which is a product of procaryotic translation and differs from the regular methionine produced in eukaryotic translation (Alberts, 2008). PAMPs are detected by pathogen recognition receptors (PRRs) present on innate immune cells such as macrophages and dendritic cells (Janeway and Medzhitov, 2002; Luster, 2002). The repertoire of PRRs is extensive and therefore recognises a diverse range of PAMPs. Moreover, a single pathogen may engage a number of PRRs via various PAMPs, hence forming a rapid and robust response. When a pathogen binds to the surface of an innate immune cell through the PAMP-PRR interaction, a neutralising or destructive response is initiated.

Phagocytosis, a central component of innate immunity, involves the internalisation and degradation of pathogens by phagocytic cells such as polymorphonuclear neutrophilic leukocytes and macrophages. This occurs in response to antigen presentation by a PAMP-PRR interaction. Macrophages amplify the inflammatory response by secreting cytokines and chemokines in addition to initiating phagocytosis. This attracts and recruits further phagocytic cells such as neutrophils and monocytes to the site of infection (Janeway and Medzhitov, 2002). Another method of phagocytosis involves opsonisation, or coating of a pathogen by antibodies (Abs) or complement proteins. This facilitates the uptake of the pathogen by a phagocytic cell. Alternative complement pathways may also be activated to destroy pathogens by lysis, and antibodies may also act to neutralise the microbe by blocking its target (Alberts, 2008). Further non-phagocytic cells such as natural killer (NK) cells and eosinophils all play specific roles in innate immunity.

1.1.2 Adaptive immunity

Although the innate immune system provides an immediate response to invading microbes, many pathogens have evolved strategies to evade complement pathways and prevent phagocytosis. The long-lasting adaptive immune system is capable of administering a

secondary, more specific, response to these pathogens. The adaptive immune system is recruited through the activation of antigen-specific lymphocytes by antigen-presenting cells (APCs) such as dendritic cells, macrophages and B cells.

Lymphocytes can be classified into two types, B or T cells which have epitope-specific receptors to recognise a large repertoire of antigen epitopes (Alberts, 2008). Adaptive immunity has two major branches: humoral and cell-mediated immunity, which are initiated by recognition of an antigen by B cell receptors (BCRs) and T cell receptors (TCRs), respectively.

The basis of cell-mediated immunity is the presentation of peptides on the surface of APCs by major histocompatibility complex (MHC) molecules (Rudolph et al., 2006; Moon et al., 2007). Foreign peptides are detected by TCRs which then initiate the proliferation and differentiation of naive T cells into effector T cells. Two main types of effector T cells develop depending on the class of pMHC presented - cytotoxic T cells (pMHC class I) and helper T cells (pMHC class II). Cytotoxic T cells directly kill cells expressing the appropriate antigen through the release of cytotoxins which lead to the caspase cascade and apoptosis, while helper T cells promote differentiation of cytotoxic T cells, mediate macrophage activity and promote B cell activation (Rudolph et al., 2006).

Humoral immunity is based on the recognition of antigens by immunoglobulins, also known as antibodies (Abs) which exist either in plasma or on the surface of B cells as part of the B cell receptor repertoire. Upon recognition of an antigen by the BCR, the complex is internalised by the B cell, expressed on the B cell surface as a peptide on a class II MHC and presented to the T cell receptors. Through binding of the peptide to the TCR, B cell activation is promoted by helper T cells. Naive B cells, once activated, proliferate and differentiate into both Ab-secreting plasma cells, which mediate clearance of pathogens, and memory B cells which provide a long-lasting response by persisting in the blood until the pathogen is encountered again.

There are five classes of human Abs: IgA, IgD, IgE, IgG and IgM. Before exposure to antigens naive B cells only express IgM, and express IgD and IgM when they mature (Edholm et al., 2011). When B cells are activated, isotype switching occurs, a mechanism enabling the Abs to switch between the five isotypes. Differentiation between the antibody classes depends on the type of antigen activating the B cell, as well as secondary signals such as pattern recognition by PRRs. Each Ab class plays its own specific role in humoral

immunity, and in combination they form a central part of adaptive immunity and are an important aspect of this report.

1.1.3 The structure and function of antibodies

Antibodies are immunoglobulins which bind to and neutralise the function of pathogens, while recruiting effector cells such as phagocytes and activating complement to promote pathogen clearance (Charles A Janeway et al., 2001). These glycoproteins form the basis of adaptive immunity and have a characteristic 'Y' shape, consisting of two light (λ or κ) and two heavy (α, δ, γ, ϵ or μ) polypeptide chains linked by covalent interchain disulphide bonds. The five classes of human Abs share the same basic structure and are categorised depending on which type of heavy chain they possess. Heavy (H) and light (L) chains both consist of a combination of variable (V) and constant (C) regions. L chains have one variable (V) and one constant (C) regions while H chains have one V and 3-4 C regions, depending on the Ab class. Antigen specificity relies on subtle differences in the V region.

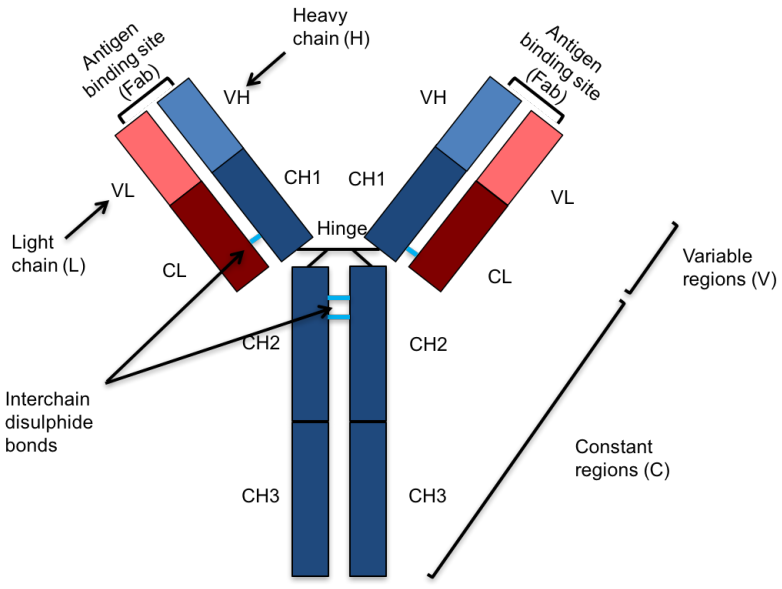


Figure 1.1: Schematic representation of an antibody, showing the heavy and light polypeptide chains which make up the characteristic 'Y' shape as well as the interchain disulphide bond structure. The two antigen-binding sites are at the tips of the arms and consist of one constant and one variable domain from each heavy and light chain.

An antibody has two distinct functional regions. The two identical arms of the 'Y' are termed the fragment antigen binding (Fab) region because of their ability to bind antigens. These arms contain variable regions which define the specificity of the antibody for the target antigen. The other fragment participates in the recruitment of effector molecules and is termed the fragment crystalisable (Fc) region. A hinge region links the Fc and Fab portions and allows flexibility within the Ab molecule. The Fc region recruits effector cells by binding to glycoproteins on the surface called Fc receptors (FcRs). Each antibody isotype binds to a specific type of FcR which can be found on the surface of multiple effector cells (Table 1.1). Depending on the effector cell recruited, Abs have a wide range of functions such as phagocytosis of Ab-coated pathogens, antibody dependent cell-mediated cytotoxicity (ADCC), mast cell degranulation and complement activation.

Table 1.1: Fc receptors (FcRs) and their corresponding antibody ligands.

Receptor	Ab Ligand	Cells
FcγRI (CD64)	IgG	Macrophages, Neutrophils, Eosinophils, Dendritic cells
FcγRIIA (CD32)	IgG	Macrophages, Neutrophils, Eosinophils, Platelets, Langerhans cells
FcγRIIB (CD32)	IgG	B Cells, Mast cells, Macrophages, Neutrophils, Eosinophils, Dendritic cells
FcγRIIIA (CD16a)	IgG	NK cells, Macrophages
FcγRIIIB (CD16b)	IgG	Eosinophils, Macrophages, Neutrophils, Mast cells, Follicular dendritic cells
FcRn	IgG	Dendritic cells, Epithelial cells, Endothelial cells, Hepatocytes, Monocytes, Macrophages
FcRL5	IgG	B cells
DC-SIGN	IgG	Macrophages, Dendritic cells
FcμR	IgM	B cells, T cells, NK cells
Fcα/μR	IgA, IgM	B cells, Macrophages
pIg	IgA, IgM	Epithelial cells
FcRL4	IgA	B cells
FcαRI (CD89)	IgA	Monocytes, Macrophages, Neutrophils, Eosinophils
FcεRI	IgE	Mast cells, Eosinophils, Basophils, Langerhans cells, Monocytes, Dendritic cells, Platelets
FcεRII	IgE	Mast cells, Eosinophils, Basophils, Langerhans cells

Immunoglobulin-M

Immunoglobulin M (IgM) is the first Ab to be produced by B cells and has an important role in the initial stages of immunity (Klimovich, 2011). The membrane-bound monomeric form of IgM forms the BCR which controls B cell responses through interactions with endogenous and exogenous ligands (Ehrenstein and Notley, 2010). It also exists on mucous membranes in a secreted form and is found in the circulation at concentrations of 1 - 2 mg/ml in blood and, with a half-life of $\sim$ 5 days, makes up about 15 % of Ig in the blood. The circulatory form can be divided into natural and immune IgM, both of which are crucial for Ab-mediated response to pathogens (Czajkowsky and Shao, 2009). The broadly reactive but low-affinity natural IgM (nIgM), produced by the B1 subset of B cells, occurs naturally in the blood of mice raised under germ-free conditions (Rapaka et al., 2010). In contrast, immune IgM is secreted by B-2 cells following exposure to pathogens and is mostly antigen-specific (Ehrenstein and Notley, 2010) although production levels fall upon development of IgG and other Abs. Immune IgM has been shown to be crucial for protection from viruses, bacteria, protozoa, fungi and helminths (Pleass et al., 2015).

IgM molecules have a subunit with the characteristic 'Y' shape seen in other Ab isotypes, formed from two light chains and two heavy chains. Each heavy chain consists of one variable region (Vμ) and four constant domains (Cμ1 - Cμ4) as well as a tailpiece (tp) located at the C terminus of the heavy chain (Sørensen et al., 1999). The tailpiece enables polymerisation and IgM exists both as pentamer with a five-subunit structure (mass $\sim$970 kDa) and as a hexamer consisting of six subunits ($\sim$1132 kDa). The pentameric form contains a 15 kDa joining (J)-chain, attached to the tailpiece by disulphide bonds (Figure 1.2). The hexameric form lacks the J chain and is less common than the pentameric form, making up 20% of the total IgM. Hexameric IgM has increased avidity for both antigen and receptor, and is 15-20 times more efficient at activating complement than pentameric IgM (Pleass et al., 2015; Randall et al., 1990; Wiersma et al., 1998). Hexameric IgM without a J-chain is unable to bind to the polymeric Ig receptor (pIgR) which mediates transport of secreted pentameric IgM across epithelia, releasing them to mucosal surfaces (Johansen et al., 2000). This suggests that the role of hexameric IgM is solely in humoral immunity (Randall et al., 1990; Wiersma et al., 1998).

The unique multimeric structure gives IgM high avidity and diverse function, including roles in infection, B cell homeostasis, inflammation, atherosclerosis and autoimmunity

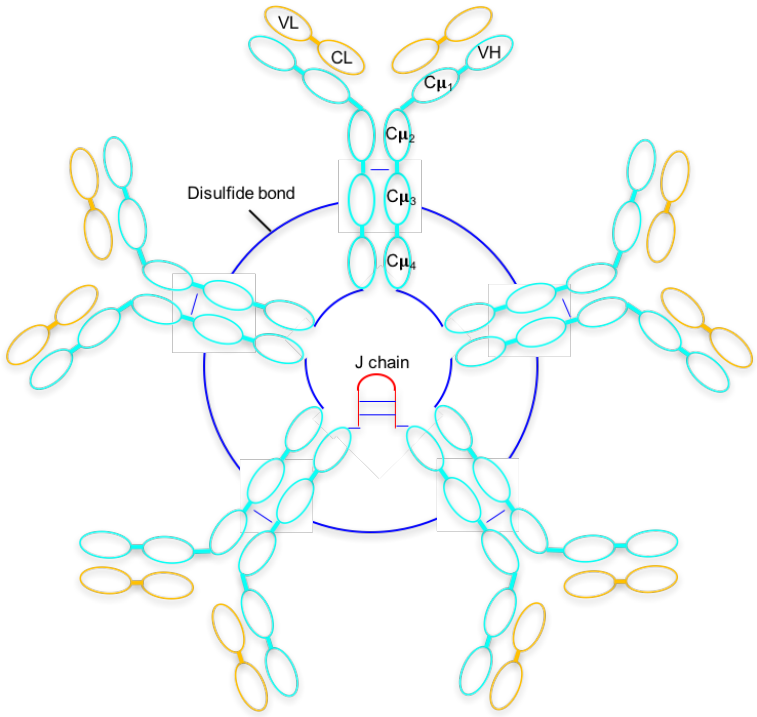


Figure 1.2: Simple schematic representation of a pentameric IgM molecule (five monomeric IgM molecules) showing the heavy and light polypeptide chains. The Fab domain contains one variable and one constant domain Cμ1. The Fc region contains three further constant domains Cμ2 - Cμ4. Disulphide bonds cross-link adjacent Cμ3 and Cμ4 domains of individual monomeric molecules and are shown in blue. The location of the J chain is given in red.

(Ehrenstein and Notley, 2010; Panda and Ding, 2015). Integral to the impressive properties of multimeric IgM is it's ability to bind C1q and recruit complement. A single pentameric IgM molecule can activate the complement cascade and lyse a red blood cell, an action that would take approximately a thousand IgG molecules to accomplish (Cooper, 1985). Due to it's size, attempts to obtain a crystal structure of IgM have been so far been unsuccessful (Müller et al., 2013). However, the structure has been modelled by Czajkowsky and Shao (2009) based on the crystal structure of IgE Fc. The model predicted an unexpected, non-planar, mushroom-shaped complex with the central portion (formed of C terminal domains) protruding above the plane created by the Fab domains (Czajkowsky and Shao, 2009). Interestingly, this structure was confirmed by cryo-atomic force microscopy and is thought to leave the C1q binding region fully exposed, explaining the mechanistic first steps of complement activation by IgM (Czajkowsky and Shao, 2009). IgM is heavily glycosylated and has five N-linked glycosylation sites present on each heavy chain as well as a single site on the J-chain (Klimovich, 2011). Glycosylation is thought to be critical

for structural integrity and stability of the Ab (Sola and Griebenow, 2009; Arnold et al., 2007).

Natural IgM is a vital part of the early innate immune response and is the first line of defence for protection against many pathogens. It is also able to recognise self-components. The multimeric structure gives the molecule poly-reactivity due to the presence of multiple antigen binding sites. This enables effective recognition of conserved structures on invading organisms. Although affinity for antigen is low, poly-reactivity also enables nIgM to bind multiple structures on the same antigen, thereby enhancing neutralisation by promoting it's elimination through clearance by phagocytic cells (Ehrenstein and Notley, 2010). High poly-reactivity also gives nIgM a superior ability to detect transformed cells in malignancy where many tumour-related epitopes exist as repetitive carbohydrate motifs (Kaveri et al., 2012). In a study of tumour specific antibodies where thousands of Abs were isolated from patients, all were found to be IgMs (Brändlein et al., 2003). Most of these IgM Abs were nIgM which bound to modified tumour-specific receptors (Brändlein et al., 2003). Similarly, nIgM taken from healthy patients was found to have tumour specificity (Brändlein et al., 2003). Natural IgM therefore has significant importance in the immunogenic response to tumours and is routinely used in diagnosis and therapy of malignancies.

Although an ineffective opsonin by itself, nIgM is highly effective in facilitating clearance of small apoptotic particles, senescent erythrocytes and mis-folded proteins through complement-dependent mechanisms (Pleass et al., 2015). This crucial function is necessary for tissue homeostasis in order to prevent uncontrolled inflammation and to suppress autoimmunity. An accelerated development of IgG autoantibodies and autoimmune disease is seen in the absence of nIgM (Boes et al., 2000).

As well as being a vital aspect of the early immune response, nIgM is considered to provide an essential boost to the adaptive immune response, particularly that of IgG. In a study of nIgM-deficient mice, production of antigen-specific IgG was lower when immunised with T-cell dependent antigen and the specificity of this IgG lower than in control mice (Boes et al., 1998). Natural IgM provides a invaluable immune-surveillance system capable of preventing dissemination of microbes, cell debris and transformed cells within the host and is able to prime the more specific response of the adaptive immune system, particularly that of IgG.

IgM-Fc receptors

The Fc region of an Ab mediates effector function through binding of host ligands and Fc receptors (FcRs). There are several known FcRs for IgM, the polymeric Ig receptor (pIgR), Fcα/μR, FcμR as well as a number of ligands known to bind IgM including mannose-binding-lectin (MBL), CD22, TRIM21 and Sp alpha (McMullen et al., 2006; Adachi et al., 2012; Mallery et al., 2010; Tissot et al., 2002). Due to the multimeric structure discussed previously, affinity for IgM is high. The pIgR binds both secretory IgM and IgA on epithelia, as discussed previously. Another FcR shared with IgA is the Fcα/μR expressed on B cells and macrophages which is involved in the priming of helper T lymphocytes and in defence against bacteria in peripheral organs (Sakamoto et al., 2001). Recently, the FcμR has been identified and is expressed in humans on B cells, T cells and NK cells (Kubagawa et al., 2009; Shima et al., 2010). In mice, the FcμR is only expressed on B cells, and studies of FcμR deficient mice have shown elevated non-immune IgM levels in serum as well as natural autoantibody levels (Honjo et al., 2012; Ouchida et al., 2012). This suggests a role for FcμR binding in IgM homeostasis, autoimmune suppression and regulation of humoral response (Honjo et al., 2012; Ouchida et al., 2012). The function of the human FcμR receptor remains unclear.

Immunoglobulin-G

Immunoglobulin G (IgG) is an Ab of molecular weight ∼150 kDa consisting of a single 'Y' subunit formed from two light and two heavy chains. Each heavy chain consists of one variable region (Vγ) and three constant domains (Cγ1 - Cγ3). IgG is produced late in the primary immune response and is the most abundant Ab found in serum and extracellular fluid (making up 75% of all Ig and found at 10 mg per ml in the circulation). IgG Abs have a wide range of functions and are involved in neutralisation of bacteria, viruses and toxins, as well as opsonising pathogens for engulfment by phagocytes and activating complement through C1q binding.

IgG can be divided into four subclasses by constant region: IgG1, IgG2, IgG3 and IgG4. The wide range of effector function seen amongst the subclasses comes largely from variation in the constant region of the Fc, and as a result the different subclasses bind to different IgG-Fc receptors (FcγR) as well as C1q (See Table 1.1 for FcRs). Structural variation is also found in the hinge region which links the Fab and Fc regions, thereby

altering the conformation of the Ab and further influencing effector function. Although deficiency in a single IgG subclass can lead to susceptibility of an individual to specific pathogens, an IgG subclass deficiency is not usually detrimental (Vidarsson et al., 2014).

IgG contains a conserved N-linked glycosylation site at Asn-297 which is located between the Cγ2 and Cγ3 domains and creates an exposed docking site for FcγR binding (Vidarsson et al., 2014). Fc glycans are known to play a role in Ab stability, and numerous studies have investigated the impact of glycosylation on structural integrity and effector function of IgG (Krapp et al., 2003; Zheng et al., 2014).

IgG-Fc receptors

IgGs are the only Abs to be passed from mother to fetus via the placenta by binding to the neonatal Fc receptor (FcRn) (Koch et al., 1967; Brambell, 1966). Binding to the FcRn also aids recycling of IgG by preventing lysosomal degradation. FcRn is unrelated to the classic FcγRs and is more similar in structure to MHC class I (Story et al., 1994; Roopenian et al., 2003). Binding occurs in the Cγ2 - Cγ3 interface of the IgG Fc domain (Wines et al., 2000). IgG binds to FcRns on epithelial cells in a variety of tissues, including the placenta where IgG are transported through vesicles and released by exocytosis into fetal blood (Alberts, 2008; Junghans and Anderson, 1996). FcRns are also located on white blood cells where they are important for IgG-mediated phagocytosis (Vidarsson et al., 2006). The half-life of IgG in the blood is greater compared to other Ab isotypes due to recycling mediated by the FcRn. The FcRn binds IgG at a slightly acidic pH, which is found in the intestinal lumen as well as in acidic endosomes. IgG internalised by pinocytosis binds the FcRn and is recycled to the cell surface and released at the basic pH of blood. This prevents lysosomal degradation and increases the half-life of IgG (Vidarsson et al., 2006; Ghetie et al., 1996).

There are six classic human FcγRs, four of which are activating Fc receptors while two, FcγRIIB and FcγRIIIB (also known as CD16b), are inhibitory Fc receptors. The inhibitory FcRIIB is a low affinity receptor and although homologous to the activating FcRs, contains a distinct ITIM sequence in the cytoplasmic domain (Ravetch and Bolland, 2001). The four activation FcRs (FcγRI, FcγRIIA, FcγRIIC and FcγRIIIA) are characterised by the presence of an ITAM motif and are found on most effector cells apart from lymphoid cells (Ravetch and Bolland, 2001). Activating and inhibitory FcγRs

are co-expressed on the surface of effector cells and the ratio of their expression determines the effector response.

As well as FcγRIIB, another low-affinity inhibitory FcR is located on the surface of dendritic cells, the dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) (Geijtenbeek et al., 2000). DC-SIGN is an important receptor in the efficacy of IVIG in controlling autoimmune disease, and binding to this receptor requires specific glycosylation patterns (Czajkowsky et al., 2015; Geijtenbeek et al., 2000).

The FcR-like protein FcRL5 expressed by B cells is homologous to FcRγI and has recently been identified as an IgG receptor. FcRL5 is thought to play a role in the regulation of the B cell response (Wilson et al., 2012).

Immunoglobulin-E

Immunoglobulin E (IgE) is a monomeric Ab of low abundance (approximately 0.1 μg per ml in circulation), consisting of two light and two heavy chains. Unlike IgG which has three, IgE has four constant domains (Cε1 - Cε4) meaning that it is heavier than IgG at ∼190 kDa. The Cε3 and Cε4 domains of IgE are homologous to the Cγ2 and Cγ3 domains of IgG. The Cε2 domains of IgE are therefore their most distinguishing feature and these replace the flexible hinge region of IgG. From the crystal structure of the IgE-Fc it is notable that the Cε2 domains bend back and make extensive contact with the Cε3 domains, causing an acute bend in the IgE molecule (Wan et al., 2002). This gives it a unique and more rigid structure than IgG.

IgE plays a role in both allergic and anti-parasitic response and is involved in the activation of mast cells and basophils leading to the consequential release of mediators such as histamine, which are associated with IgE-dependent allergic reactions (Kinet, 1999). When FcεRs are activated in the lungs, the release of toxic mediators are induced, causing the symptoms of asthma (Gould and Sutton, 2008).

IgE is highly susceptible to cleavage and inactivation by proteases produced by parasitic helminths (Quinn et al., 2016). Several splice variants of IgE are known to exist, including one such variant called IgE-tailpiece (IgE-tp), which has eight novel residues including a cysteine in place of the two carboxy-terminal amino acids in classical IgE (Quinn et al., 2016). This variant has been shown to associate with α-1-antitrypsin (A1AT) in plasma which protects the Ab from proteolysis while enhancing the interaction with FcεRI (Quinn

et al., 2016).

IgE-Fc receptors

Two IgE-Fc receptors are known, the high affinity FcεRI and the lower affinity FcεRII which are present on most immune cells. The high affinity FcεRI captures both Cε3 domains of IgE and the subsequent and exceptionally slow rate of dissociation on mast cells and basophils is responsible for persistent sensitisation of these cells to allergic challenge (Stone et al., 2010).

Immunoglobulin-A

The highly hydrophilic immunoglobulin A (IgA) is the most abundant Ab in secretions at mucosal sites (making up 10 % of Ig in total). Consisting of the characteristic 'Y' structure common to most Ab molecules, IgA contains two heavy and two light chains. A molecule of IgA has three constant regions (Cα1 - Cα3) and, like IgM, contains a tailpiece and J-chain at the heavy chain C terminus (Kerr, 1990). This also gives IgA the ability to polymerise, and although it is primarily found in monomeric form it can also take a dimeric form (Atkin et al., 1996). In serum, catabolic rates of IgA are five times faster than IgG which is synthesised at the same rate of IgA, therefore the concentration of IgA is about one-fifth the concentration of IgG (Kerr, 1990). In secretions, IgA synthesis rates are far higher than any other Ab class. IgA is therefore overall the most highly synthesised Ab isotype with more IgA produced in total than all of the other Ab classes combined (Kerr, 1990).

IgA does not have the ability to activate complement and it lacks residues in the Fc region which have been identified in IgG and IgM as critical for C1q binding. Two distinct subclasses of IgA exist: IgA1 (90%) and IgA2 (10%). IgA1 has an extended hinge region compared to IgA2 which has an additional 13 amino acid, highly O-glycosylated region. This extended hinge region makes IgA1 susceptible to cleavage by proteases and therefore IgA2 dominates at mucosal sites, whereas IgA1 is more common in serum. Like IgM, IgA is capable of travelling across epithelial boundaries through binding of the J-chain to the polymeric Ig receptor (pIg). It is thought that the primary function of IgA is in the neutralisation of pathogens at mucosal sites such as gastrointestinal, genitourinary and respiratory tracts (Hayes et al., 2014).

IgA-Fc receptors

Due to it's ability to polymerise and the presence of the J-chain, secretory IgA binds the polymeric Ig receptor (pIg) through the Cα3 domain, leading to transport across epithelial surfaces (Brandtzaeg and Prydz, 1984). Another receptor shared with IgM is the Fcα/μR expressed on B cells and macrophages. For this receptor interaction, the Cα2 - Cα3 domain interface is critical while the J-chain is not considered essential for binding (Sakamoto et al., 2001). IgA has a high level of glycosylation although its glycans are not required for receptor interaction (Mattu et al., 1998).

A receptor unique to IgA, is the FcαRI (CD89), thought to play a role in mucosal immune defence (Bakema and van Egmond, 2011). FcαRI has homology with FcγRs and FcεRI, although the binding site at the Cα2 - Cα3 interface is distinct from the homologous regions in IgG and IgE at the hinge proximal regions of Cγ2 and Cε3 respectively (Woof, 2002).

The FcR-like protein FcRL4 expressed by B cells has recently been identified as an IgA receptor. FcRL4, like the IgG receptor FcRL5, is thought to play a role in the regulation of the B cell response (Wilson et al., 2012).

Immunoglobulin-D

Co-expressed with IgM by mature B cells before isotype switching triggers expression of the other Ab classes, immunoglobulin D (IgD) is a monomeric Ab of relatively low abundance (approximately 30 μg per ml in circulation). The role of IgD is not fully understood. IgD exists in both membrane-bound form and in serum and it consists of two light and two heavy chains. Although IgD has three constant domains (Cδ1 - Cδ3) similar to IgG, it has a long hinge region which gives it a unique flexible 'T' shape compared to the traditional 'Y' shape of other Ab isotypes. This makes the molecule highly flexible thus enhancing antigen binding, although making it susceptible to proteolytic degradation (Chen and Cerutti, 2011; Vladutiu, 2000).

IgD-Fc receptors

The FcR for IgD remains elusive although there is evidence for the presence of FcδRs on B cells, basophils and mast cells (Chen and Cerutti, 2011).

1.2 Fc Fusion Technology

Fc fusion therapeutics are a fast growing field of bio-engineering. Monoclonal antibodies (mAbs) have dominated the field of therapeutics in recent years mainly due to the biological and pharmacological properties resulting from the effector function of the Fc region. A new generation of therapeutics whereby the Fc region of an IgG antibody is fused to an active protein drug, are proving successful. The IgG Fc allows interaction with FcRs on immune cells and gives the protein of interest antibody-like properties. The IgG Fc also binds to the neonatal Fc receptor (FcγRn) an action which facilitates recycling, thus protecting the active protein from endosomal degradation. The addition of the IgG Fc therefore increases the serum half-life of the protein of interest, extending exposure to the target by limiting renal clearance, consequently improving pharmacological effect and its therapeutic potential. This technology has opened up therapeutic use and drug delivery of smaller active proteins in particular as clearance of such molecules (< 60 kDa) is rapid with half-life ranging from between a few minutes to hours. Fc-fusion proteins also form homodimers which improves avidity of the active protein giving it further therapeutic advantage (Figure 1.3).

Other fusion proteins which are designed to improve the half-life of the fused protein include HSA-fusions and Transferrin-fusion proteins. The 66.5 kDa human serum albumin (HSA) is also known to bind the FcRn, with a slightly lower affinity than IgG and therefore shorter half-life (456 hours compared to 480 hours for IgG) (Strohl, 2015). HSA-fusion proteins are also currently in development and are showing some success in clinical trials with Tanzeum (GlaxoSmithKline) receiving FDA approval in April 2014 (Strohl, 2015). Novozymes have also shown that an engineered HSA is able to improve half-life by improving the affinity for FcRn 12-fold (Andersen et al., 2014). This has potential to improve the half-life of HSA-fusion therapeutics. Further fusion proteins designed to improve half-life include fusion of Transferrin or CTP although so far none have FDA approval (Strohl, 2015).

Fc-fusions have additional practical advantages which add to their success over other fusion proteins. Cost-effective and efficient purification by protein A/G affinity chromatography streamlines the manufacturing process while the attachment of the IgG Fc domain improves stability and solubility, further simplifying the manufacturing process. As a result, Fc-fusion proteins are fast gaining commercial success and there are currently

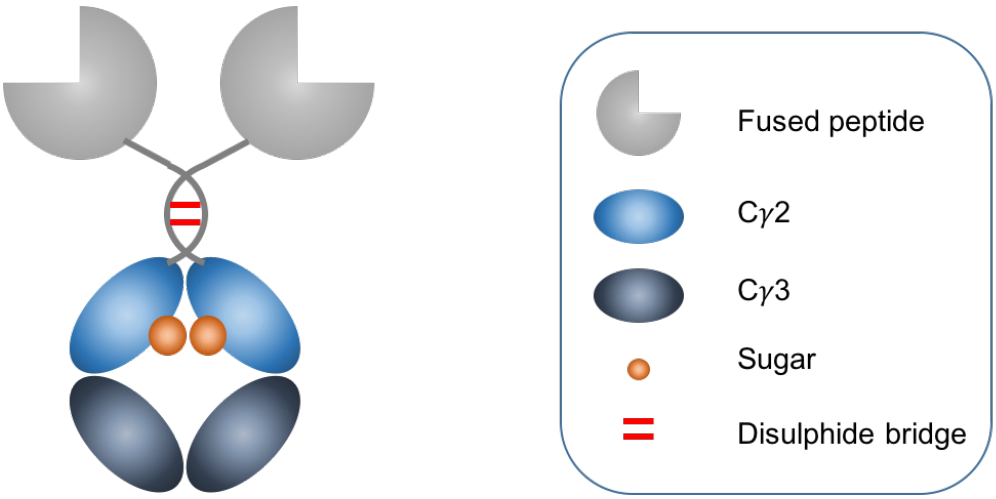


Figure 1.3: **Schematic representation of a dimeric Fc-fusion protein.** The fused peptide is covalently attached to the Cγ2 via a hinge region which forms a disulphide bridge and produces a dimeric fusion protein. Adapted from Czajkowsky et al. (2012).

nine approved by the FDA for use in drug delivery (Table 1.2). Like many mAbs, most of these Fc-fusion proteins target receptor-ligand interactions acting either as antagonists or as agonists to manipulate immunity by either blocking or stimulating receptor function (Czajkowsky et al., 2012).

Table 1.2: FDA approved hIgG Fc-fusion proteins currently in the clinic.

Trade Name	Base protein	Approved use	Year of Approval	Molecular weight (kDa)
Eloctate (aloctate)	Antihemophilic factor	Hemophilia A	2014	220
Alprolix	Coagulation Factor IX	Hemophilia A	2014	98
Nulojix (belatacept)	CTLA-4	Organ rejection	2011	90
Eylea (aflibercept)	VEGFR1/ VEGFR2	Age related macular degeneration	2011	115
Arcalyst (rilonacept)	IL-1R	Cryopyrin-associated periodic syndromes	2008	251
NPlate (romiplostim)	Thrombopoietin-binding peptide	Thrombocytopenia	2008	59
Orencia (abatecept)	Mutated CTLA-4	Rheumatoid arthritis	2005	92
Amevive (alefacept)	LFA-3	Psoriasis and transplant rejection	2003	73
Enbrel (etanercept)	TNFR	Rheumatoid arthritis	1998	51

Currently, all FDA-approved Fc-fusion proteins use the Fc region from the IgG1 subclass. This is due to its increased half-life over other subtypes as well as its superior engagement of FcγRIIA which makes it the preferred subclass for use in delivery of antigen to APCs (Czajkowsky et al., 2012). It has been suggested that IgG3 could be used to improve complement activation. However, the increased length of the IgG3 hinge region means that it is more susceptible to proteolytic cleavage than other subclasses and this technical difficulty would have to be overcome for its use in Fc-fusion proteins (Czajkowsky et al., 2012).

The recent success of Fc-fusion proteins has mostly been focussed on drug delivery, although direct targeting of FcRs by the Fc-fusion protein gives promise for Fc-fusion based vaccines (Czajkowsky et al., 2012). Once the target FcRs are identified, the need for receptor cross-linking in enhanced cell signalling highlights the importance of valency particularly when targeting low-affinity receptors. It has recently been shown that Fc-fusions can be modified to polymerize into a hexameric form with twelve fused partners (Mekhaiel et al., 2011). The Hexa-Fc scaffold has thus been developed to oligomerise monomeric IgG Fc into hexameric oligomers (Czajkowsky et al., 2015). This construct is able to enhance binding to low-affinity receptors with increased avidity, thus opening up the use of Fc-fusion proteins as vaccines (Czajkowsky et al., 2015).

Further modifications to the Fc-scaffold may improve the therapeutic potential of an Fc-fusion protein, including modifications to the hinge region or to the fused protein. These have been comprehensively reviewed by Czajkowsky et al. (2012).

1.3 Parasite Fc-Binding Proteins

We have seen that the function of the five Ab classes and the recruitment of effector cells relies on Ab Fc binding to glycoprotein FcRs which are present on immune cell surfaces. A successful field of bio-engineering takes advantage of these interactions to develop successful Fc-fusion based therapeutics and has further potential for the development of vaccines.

In some cases, an FcR-like pathogen is also able to take advantage of this system; hijacking the Ab by binding to the Fc portion of a non-immune antibody instead of the antigen-specific Fab region. This allows the pathogen to evade immunity as the Ab is prevented from interacting with host FcRs, the effector function is blocked and the pathogen is masked in host non-specific Ab. Parasites, which rely on remaining undetected within their host for survival, are particularly proficient in this immune evasion strategy.

A large number of human parasites are able to bind the Fc portion of immunoglobulin non-specifically and prevent downstream Fc-mediated destruction. This strategy, mediated by Ig-binding molecules expressed on the surface of the parasite is also well documented in bacteria and viruses such as *Staphylococcus aureus* which expresses protein A, herpes simplex virus which expresses the glycoprotein complex gE-gI, a known Fcγ receptor, as well as IgA-binding proteins identified in many strains of *Streptococcus* (Watkins, 1964; Lubinski et al., 2011; King and Wilkinson, 1981; Pleass et al., 2001; Nezlin and Ghetie, 2004; Pleass and Woof, 2001). In parasites, Fc-binding proteins which play a role in evasion of the immune response may also add to the infectivity and persistence of the parasite in circulation.

IgM Fcμ-binding proteins are less well documented than for IgG and IgA. This may be due to difficulties in distinguishing between low-affinity natural IgM Fab binding and Fc-receptor interactions (Czajkowsky et al., 2010). Several protozoa are known to utilise IgM-binding proteins, including *Toxoplasma gondii* and pathogenic species of *Trypanosomitidae* (Vercammen et al., 1999; Vincendeau and Daëron, 1989; Czajkowsky et al., 2010). More recently, proteins from the *Plasmodium falciparum* erythocyte membrane protein 1 (PfEMP1) family, expressed on the surface of *P. falciparum*-infected erythrocytes (IEs) have been shown to bind natural IgM via the Fc. Only certain *P. falciparum* strains have been shown to bind IgM, but binding phenotypes are linked to the most serious clinical effects of malaria. PfEMP1 proteins are known to contribute to

the pathogenesis of malaria by mediating cytoadhesion (Smith et al., 2013). The role of IgM-binding in PfEMP1 is discussed further in Section 1.5.3.

Recently, *P. falciparum* proteins located on the surface of the merozoite, the infective form of the parasite, have also been shown to bind natural IgM via the Fc (Crosnier et al., 2016). The role of these merozoite proteins is unclear but interestingly they contain a Duffy-Binding-Like (DBL) domain which is also present in PfEMP1s and known to be the FcR for many IgM-binding isolates as well as the binding site for other cytoadherant phenotypes. These binding domains will be a key focus of this report. In order to understand why proteins expressed at different stages in parasite development bind to IgM via the Fc and contain homologous binding regions, it is first necessary to understand the immune response at various stages of the parasite's complex lifecycle and the immune evasion tactics employed by the parasite. This will be discussed in Section 1.4.

1.4 Immunity and Malaria

With around half of the world's population at risk, malaria is a major global health problem. Caused by the highly successful *Plasmodium* parasite, which is injected into a human host by the *Anopheles* mosquito upon taking a blood meal, in 2013 there were an estimated 198 million clinical cases of malaria worldwide, 584,000 of which resulted in death (WHO, 2015). Malaria immunity develops slowly and as a result 78% of malaria deaths occur in children under five years of age. Moreover, 90% of all malaria deaths occur in Africa where many mothers are unable to afford or access treatment, therefore prevention is critical (WHO, 2015; Miller et al., 2002). Despite some recent success in prevention and control measures (Yewhalaw et al., 2011; Greenwood and Targett, 2011), mosquito-insecticide resistance and drug-parasite resistance threaten to undermine these efforts. Antibody response is an important aspect in protective immunity to *Plasmodium*, particularly in the blood stages of infection where several parasite antigens that are known to cause clinical symptoms drive strong antibody responses. Antigenic variation means that immunity to a malarial infection is short-lived and leads only to partial immunity which is unable to protect an individual against a new infection. Malaria vaccine development is urgently needed but the complexity of the parasite lifecycle and intricacy of the host-parasite interaction means that it is a great scientific challenge to develop an effective vaccine.

There are more than 100 species of *Plasmodium*, five of which cause malaria in humans (*Plasmodium falciparum*, *P. vivax*, *P. malariae*, *P. knowlesi* and *P. ovale*). While *P. falciparum* is the most virulent species and causes the highest rates of complications and mortality, *P. vivax* is the most widely distributed species worldwide. In Africa, however, the risk of infection with *P. vivax* is low due to the absence of the Duffy gene in many populations, which is required for infection of red blood cells. Here, where the vast majority of malaria deaths occur, *P. falciparum* poses a huge burden on the poorest and most vulnerable communities.

1.4.1 Lifecycle of *Plasmodium falciparum*

A human host becomes infected with the malarial parasite *P. falciparum* when a parasite-laden *Anopheles* mosquito takes a blood meal. During the mosquito stage of infection, gametes reproduce in the midgut to produce diploid zygotes. These develop

within the midgut into ookinetes which burrow through the midgut wall and form oocysts on the other side. As oocysts grow and divide, thousands of active haploids called sporozoites are produced which burst into the body cavity of the mosquito where they travel to and invade the salivary glands. It is from the saliva of the infected mosquito that sporozoites are injected into the human host's skin where they migrate into the blood stream.

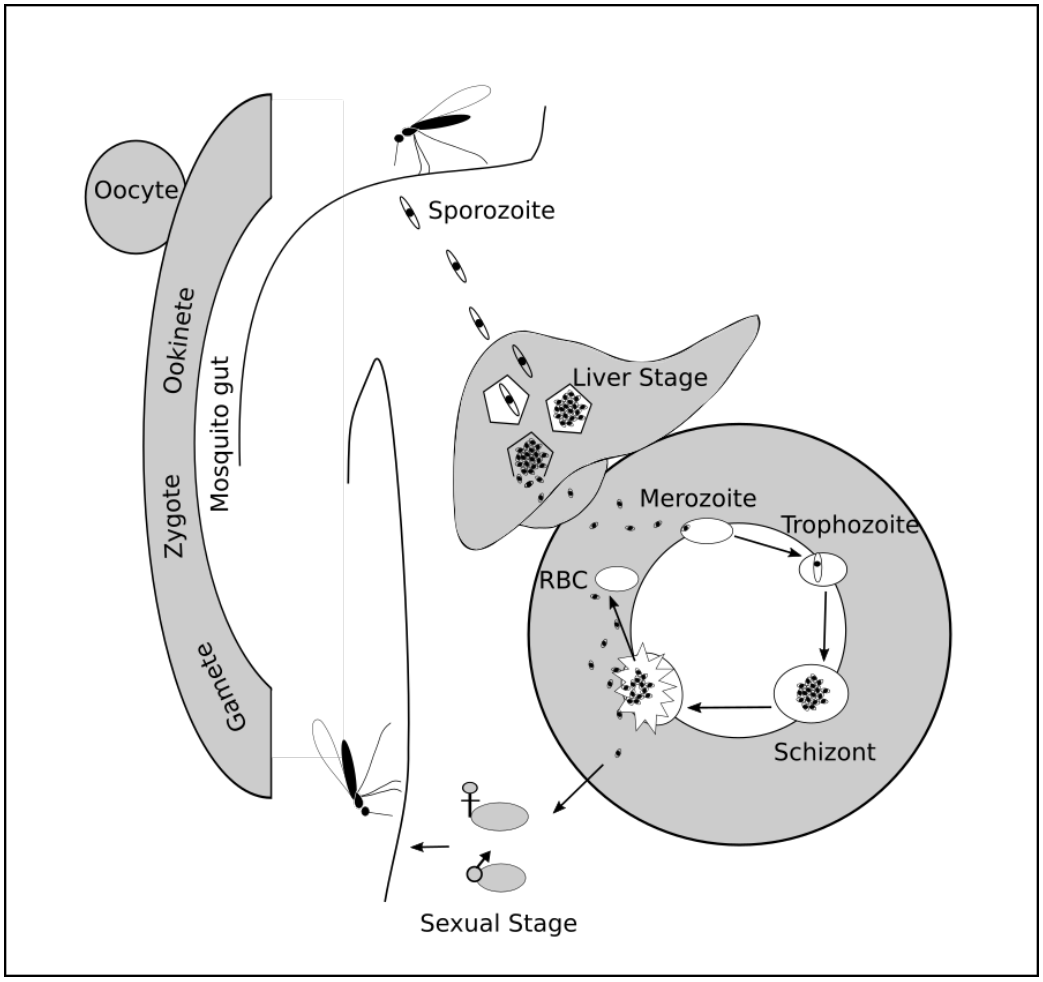


Figure 1.4: **Lifecycle of *Plasmodium falciparum* within the mosquito and human hosts.** Upon taking a blood meal, an *anopheles gambiae* mosquito injects sporozoites into the human host. These migrate to and invade liver cells where they differentiate and multiply, generating merozoites, the invasive form of the parasite. Merozoites subsequently infect and then multiply inside erythrocytes or differentiate into gametocyes which are injested by a mosquito upon taking a blood meal. The change of environment in the mosquito midgut induces gamete formation and the life cycle repeats.

Sporozoites are transported through the blood stream to the liver sinusoids where they migrate across the sinusoidal wall in order to invade hepatocytes. Invasion occurs rapidly,

approximately within 60 minutes of inoculation (Baron et al., 1996). A single sporozoite multiplies and re-differentiates within an infected hepatocyte over the course of 5 - 10 days to generate thousands of merozoites, which are the invasive form of the parasite (Maier et al., 2009). Subsequently, bulging hepatocytes burst and merozoites are released into the blood stream where they rapidly begin a cycle of invasion and infection of host erythrocytes.

Merozoites within an infected erythrocyte consume haemoglobin, giving them energy to develop into trophozoites. Upon gorging itself on haemoglobin, a trophozoite undergoes schizogony and develops into a schizont, consequently another round of asexual replication then produces multiple merozoites per schizont. A bulging schizont ruptures 48 hours after infection of the erythrocyte, releasing merozoites into the blood stream once again where they infect further erythrocytes. The cycle of infection, multiplication and bursting of erythrocytes continues, producing a high parasite burden on the host. The blood stage parasites are responsible for clinical manifestations of the disease and the cycle of fresh merozoites bursting into the blood stream responsible for the waves of parasitemia.

A few blood stage merozoites don't develop into schizonts and multiply, instead they differentiate into gametocytes. Upon ingestion when a mosquito takes a blood meal from an infected human host, the change in environment in the mosquito midgut induces gamete formation and the life cycle repeats (Figure 1.4).

Plasmodium falciparum multiplies rapidly within a human host during cycles of erythrocyte invasion by merozoites. The host first encounters the parasite during the symptomless pre-erythrocytic phase and the parasite is exposed to the immune system again in the form of the infective merozoite. The immune response to malaria is complex and stage-specific. Individuals exposed to repeat infection do eventually develop immunity, although as it is slow to develop children are of high risk.

1.4.2 Innate immunity and malaria

A co-ordinated increase is seen in the levels of pro-inflammatory cytokines in malaria-naive individuals at the time that the parasite emerges from the liver and at the first infection of erythrocytes (Hermsen et al., 2003). This confirms the *in vitro* observation that infected erythrocytes (IEs) induce tumour-necrosis factor (TNF) and the production of other pro-inflammatory cytokines by peripheral-blood mononuclear cells (PBMCs) of naive

donors within 10 hours (Stevenson and Riley, 2004; Scragg et al., 1999). Recently, it has been suggested that innate immunity is triggered when parasite density crosses a threshold value (Stevenson and Riley, 2004). This means that blood parasite densities oscillate between a high level (where innate responses are triggered and parasites are partially cleared) and low level (where innate immune responses are not triggered) (Stevenson and Riley, 2004). This implies that innate responses are essential during the initial phase of infection and allow the host time to develop an adaptive response. From an evolutionary perspective, this is advantageous for the parasite as well as the host. A reduction in the virulence of the initial infection means a decreased incidence of early host death, the parasite therefore has more chance of being passed to the next mosquito host (Stevenson and Riley, 2004).

There have been relatively few investigations into the role of innate immunity in malaria. It has been suggested that regulatory cytokines produced by innate immune cells in response to malaria modulate the subsequent adaptive immune response (Stevenson and Riley, 2004). This may also help provide a first step in downstream T-cell activation (Stevenson and Riley, 2004). Macrophages, dendritic cells (DCs) and B cells have been implicated as the antigen-presenting cells (APC) involved in T cell activation in response to malaria (McCall and Sauerwein, 2010; Stevenson and Riley, 2004).

Macrophages have also been implicated in the phagocytosis of IEs in the absence of malaria-specific antibody (Stevenson and Riley, 2004). Scavenger receptors including the class B receptor CD36 may have a role in opsonin-independent phagocytosis of IEs by monocytes in non-immune individuals (Serghides et al., 2003). Dendritic cells are central to both innate and adaptive immunity and express toll-like-receptors (TLRs) which play a role in both activating innate immunity and modulating adaptive immune response to microbial pathogens, including intracellular protozoan parasites (Stevenson and Riley, 2004). Their exact role in malaria is unclear, however TLR-mediated signals have been implicated in anti-parasite mechanisms (Stevenson and Riley, 2004). Investigations into DC activation by *P. falciparum* have had conflicting conclusions as to whether parasites interact with DCs to promote inflammatory response (Urban et al., 1999; Ocana-Morgner et al., 2003; Seixas et al., 2001). It has been suggested that an initial, transient period of conventional APC/DC activation may be followed by a period of down-regulated or absent pro-inflammatory signals, the relevance of which is unclear (Stevenson and Riley,

2004).

γδT cells bridge the responses between the innate and adaptive immune systems and have been reported to increase in number in response to infection with *P falciparum* (Hviid et al., 2001). The clinical relevance of this is unclear, however, they have been reported to have anti-parasitic properties (Jagannathan et al., 2014; Elloso et al., 1994). One suggestion is that the role of γδT cells is to activate further cells such as T cells and NK cells (Stevenson and Riley, 2004).

Natural Killer (NK) cells are often the first cells to respond to *in vitro* exposure of PBMCs to IEs with rapid IFN-γ production (Artavanis-Tsakonas and Riley, 2002). They are also known to be involved in controlling early parasitemia in murine models. The IFN-γ response is IL-12 dependent which suggests that the source is NK cells. NK cell activation in response to *P. falciparum* infection produces IFN-γ 24 - 48 hours before γδT cells and NKT cells respond, and activation of γδT cells and NKT cells correlates highly with the NK response suggesting that this response initiates a further cascade of innate responses (Stevenson and Riley, 2004; Artavanis-Tsakonas and Riley, 2002). However, high levels of IFN-γ are associated with pathology and clinical immunity is associated with decreased IFN-γ production (Riley, 1999).

The impact of the innate response on the outcome of infection by *P. falciparum* is unclear due to relatively few investigations. A rapid pro-inflammatory response might be able to control early parasitemia until the adaptive response is initiated. A potent pro-inflammatory response may also promote the development of severe malaria by overstimulating the adaptive response. In reality, it is possible that the innate response has the potential to be both initially beneficial and subsequently harmful unless modulated by the adaptive immune response (Stevenson and Riley, 2004).

Natural IgM and *Plasmodium falciparum*

The role of natural IgM in the innate immune response to malaria is unknown. However, given its exceptional ability to clear small apoptotic particles it has been suggested that natural IgM could be effective in the clearance of particle remnants produced upon erythrocyte rupture (Pleass et al., 2015).

In order to survive and establish infection, *P. falciparum* parasites have evolved highly effective immune evasion mechanisms and have found ways to use natural IgM to their own

benefit (Pleass et al., 2015). These will be discussed in Section (1.5) along with numerous other ingenious tactics.

1.4.3 Adaptive immunity and malaria

In malaria endemic regions, repeated exposure to *P. falciparum* infection over many years eventually results in relative immunity or "premunition". However, because of the slow development time malaria deaths are highest in children who have not yet developed immunity. The adaptive response of B cells and antibodies forms the basis of naturally acquired immunity to malaria.

Antibodies use a variety of mechanisms to protect against malaria both directly and in collaboration with effector cells. By binding to surface-expressed antigens, protective antibodies are able to directly block invasion of erythrocytes as well as blocking the release of merozoites from schizonts. Antibodies are also able to bind phagocytes via the Fc receptor, limiting parasite growth. Antibodies further promote parasite clearance by stimulating the splenic removal of IEs as well as initiating opsonisation, therefore enhancing phagocytic activity and complement-mediated clearance.

Immune responses to pre-erythrocytic stages are thought to have little involvement in adaptive immunity (McCall and Sauerwein, 2010). When the liver is bypassed and blood-stage parasites directly injected, immune individuals are still able to reduce parasitemia and avoid symptoms. In contrast, immunisation with whole irradiated sporozoites or pre-erythrocytic antigens can induce immunity (Kumar et al., 2006). During erythrocytic stages, there is more exposure to the parasite and a more extensive immune response is established. Potential targets of an immune response at this stage are parasite-encoded surface-expressed proteins revealed by free merozoites in the bloodstream as well as intra-erythocytic parasites. Protective immunity is thought to be largely mediated by IgG1 and IgG3 subclasses, as treatment of patients with IgG extracted from immune individuals reduces parasitemia and improves clinical symptoms. Parasite-specific IgG molecules target clonally variant surface antigens (VSA) expressed by the parasite in the blood stages (Hviid, 2005).

Little is known about the role of immune IgM in malaria, and although protective immunity is mainly associated with IgG, several studies have implicated IgM (Bolad et al., 2005; Brown et al., 1986; Boudin et al., 1993). Murine studies suggest that parasite-specific

IgM is a potent adjuvant (Couper et al., 2005; PG et al., 1982) while immune complexes containing IgM stimulate the development of acquired T-cell-mediated immunity (Pleass and Holder, 2005).

Transcriptional switching occurs between members of gene families such as the *Var* gene family. *Var* genes code for expression of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), a family of proteins located on the surface of an infected erythrocyte. Switching is a tactic employed by the parasite to avoid immune recognition by parasite-specific antibodies. This process, known as antigenic variation, gives rise to a wide range of proteins causing parasites with different antigenic and phenotypic characteristics to appear at different times within a population (Recker et al., 2011). In order to develop immunity to the parasite, the host has to build a large repertoire of antibodies recognising the range of antigens expressed by the parasite, a process which takes years and involves repeat exposure. The immunity developed is also non-sterile and a state of relative immunity is reached instead, where low levels of parasitemia are unlikely to cause clinical symptoms.

The parasite lifecycle is complex but the human host has a sophisticated immune system consisting of the early innate response and the deadly adaptive response. In order to be successful the parasite must remain undetected in the host and it has developed ingenious immune evasion mechanisms. Antigenic variation enables the parasite to slow down immune detection while it multiplies inside infected erythrocytes. The parasite has evolved further tactics to avoid clearance of the infected erythrocyte should it be identified. These must be understood in order to develop effective vaccines to tackle the disease burden.

1.5 *Plasmodium falciparum* Immune Evasion

The persistence of an infection by *Plasmodium falciparum* as well as the ability to cause recurrent infections is partly due to the parasite's superior ability to evade the host immune response. The parasite is vulnerable whilst moving between cells within the host, however the parasite moves very rapidly within the bloodstream and swiftly infects cells, removing itself from the bloodstream and out of danger from antibodies. Upon invasion of erythrocytes, the parasite undergoes developmental changes and consequently the erythrocyte becomes more rigid, knob-like structures form on the membrane and parasite-expressed proteins appear on the membrane surface, hence an infected erythrocyte is readily distinguishable from an uninfected erythrocyte.

The parasite has itself developed mechanisms to avoid recognition by the immune system. Antigenic variation in *P. falciparum* infection is one reason that the parasite is able to proliferate so successfully and has the ability to reinfect time after time before the host can develop immunity. The wide range of protein variants expressed at critical stages in the lifecycle keeps the parasite one step ahead of the host immune response. *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) is a highly-variant family of proteins encoded by the *var* genes. The *var* gene family consists of 60 *var* genes per parasite genome which code for PfEMP1. PfEMP1s are expressed on the surface of infected erythrocytes (IEs) following invasion by the merozoite. Transcriptional switching of this protein family slows immune recognition of the IE and allows the parasite time to develop before bursting out and infecting more erythrocytes. Susceptibility to malaria corresponds with gaps in an individual's repertoire of PfEMP1-specific Abs (Hviid, 2010). The long acquisition time of immunity is partly due to the intraclonal and interclonal variation of PfEMP1 proteins, as well as the parasite's ability to express multiple PfEMP1s at one time in a mutually exclusive manner and its ability to quickly switch expression among the different variants (Barfod et al., 2011). Moreover, a variant of PfEMP1 may only be expressed for a short period of time before this switching occurs, limiting its exposure to the immune system.

PfEMP1 are also highly cytoadherent which gives them further advantages over the immune response. Cytoadherence prevents parasite clearance and PfEMP1, a known virulence factor, contributes to the severity of the clinical disease. The sequestration of IEs within host tissues and organs and the coating of an IE with uninfected erythrocytes

(known as rosetting) prevents clearance of the parasite by the spleen and leads to severe malaria pathogenesis. These surface proteins also aid the parasite in avoiding further clearance of the IE by binding to the Fc of natural IgM, preventing specific binding from more destructive IgG molecules. Protein variants which bind natural IgM are also present in the Merozoite Surface Protein (MSP) family located on the merozoite surface. The variant of surface protein expressed therefore determines pathogenesis and clinical disease. Antigenic variation of cytoadherent surface proteins is a deadly combination and the reason that the parasite thrives within the human host. Clinical disease depends on the cytoadherent property of the variant expressed and the pathogenesis that it produces.

1.5.1 Sequestration of parasites

One characteristic of the PfEMP1 protein family is its cytoadhesive property, an attribute linked to clinical malaria symptoms (Hviid, 2010). PfEMP1 variants act as ligands, adhering to host receptors such as chondroitin sulphate A (CSA), cluster of differentiation 36 (CD36) and intercellular adhesion molecule-1 (ICAM-1) thus mediating tissue-specific sequestration of IEs (Barfod et al., 2011). Sequestration allows the parasite to avoid immune evasion by avoiding clearance by the spleen and leads to the accumulation of IEs in various tissues. Consequently, this pathogenesis leads to life-threatening inflammation and circulatory disturbances (Barfod et al., 2011). For example, the adherence of IEs to I-CAM1 has been linked to cerebral malaria, while adherence to CSA is associated with placental malaria.

1.5.2 Rosetting

Another phenotype arising from the adhesive nature of PfEMP1 is rosetting, which is the coating of *P. falciparum*-infected erythrocytes with uninfected erythrocytes. This is a virulent phenotype and has long been associated with the occurrence of severe malaria due to microvascular obstruction caused by high densities of rosettes (Juillerat et al., 2011). PfEMP1 has been identified as the rosetting ligand and numerous receptors on the erythrocyte have been implicated in binding (Chen et al., 1998; Mercereau-Puijalon et al., 2008).

1.5.3 Natural IgM Fc-binding

A less characterised adhesive property of *P. falciparum* is the ability of PfEMP1 proteins to bind the Fc portion of natural IgM (Czajkowsky et al., 2010). It has been discussed previously that in order to evade destruction mediated by the Fc region of an Ab, many pathogens have evolved Fc-binding proteins (Czajkowsky et al., 2010). The Fc portion of an antibody is implicated in effector functions through binding to Fc receptors on effector cells. Binding of the protein to the Fc portion interferes with effector functions such as phagocytosis and cytotoxicity by preventing pathogen-specific Abs from interacting with host Fc receptors. It has been shown that in the case of *P. falciparum*, the PfEMP1 protein VAR2CSA (involved in placental pathogenesis) binds to non-specific IgM through the Fc portion of the antibody. This masks the infected red blood cell and blocks subsequent binding of specific human monoclonal IgG, therefore evading the specific immune response.

PfEMP1 proteins are a well-recognised virulence factor for *P. falciparum*, known for their cytoadherent properties amongst variants whose expression can be switched to avoid recognition. Less characterised Fc-binding proteins are DBLMSP1 and DBLMSP2 which are expressed on the surface of the merozoite. These are part of the Merozoite Surface Protein 3 (MSP3) family of proteins and also have the ability to bind to the Fc portion of natural IgM (Crosnier et al., 2016). Interestingly, these proteins contain a Duffy-binding-like (DBL) domain, a highly conserved region which is also present in PfEMP1. These have been well characterised as binding domains in PfEMP1 and are not present in any other MSP3 proteins. DBL domains therefore have potential as vaccine candidates and are the focus of this report.

1.6 *Plasmodium falciparum* Vaccine Development

Malaria is the leading cause of mortality and morbidity worldwide and the malaria burden is particularly heavy in vulnerable communities in subsaharan Africa where *P. falciparum* is widespread. A malaria vaccine is urgently needed, however vaccine development is difficult due to the complexity of the parasite lifecycle which limits immune exposure as well as effective immune evasion mechanisms. The clinical outcome of malaria depends on many factors and consequently severe malaria is difficult to predict. However, key targets of natural immunity provide a starting point for vaccine development.

1.6.1 Current progress in vaccine development

In 2006, the World Health Organisation (WHO) set out a framework for vaccine development in the Malaria Vaccine Technology Roadmap, which was then updated in 2013 in response to changing epidemiological and control status (Malaria Vaccine Funders Group, 2013). The current Roadmap consists of a long and a short-term strategic goal. The long-term objective is to develop a vaccine by 2025 with a protective efficacy of >80%. The aim is for this vaccine to provide protection against clinical disease for longer than 4 years. A short-term landmark goal required a first-generation vaccine to be developed and licensed by 2015. In July 2015, following relative success of RTS,S in phase 3 clinical trial which showed a substantial reduction in clinical malaria over a 3 - 4 year period, marketing authorisation for RTS,S was submitted (Kaslow and Biernaux, 2015). RTS,S is a promising pre-erythrocytic vaccine candidate which targets circumsporozoite protein (CSP) which is expressed on the surface of the sporozoite and is essential for parasite development as well as invasion of mosquito and human host tissues (Mueller et al., 2010). This reduces the number of parasites in the blood by blocking invasion and inhibiting development of the parasite. The overall vaccine efficacy of RTS,S, however, falls short of the target 50% protective efficacy against severe disease lasting longer than one year (Kaslow and Biernaux, 2015).

The liver stages are essential in the initial development of the parasite within the human host, the pre-erythrocytic phase is therefore an excellent target for early intervention in order to prevent clinical disease and transmission. A sporozoite-based vaccine could also have the potential to limit infection in both the mosquito vector and the human host. The relatively short developmental period of 8 - 15 days and the low levels of parasitemia

mean that targets of natural immunity in the liver stages are limited. However, surface and secretory proteins expressed by the sporozoite may be essential for either parasite development or invasion of host tissues and therefore have potential as vaccine candidates.

Efforts to improve on the efficacy of RTS,S include more than twenty vaccine candidates in clinical trial (Ouattara and Laurens, 2015). The success of recent candidates has mostly been limited with few progressing past early stages. Recently, immunisation with radiation-attenuated sporozoites (RAS) has been shown to induce sterile immunity in patients and has raised hopes for another pre-erythrocytic vaccine candidate PfSPZ with excellent efficacy (Hoffman et al., 2002; Seder et al., 2013). PfSPZ uses irradiated sporozoites to invoke a natural immune response without causing clinical disease. One problem of PfSPZ for mass vaccination is the need to administer intravenously as a clinical trial given under the skin in 2011 reported disappointing results (Seder et al., 2013). The lack of infrastructure to support this requirement also poses further problems as the vaccine must be stored frozen in liquid nitrogen vapour phase. Although the vaccine efficacy of PfSPZ is promising, work must be done to find a way around the logistical challenges faced in order to harness the vaccine's potential.

Transmission blocking vaccines (TBV) where immunity is raised to mosquito stage antigens, are aimed at preventing the development of the parasite within the mosquito host. This prevents the parasite from being passed on from an immunised individual and infecting a further host. TBVs may be used alongside other control and prevention measures such as drainage and insecticides to reduce transmission rates, as well as to complement the use of human lifecycle stage vaccines. TBVs based on sexual stage parasite surface proteins such as Pfs25, Pfs28, Pfs30, Pfs48 have had limited success, proving difficult to manufacture and in some cases causing serious side effects (reviewed by Dinglasan et al. (2013)). The current leading TBV is based on the mosquito-stage antigen anopheline midgut analyl aminopeptidase N1 (AnAPN1), located on the surface of the mosquito midgut (Atkinson et al., 2015). Vaccination involves immunisation of the human host, acquired antibodies are then passed to the mosquito along with the parasite. Some, but not all of these antibodies have been shown to be effective at preventing the parasite from entering cells of the midgut wall. The antibody response therefore needs to be focussed and current vaccine development aims to identify the transmission-blocking region of the protein (Atkinson et al., 2015). Combined use of TBVs as well as human

lifecycle stage vaccines may be necessary to reach long term vaccine goals. By reducing transmission within a community, protective efficacy of other vaccine components may be improved and the spread of parasites resistant to such vectors may also be reduced.

1.6.2 Potential vaccine candidates

Further surface and secreted sporozoite proteins also have potential as vaccine candidates, such as thrombospondin-related anonymous protein (TRAP) which has a role in sporozoite motility and the invasion of liver cells (Akhouri et al., 2008).

Pre-erythrocytic stage vaccines have the potential to make a significant public health impact. In order to meet long-term goals, however, a second-generation vaccine must be developed in addition to these. Another approach is to target the blood stage of the infection where parasites become symptomatic and cause clinical disease.

Clinical symptoms correspond with invasion of erythrocytes by merozoites. One tactic for vaccine development is to block invasion and prevent disease progression and clinical symptoms. RH5 of the reticulocyte binding protein-like homologue family is a critical mediator of erythrocyte invasion through its interaction with basigin (Ord et al., 2015). Antibodies raised against RH5 have shown inhibition of erythrocyte invasion in several studies (Douglas et al., 2011; Ord et al., 2012; Reddy et al., 2014) showing RH5's promise as a vaccine candidate (Ord et al., 2015).

It is during the blood stages, where the parasite is repeatedly exposed to the immune system, that the main targets of natural immunity have been identified. Proteins exposed to the immune system on the surface of both the merozoite as well as the merozoite-infected erythrocyte induce a natural immune response and are therefore ideal vaccine candidates.

The main cause of severe disease is thought to be due to the cytoadherent property of proteins expressed on the surface of infected erythrocytes. Vaccines blocking cytoadherence would have a therapeutic effect and significantly reduce severity of disease, preventing further deterioration in patients exhibiting the first symptoms of clinical disease. PfEMP1 is exposed on the surface of infected erythrocytes and is a well-known target of natural immunity. PfEMP1 variants have been identified as a virulence factor, the pathogenesis of which relies on its cytoadherent property (Miller et al., 2002). A PfEMP1-based cytoadherence-blocking vaccine may therefore improve clinical symptoms and reduce disease severity.

DBL domains are conserved binding domains present in PfEMP1 proteins on the erythrocyte surface. There is hope that DBL domain fragments may be identified with immunogenicity for inducing anti-adhesion antibodies. Two fragments from the PfEMP1 variant VAR2CSA which have been found to have high affinity for CSA are expected to enter clinical studies in the near future (Fried and Duffy, 2015). These fragments show promise for a vaccine giving protection against placental malaria which is a major cause of maternal, fetal and infant mortality. However, susceptibility to malaria corresponds with gaps in an individual's repertoire of PfEMP1-specific Abs (Hviid, 2010). Due to a large number of PfEMP1 alleles responsible for a range of cytoadherent properties, targeting PfEMP1 directly may have limited success as a single vaccine may only be protective against a single variant. Although polymorphic, DBL domains can be categorised into groups based on sequence homology blocks. If a common binding region can be identified across the catalogue of PfEMP1 DBL domains, there may be potential to develop a single PfEMP1 based vaccine with immunogenicity to all cytoadherent, virulent alleles.

Conserved binding regions in PfEMP1, called DBL domains, are also present in two merozoite surface proteins from the MSP3 protein family. MSP3 is a multi-gene family of proteins which have long been associated with clinical immunity. IgG3 antibodies to the highly conserved C-terminal region have been shown to be associated with reduced incidence of malaria (Polley et al., 2007). The presence of DBL domains in these MSP3-family proteins implies functional similarities with PfEMP1 as well as with parasite invasion ligands such as EBA-175 (Chiu et al., 2015). MSPDBL proteins have recently been shown to induce a powerful immune response. Antibodies to MSPDBL1 have been shown to inhibit parasite growth *in vitro* while antibodies specific to the DBL domain of MSPDBL2 have been associated with reduced risk of clinical malaria. These antibodies have also been shown to opsonise merozoites for FcγR-mediated phagocytosis by monocytes (Chiu et al., 2015). MSPDBL1 and MSPDBL2 have also been found to bind unknown receptors on the erythrocyte surface through the DBL domain, which suggests that they may be involved in erythrocyte invasion (Chiu et al., 2015). A target of protective antibodies, MSPDBL1 and MSPDBL2 have potential for vaccine development as an MSPDBL-based vaccine may act to prevent erythrocyte invasion.

A further member of the MSP3-family of proteins, MSP3.1 has also been identified as a target of protective Abs against *P. falciparum* by screening a genomic expression library

using a method called Antibody Dependent Cellular Inhibition (ADCI) (Druilhe et al., 2005). As well as being shown to induce anti-parasitic effects in the presence of monocytes both *in vitro* and *in vivo*, MSP3.1 is associated with protective immunity as well as an improved prognosis of drug-treated cerebral malaria (Demanga et al., 2010; Druilhe et al., 2005; Sirima et al., 2011). The MSP3-family also share a NLRNA/NLRNG N-terminal signature signal sequence. Unlike the PfEMP1 family of proteins, MSP3 proteins are simultaneously expressed on the merozoite surface. As well as showing sequence homology between members of the family, they are also highly conserved between different parasite isolates. MSP3.1 shows sequence conservation in the C-terminal region with many field isolates from Africa and Asia, and has shown significant promise in clinical trials (Singh et al., 2009). Moreover, antibodies raised both naturally and by immunisation to a member of the MSP3-family show cross-reactivity with other members of the family (Singh et al., 2009; Demanga et al., 2010) further indicating the potential of this protein family in vaccine development.

MSPDBL1 and MSPDBL2 proteins, from the MSP3 protein family expressed on the merozoite surface, have recently been identified as candidates for vaccine development. Thought to be involved in erythrocyte invasion, interestingly they are the only proteins expressed on the merozoite surface to contain a DBL domain, a structurally conserved binding region also present in PfEMP1 and known to be responsible for its cytoadherent property. Due to the difficulties faced by antigenic variation and variant switching in PfEMP1 proteins, these proteins may be a better candidate in the race to develop an erythrocytic stage vaccine and meet long-term objectives. A MSPDBL-based vaccine able to prevent invasion of erythrocytes would limit parasitemia, slowing disease progression and reducing severity of disease. These proteins require more attention and are a particular focus of this report.

1.7 MSPDBL1 and MSPDBL2

The potential of MSPDBL1 and MSPDBL2 as vaccine candidates has only recently been realised, as it has been suggested that naturally acquired antibodies specific to MSPDBL2 contribute to protective immunity (Tetteh et al., 2013; Chiu et al., 2015). MSPDBL1 and MSPDBL2 are interesting proteins that differ in structure from the rest of the MSP protein family in that they contain a DBL domain. This domain is responsible for the cytoadherence of PfEMP1 proteins on the surface of the erythrocyte, is similar to the *Plasmodium vivax* Duffy binding protein, and is found in other *P. falciparum* invasion ligands such as Erythocyte Binding Antigen-175 (EBA-175), EBA-140 and EBA-181 (Chiu et al., 2015). MSPDBL1 has been shown to inhibit merozoite invasion into erythrocytes *in vitro* (Sakamoto et al., 2012) and more recently, the MSP1 complex has been shown to act as a platform for the display of MSPDBL1 and MSPDBL2 on the merozoite surface for binding to erythrocyte receptors (Lin et al., 2014). MSPDBL1 and MSPDBL2 appear to be involved in the initial attachment of the merozoite to erythrocytes during invasion, a critical stage in the *P. falciparum* lifecycle.

MSPDBL1 and MSPDBL2 bind to the Fc of natural IgM via their DBL domains (Crosnier et al., 2016). This is a binding characteristic seen in many pathogens and is thought to prevent recognition of the parasite by the more specific IgG antibody. This mechanism needs to be further understood in order to develop an effective invasion-blocking vaccine. Furthermore, the IgM binding property could be used to control IgM-mediated diseases and complement activation. DBL-Fc fusion protein therapeutics which bind IgM with high affinity could be used as therapeutics to prevent diseases such as IgM-associated peripheral neuropathies, to reduce systemic deposition of monoclonal IgM and its associated symptoms, or reduce complement activation in conditions such as type II cryoglobulinemia.

1.7.1 Duffy-Binding-Like protein domains

Modifications to both the merozoite and erythrocyte membrane surfaces by the MSP and PfEMP1 protein families are key stages in disease progression, targets for natural immunity and therefore targets for vaccine development.

Duffy-Binding-Like (DBL) protein domains, named after the Duffy binding proteins in *P. knowlesi* and *P. vivax* which bind the Duffy antigen receptor for chemokines (DARC),

are conserved domains common to both MSPDBL1, MSPDBL2 and PfEMP1 and are believed to be essential to their function (Hodder et al., 2012; Wickramarachchi et al., 2009). DBL domains are cysteine-rich modules, found in var and EBL genes and are classified into six types (α, β, γ, δ, ϵ, and ξ). Although conserved within classes of homology blocks, DBL domains have a diverse function and there is great variation in the host receptors with which they interact (Rask et al., 2010; Hodder et al., 2012). MSPDBL1 and MSPDBL2 both have a single, highly polymorphic, DBL domain while multiple DBL domains are present in PfEMP1 (Wickramarachchi et al., 2009).

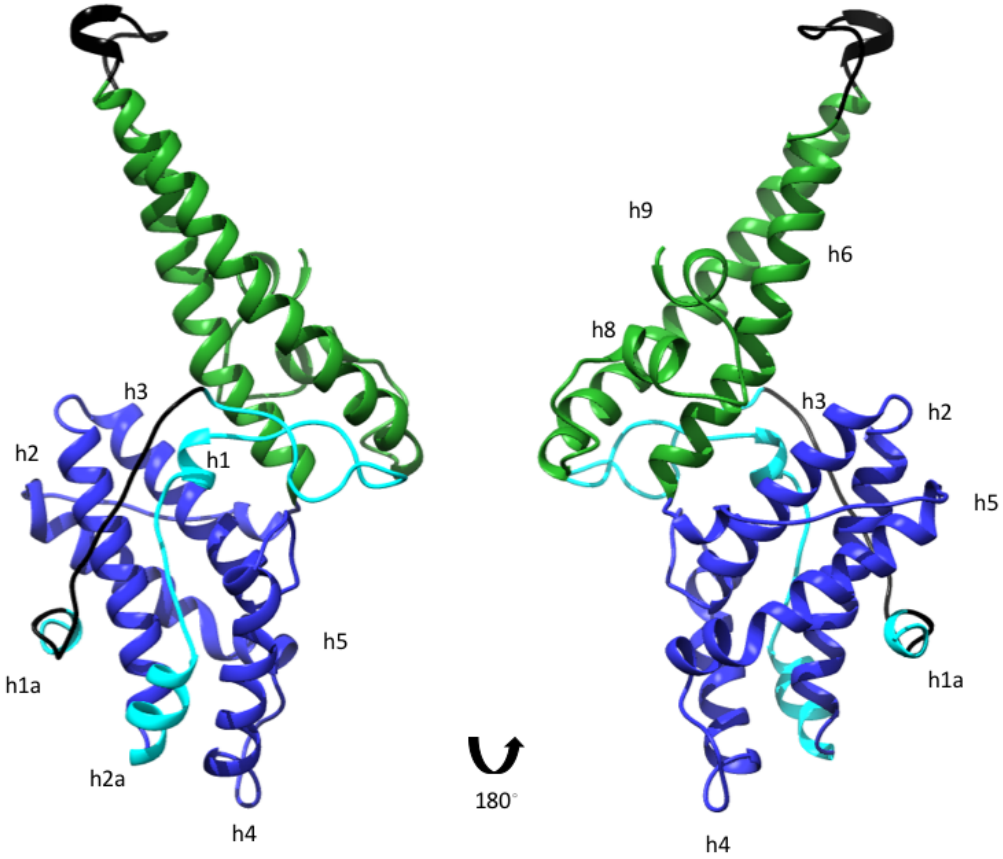


Figure 1.5: Ribbon diagram showing the structure of MSPDBL2. Helices are labelled h1-h9 and regions corresponding to subdomains 1-3 are coloured cyan, blue and green, respectively. Only electron density corresponding to Cys-177 was observed in the region between residues Ser-172 and Asn-185 and no electron density was observed for the loop (Lys-375, Val-387) between helices 5 and 6 (Hodder et al., 2012). These regions were modelled using SWISS-MODEL, and are coloured in black (Arnold et al., 2006). (Figure adapted from Hodder et al. (2012))

Both DBL domains have recently been expressed and shown to be functional as erythrocyte-binding domains with similar properties to the full-length proteins (Hodder et al., 2012). Furthermore, the crystal structure of MSPDBL2 has been solved and from this, the structure of MSPDBL1 modelled. The crystal structure reveals a canonical DBL fold consisting of a boomerang-shaped α-helical core formed from three subdomains (Hodder et al., 2012) (Figure 1.5).

Although DBL variants have high sequence diversity, their structural similarity suggests that structure is critical for cytoadherence. This binding property and their involvement in erythrocyte invasion makes them of particular interest in understanding the virulent nature of *P. falciparum* and in order to develop effective vaccines to reduce the parasite burden and lower mortality and morbidity rates particularly in sub-saharan Africa.

1.8 Summary

Antibodies form a central part of the immune response to a pathogen, identifying the threat and recruiting effector cells in order to neutralise and destroy the invading organism. Effector cells are recruited through binding of the antigen-bound Ab to FcRs present on the cell surface. An antibody consists of two regions, the Fab and Fc. The Fab arms of the 'Y' shaped Ab molecule is antigen specific whereas the constant Fc region binds to the FcR on effector cells.

Monoclonal antibodies have dominated the field of therapeutics in recent years, although a relatively new range of therapeutics harnesses the interaction of the IgG Fc with the neonatal receptor (FcRn) in order to improve the half-life of a fused immunogenic protein. Fc-fusion proteins are proving to be successful therapeutics, with numerous Fc-fusions reaching FDA approval in recent years. Recently, engineering of the Fc-scaffold has opened up the use of Fc-fusions in vaccines where cross-linking of multiple receptors is important in order to prime the immune response. The Hexa-Fc scaffold is a hexameric Fc-fusion scaffold which shows improved binding to FcRs and may enable delivery of twelve fused proteins.

The pressing need for new vaccine development in malaria in order to meet long-term targets is clear, with limited vaccine candidates progressing past phase two clinical trials. Of the most promising candidates, RTS,S has showed limited efficacy and the highly effective PfSPZ has practical limitations for use in the field. Clinical symptoms correspond with the blood stages of infection and it is anticipated that a new blood stage vaccine may prove effective in targeting clinical disease. In the blood stages of infection, immunogenic proteins are present on the surface of both infected erythrocytes and merozoites. Cytoadherence of infected erythrocytes in various tissues is associated with virulence, and conserved regions called Duffy-Binding-Like (DBL) domains have a pivotal role in cytoadherence of the deadly *P. falciparum*.

Two merozoite surface proteins, MSPDBL1 and MSPDBL2, are the only MSPs to contain DBL domains. These domains, like those on the surface of erythrocytes, are cytoadherent and have recently been shown to bind the Fc region of natural IgM. These are the only DBL domains present on the merozoite surface and identification of binding sites present in these domains may lead to identification of residues critical to IgM-binding in DBL domains of the surface of infected erythrocytes. Critical binding regions may

be used to raise binding-blocking antibodies while the DBL-Fc fusion proteins may be developed for therapeutic use for treatment of IgM-mediated conditions.

1.9 Aims

The two merozoite DBL domains will be expressed as Fc-fusion proteins for two reasons:

1. To further investigate their IgM-binding properties.

2. As novel DBL-Fc fusion proteins which can be further developed for vaccine use.

The central aim of this project is to produce Fc-fusion proteins from the *P. falciparum* merozoite surface proteins MSPDBL1 and MSPDBL2. Recombinant protein will then be used to investigate the binding properties of the DBL domains, two blood stage vaccine candidates, and to determine residues and structural properties critical for binding. This is important because it would allow us to formulate with improved accuracy the properties of these DBL domains that are a potential target for vaccination.

In order to achieve these aims, methods will be developed to express the DBL domains in a mammalian expression system. The DBL domains will be fused to an IgG1-Fc domain which, while convenient for protein purification, will provide proof-of-concept for an Fc-fusion protein based on these DBL domains. Such a protein with increased plasma half-life will have prolonged therapeutic activity.

In order to identify residues critical for binding, a library of DBL mutants from MSPDBL1 and -2 will be produced and their IgM-binding properties investigated. A comprehensive analysis of the structural and sequential similarities between MSPDBL1 and MSPDBL2 DBL domains along with eight other known IgM-binding DBL domains from PfEMP1 in contrast with fifteen non-binding DBL domains will provide further insight into the IgM-binding property.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Oligonucleotides

Oligonucleotides were ordered lyophilised from Eurofins (UK) and stored as 100 μM stock. Polymerase chain reaction (PCR) and Mutagenesis primers (Tables 2.1 and 2.3) were used at 10 μM working dilution, and sequencing primers (Table 2.2) were sent to Source Bioscience at concentration of 3.2 pmol/μl.

Table 2.1: PCR primers for amplification of MSP3.4 and MSP3.8 DBL domains.

Primer	Sequence	Annealing Temperature (°C)
3.4 Forward	atc ggc cat ggt ttt ctg caa ggg cat caa	68.1
3.4 Reverse	gtt ttg tca gat ctg ttg tcc agg tcc	65.0
3.8 Forward	cgg aat tcg atc tgc aag gac ttc agc	61.0
3.8 Reverse	ttg tca gat cta acc atg gcg tcc tgc a	66.6

Table 2.2: Sequencing primers

Primer	Sequence	Annealing Temperature (°C)
M13 F	tgt aaa acg acg gcc agt	48.0
M13 R	cag gaa aca gct atg acc	48.0
3.4MutF	gcc gtt aca gat cca agc tg	54.0°C
3.4MutR	cca cgc atg tga cct cag	55.0°C
Fc mut-1:5'	acc ctg ctt gct caa ctc t	57°C
Fc mut-1:3'	tgg ttt gtc caa act cat caa	57°C

Table 2.3: Mutagenesis primers

Primer	Sequence	Annealing Temperature (°C)
3.4 Forward	atc ggc cat ggt ttt ctg caa ggg cat caa	68.1°C
3.4 Reverse	gtt ttg tca gat ctg ttg tcc agg tcc	65.0°C
3.8 Forward	tcg cgg cgc aat tcg atc tgc aag gac	69.5°C
3.8 Reverse	cgg aat tcg atc tgc aag gac tta gc	66.6°C

2.1.2 Plasmids

Commercially available plasmids were used for cloning (Table 2.4). See Appendix A for plasmid maps and sequences.

Table 2.4: Commercial plasmids used for cloning

Plasmid	Features	Supplier
pFUSE-hIgG1-Fc2	hIgG1-Fc (human), hEF1-HTLV prom, IL2 ss, MCS, SV40 pAn, ori, CMV enh / hFerL prom, EM2KC, Zeo, βGlo pAn	Invivogen
pCR2.1-TOPO	LacZα fragment, M13 Reverse priming site, Multiple cloning site, T7 promoter/priming site, M13 Forward priming site, f1 origin, Kanamycin resistance, Ampicillin resistance, pUC origin	Life Technologies

2.1.3 Competent Cells

The following competent cells were used for transformations.

Table 2.5: Commercial competent cells

Name	Transformation Efficiency (transformants/µg control DNA per 50 µL reaction)	Supplier
One Shot TOP10	$> 1x10^9$	Life Technologies
Subcloning Efficiency Dh5α	$> 1x10^6$	Life Technologies
XL1-blue	$> 1x10^8$	Agilent

2.1.4 Antibodies

The following antibodies were used for Western blots, immunoblots and Enzyme linked immunosorbent assays (ELISAs).

Table 2.6: Primary antibodies

Antibody	Working Dilution	Supplier	Catalogue No
hIgM	5 µg/ml	Sigma	I8260
Gammagard IVIG	1 mg/ml	Baxter	0944-2656-03

Table 2.7: Antibodies for Westerns and Immunoblots

Antibody	Host	Dilution	Supplier	Cat No
αhIgG (Fc specific) AP	Goat	1:5000	Sigma	A9544
αrbIgG (whole molecule) AP	Goat	1:15,000	Sigma	A9919
αrbIgG (whole molecule) AP	Goat	1:15,000	Sigma	A3687
αmsIgG AP	Goat	1:10,000	Thermo Scientific	31322
Polyclonal αMSP3.4			Sanger (UK)	
Polyclonal αMSP3.8			Sanger (UK)	
Monoclonal αMSP3.4			Sanger (UK)	
Monoclonal αMSP3.8			Sanger (UK)	

Table 2.8: Antibodies for ELISAs

Antibody	Host	Dilution	Supplier	Cat No
αhIgG (Fc specific) AP	Goat	1:5000	Sigma	A9544
αhIgM (μ-chain specific) AP	Goat	1:5000	Sigma	A9794
αhIgM (μ-chain specific)	Mouse	1:5000	Sigma	I6385
αhIgM (μ-chain specific)	Goat	1:5000	Sigma	I2386
αhIgG (Fc specific)	Goat	1:1000	Life Technologies	H10000
Monoclonal αhIgG (Fc specific)	Mouse	1:1000	Sigma	I6260
αhIgG (Fc specific) HRP	Goat	1:500	Sigma	A0170

2.1.5 Kits

The following commercials kits were used according to manufacturer's instructions.

Table 2.9: Commercial Kits

Kit	Supplier	Cat No
High Pure PCR cleanup Micro Kit	Roche	04983955001
QIAprep Spin Mini prep Kit	Qiagen	27106
QIAGEN Plasmid Maxi Kit	Qiagen	12165
QIAGEN Plasmid Midi Kit	Qiagen	12145
Genemorph II Random Mutagenesis Kit	Agilent Technologies	200550

2.1.6 Reagents

The following reagents were used according to manufacturer's instructions.

Table 2.10: PCR and ligation reagents

Reagent	Supplier	Cat No
rSAP and CutSmart buffer	NEB	M0371S
Antarctic phosphatase and standard buffer	NEB	M0289
Taq polymerase and standard Taq buffer	NEB	M0273
Phusion High Fidelity DNA polymerase	NEB	M0530
dNTPs	NEB	N0447
T4 DNA ligase and buffer	NEB	M0202

Table 2.11: DNA and protein standards and dyes

Standard	Supplier	Cat No
Gel filtration standard	Bio-Rad	151-1901
SeeBlue Pre-stained Protein Standard	Life Technologies	LC5625
1kbp DNA gel ladder	NEB	N3232
Gel loading dye (6x), Blue	NEB	B7021S
Ethidium bromide	Sigma	E7637
Coomassie Brilliant Blue R-250	Bio-Rad	161-0436
Coomassie Brilliant Blue R-250 Destaining	Bio-Rad	161-0438

Table 2.12: Substrates for ELISA, Western blot and Immunoblot

Name	Supplier	Cat No
SigmaFast p-NPP tablets	Sigma	N2770
SigmaFast BCIP/NBT	Sigma	B5655
5'5'3'3' Tetramethylbenzidine dihyrochloride	Sigma	T3405

Table 2.13: Transformation and transfection reagents

Name	Supplier	Cat No
FuGene	Promega	E2311
IPTG	Life Technologies	15529-019
X-Gal	Sigma	B4252

2.1.7 Restriction Enzymes and Buffers

Restriction enzymes were used according to manufacturer's instructions alongside their recommended buffers.

Table 2.14: Restriction Enzymes and corresponding buffers

Enzyme	Restriction Site	Buffer	Supplier	Cat No
NcoI	5' C$^{\downarrow}$CATGG 3' 3' GGTAC$_{\uparrow}$C 5'	NEBuffer 3.1	NEB	R0193
BglII	5' A$^{\downarrow}$GATCT 3' 3' TCTAG$_{\uparrow}$A 5'	NEBuffer 3.1	NEB	R0144
EcoRI	5' G$^{\downarrow}$AATTC 3' 3' CTTAA$_{\uparrow}$G 5'	NEBuffer EcoRI	NEB	R0101
EcoRV	5' GAT$^{\downarrow}$ATC 3' 3' CTA$_{\uparrow}$TAG 5'	NEBuffer 3.1	NEB	R0195

2.1.8 General Buffers

- 10x Phosphate Buffered Saline (PBS) contains 1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4.
- Phosphate Buffered Saline + 0.05% Tween (PBST).
- Blocking buffer consists of PBST + 5% dried milk powder.

2.1.9 Gel Electrophoresis Buffers

- Tris-borate-EDTA (TBE) buffer contains 89 mM Tris base, 89 mM boric acid, 2 mM EDTA.
- Tris-EDTA (TE) buffer contains 10 mM Tris-HCl pH 8 + 1 mM EDTA.
- Loading dye contains 0.25% bromophenol blue w/v, 0.25% xylene cyanol w/v, 50% glycerol, 1% Tris-HCl pH 8.

2.1.10 ELISA Buffers

- 0.5 M carbonate-bicarbonate buffer; dissolve one capsule in 100 ml deionised water (Sigma, C3041). This is used as a coating buffer for ELISA plates.

- 0.05 M phosphate-citrate buffer; dissolve one capsule in 100 ml deionised water (Sigma, P4809). This is used to develop the substrate 5'5'3'3' tetremethylbenzidine dihydrochloride according to manufacturer's instructions.

2.1.11 SDS-Page Electrophoresis Buffers

- MOPS-SDS running buffer (50 mM MOPS, 50 mM Tris base, 0.1% SDS, 1 mM EDTA, pH 7.7), bought as a 20x stock (Life Technologies, NP0001).

2.1.12 Western Blot Buffers

- Transfer buffer (25 mM bicine, 25 mM bis-tris, 1 mM EDTA pH 7.2, 20% methanol), bought as a 20x stock (Life Technologies, NP0006).

2.1.13 Fast protein liquid chromatography (FPLC) Buffers

All FPLC buffers were filtered using Steritop GP Sterlisation Unit 0.22 μm PES membrane (Millipore).

- 20 mM sodium phosphate, pH 7 (mix 20 mM NaH_2PO_4 (monobasic) and 20 mMNa_2HPO_4 (dibasic) stock solutions to the required pH).
- Hank's Balanced Salt Solution (HBSS) (Life Technologies, 14185045).
- 0.1M Glycine - HCl, pH 2.5 - 2.7.

2.1.14 Antibiotics

- zeocin (Invivogen, ant-zn-ip).
- ampicillin (Sigma, A0166).

2.1.15 Cell culture media

- Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies) plus 10% ultra-low IgG foetal bovine serum (Life Technologies), 1% L-glutamine (Life Techologies), 1% penicillin/streptomycin (Life Technologies).
- Ham's F10 medium (Life Technologies) plus 10% ultra-low IgG foetal bovine serum (Life Technologies), 1% penicillin/streptomycin (Life Technologies).

- DMEM or Ham's F10 selection media were supplemented with zeocin at 400 μgml^{-1}.

2.1.16 Microbiological media

- Terrific Broth (TB) containing enzymatic casein digest 12 mgml^{-1}, yeast extract 24 mgml^{-1}, K_2HPO_4 9.4 mgml^{-1}, KH_2PO_4 2.2 mgml^{-1}, glycerol 10 mgml^{-1}
- Lysogeny Broth (LB) containing bacto-tryptone 10 mgml^{-1}, bacto-yeast extract 5 mgml^{-1}, NaCl 10 mgml^{-1}.
- For agarose plates, the media were supplemented with bacto-agar 15 mgml^{-1}.
- TB or LB selection media were supplemented with ampicillin 50 μgml^{-1} or zeocin 25 μgml^{-1}.

2.1.17 Software

Molecular visualisation software

- PyMOL v 1.7.4.5 Edu Enhanced for Mac OS X
- Chimera v 8.6.1
- MODELLER v 9.16

Software for analysis of DNA

- 4Peaks sequence viewer
- EnzymeX
- Sequence Manipulation Suite: bioinformatics.org
- BLAST (NCBI)
- Primer3plus (Andreas Untergasser, Harm Nijveen, Xiangyu Rao, Ton Bisseling, Ren Geurts, and Jack A.M. Leunissen: Primer3Plus, an enhanced web interface to Primer3 Nucleic Acids Research 2007 35: W71-W74; doi:10.1093/nar/gkm306)
- MEGA v 6

Plasmid map design

- Savvy (Bioinformatics.org, Scalable Vector Graphics Plasmid Map)

2.2 Molecular Biology Methods

2.2.1 Primers

All primers were synthesised by Eurofins UK. Stock solutions of 100 µM were stored long-term and PCR reactions composed from a working solution of 10 µM.

When designing primers, the base composition was considered as well as the primer length and melting temperature. Primer pairs were designed with melting temperatures (T_m) within 5°C of each other. Primers were also checked to avoid self-complementarity and excessive repeats and runs.

2.2.2 PCR Protocol

PCR amplifications were performed in 25 µl volumes. When larger volume reactions were needed, a larger volume reaction mix was made and this split into multiple tubes with total 25 µl volume in each before amplification in the thermocycler. For each reaction, the appropriate polymerase and buffer was used, with 1mM deoxyribonucleotide triphosphates (40mM dNTPs mix) and the reaction performed according to standard procedures using a Progene thermocycler (Techne, Cambridge, UK).

2.2.3 Purification of PCR product from solution

In order to remove salts, primers, unincorporated nucleotides, and DNA polymerase, PCR products were routinely purified. This was done using one of two methods:

1. High Pure PCR cleanup micro kit (Table 2.9) was used according to the manufacturer's instructions. This involved adding a binding buffer which contains the chaotropic salt guanidine thiocyanate and enables binding of DNA of over 100 bases in length to glass fibres packed into a filter tube. Bound DNA was then washed and purified using a series of wash and spin steps before being eluted by a low salt elution buffer.

2. Ethanol precipitation. 1/10 volume of sodium acetate pH 5.2 was added to the sample followed by 2.5 volumes of cold 100% ethanol and the solution mixed well. This was then incubated at −20°C for a minimum of 20 minutes to precipitate the DNA. The DNA was pelleted by spinning in a tabletop centrifuge at maximum speed for 15 minutes and the supernatant decanted carefully. The DNA was washed by

adding 1 ml 70% ethanol, mixing well and spinning briefly before decanting the supernatant. The pellet was then air-dried and resuspended in dH_2O.

2.2.4 Restriction Digest

Plasmids and inserts were digested using restriction enzymes (NEB) in order to cut out compatible sticky ends ready for the ligation step. All digests were performed according to the manufacturer's instructions.

2.2.5 DNA gel electrophoresis

DNA gel electrophoresis was routinely done using on 0.8 - 1% TBE agarose gels (UltraPure, Invitrogen) (Section 2.1.9) containing 40 µg/µl Ethidium Bromide. Samples were loaded in Bromophenol blue dye (Section 2.1.9) and a 1 kbp ladder from NEB (Figure 2.1) used as a molecular marker. The gel was then electrophoresed in TBE running buffer at 80 V. The DNA bands were analysed using a UV illuminator.

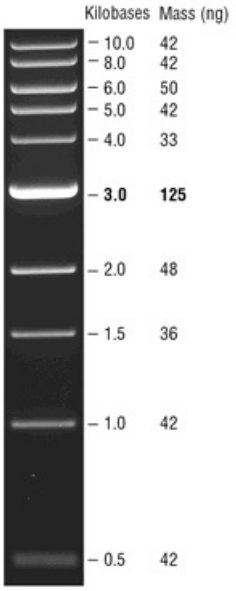


Figure 2.1: NEB 1 kb ladder used to size DNA fragments on agarose gel.

2.2.6 Extraction of DNA from agarose gel

DNA was routinely extracted from agarose gel using the High Pure PCR cleanup micro kit (Table 2.9). This was done according to the manufacturer's instructions and involved cutting out the DNA band, and heating it at 56°C in binding buffer to dissolve the gel. Binding buffer contains the chaotropic salt guanidine thiocyanate and enables binding of

DNA of over 100 bases in length to glass fibres packed into a filter tube. Bound DNA was then washed and purified using a series of wash and spin steps before being eluted by a low salt elution buffer.

2.2.7 Vector de-phosphorylation

In order to prevent vector DNA from self-ligating it was dephosphorylated. This was done using either Antarctic phosphatase or Shrimp Alkaline phosphatase (rSAP) according to manufacturer's instructions.

2.2.8 Ligation Protocol

For ligation reactions typically 90 ng of vector and 60 ng of insert were used in line with the recommended vector:insert ration of 1:3. T4 DNA ligase and T4 DNA ligase buffer (Invitrogen) were used according to manufacturer's instructions. Ligation reactions were held at 16° C for 16 hours using a Progene thermocycler.

2.2.9 Making competent *E.coli* cells

50 ml of LB medium (Section 2.1.16) was inoculated using glycerol stock of *E.coli* cells (XL1-blue, Dh5α or TOP10). These were grown at 37°C in an orbital shaker set to 225 rpm until an OD of 0.6 - 0.8 was achieved. Bacterial cultures were spun at 2,600 xg for 5 mins in a chilled (4°C) centrifuge and the pellet was resuspended in 25 ml of 100 mM ice-cold $MgCl_2$. Cells were spun at 2,600 xg (4°C) for 5 mins and the supernatant discarded. The pellet was resuspended in 25 ml of 100 mM ice-cold $CaCl_2$ and the cell suspension was incubated on ice for 60 mins. The suspension was then spun at 2,600 xg (4°C) for 5 mins and the supernatant discarded. The pellet was resuspended in 5 ml ice-cold 85 mM $CaCl_2$ containing 15% glycerol. 100 µl aliquots were stored at −80°C.

2.2.10 Transformation of competent *E.coli* cells

Competent *E.coli* cells were defrosted on ice and 1–5 µl of the ligated construct added and the mixture incubated on ice for 30 mins. The cells were then given a heat shock of 42°C for 30 seconds in order to create pores in the cell membranes. The cells were then returned to ice for 2 mins. 200 – 500 µl of SOC media was added and then the cells shaken at 225 rpm, 37°C for 1 hour.

5–200 µl of the transformation reaction was spread onto pre-warmed agarose plates containing the appropriate antibiotic for selection and plates incubated at 37°C overnight for colonies to form.

2.2.11 Mini-prep

Plasmid DNA was routinely purified using the QIAprep Spin Miniprep kit (Table 2.9). A single bacterial colony containing the plasmid DNA was grown in 5 ml media containing the appropriate antibiotic required for selection (see Section 2.1.16 for bacterial growth and selection media routinely used) for 12-16 hours at 37°C with vigorous shaking (225 rpm). The bacteria was harvested and the plasmid DNA purified according to the manufacturer's instructions. This involved the preparation and clearance of lysate, followed by purification on a silica membrane and finally elution of the DNA from the membrane. First, bacterial cells were pelleted by centrifugation and resuspended in buffer containing RNase. The cells were lysed in NaOH/SDS buffer in order to release the cell contents whilst at the same time, denaturing chromosomal and plasmid DNA as well as proteins. The product was then neutralised and adjusted to high-salt concentration. This causes denatured proteins, chromosomal DNA, cellular debris, and SDS to precipitate, while renaturing the smaller plasmid DNA which stayed in solution. The solution was then passed across a silica membrane to which the plasmid DNA bound. Bound DNA was then washed and purified using a series of wash and spin steps before being eluted by a low salt elution buffer. The concentration of the purified plasmid DNA was then established using a nanodrop spectro-photometer (Section 2.2.15).

2.2.12 Screening Positive Colonies

Blue-white screening

For TOPO transformation, blue-white screening was used to select positive colonies. This works via a mutant form of the *E. coli* enzyme β-galactosidase that has its N-terminal residues deleted (the ω-peptide) and is inactive. The enzyme can be restored to it's active state by the presence of an α-peptide. When a vector containing this α-peptide is expressed by a cell containing the ω-peptide, they form a functional β-galactosidase enzyme. This enzyme hydrolyses the chromogenic substrate X-gal, which subsequently confers a blue colour on the bacterial colony.

This process is interrupted if DNA is inserted into the plasmid LacZα gene, stopping production of the α-peptide. Consequently, *E. coli* that contain the vector plus insert can't produce a functional β-galactosidase and the colony will appear white.

Using this method, white colonies were picked for screening and grown overnight in media containing 50 µg/ml ampicillin. Plasmid DNA was then isolated by mini-prep (Section 2.2.11).

Restriction Digest screening

For TOPO transformation, once the white colonies had been selected and the DNA isolated by mini-prep, plasmid DNA was then digested using restriction enzymes to identify whether the insert was present.

For pFUSE and Mutazyme transformations, all large colonies were selected and plasmid DNA isolated by mini-prep. This plasmid DNA was then digested using restriction enzymes to identify whether the insert was present.

The restriction digest solution consisted of the following:

- Restriction enzyme 1 — 1 µl
- Restriction enzyme 2 — 1 µl
- Compatible Buffer — 5 µl
- BSA (10x) — 5 µl
- Plasmid DNA — 1 µg
- dH_2O — up to 50 µl

Restriction digests were incubated at 37°C for 2 hours and then the product visualised by gel electrophoresis.

2.2.13 Sequencing

Plasmid DNA was sequenced using Sanger sequencing (Source Bioscience, Nottingham, UK). Sequences were then analysed using 4Peaks and the Basic Local Alignment Search Tools (BLAST) available at `http://blast.ncbi.nlm.nih.gov/Blast.cgi` as well as the 6 frame translation tool available at `http://www.bioinformatics.org/sms2/`

`translate.html`. 3' sequences were reverse complemented and aligned with the 5' sequence in order to clarify any misreads in either strand. In order to speed up this process, a script has been written in Python to identify the sequence number of the nucleotide in question so that it can be located quickly in the ab1 file.

2.2.14 Midi-prep and Maxi-prep

Plasmid DNA was routinely purified using the QIAprep Spin Midiprep kit and QIAprep Spin Maxiprep kit (Table 2.9). A single bacterial colony containing the plasmid DNA was grown in 5 ml media containing the appropriate antibiotic required for selection (see Section 2.1.16 for bacterial growth and selection media routinely used) for 8 hours at 37°C with vigorous shaking. The culture was then transferred to a further 50 ml (midi) or 200 ml (maxi) of media containing the antibiotic and grown for a further 12-16 hours. The bacteria was harvested and the plasmid DNA purified according to the manufacturer's instructions. This involved the preparation and clearance of lysate, followed by binding to a silica membrane, elution of the DNA from the membrane and finally the concentration and desalting of DNA by isopropanol precipitation.

First, bacterial cells were pelleted by centrifugation and resuspended in buffer containing RNase. The cells were lysed in NaOH/SDS buffer in order to release the cell contents whilst at the same time, denaturing chromosomal and plasmid DNA as well as proteins. Acidic potassium acetate was added to neutralise the lysate. The lysate was then cleared by passing the solution through a QIAfilter cartridge. Cleared lysate was loaded onto a QIAGEN-tip and passed through by gravity flow. As plasmid DNA was then bound to the membrane, contaminants were removed by washing with a medium-salt buffer before elution by a high-salt buffer. The eluted plasmid DNA was desalted and concentrated by isopropanol precipitation, washed with ethanol and dissolved in TE buffer. The concentration of the purified plasmid DNA was established using a nanodrop spectro-photometer.

2.2.15 Nanodrop

Nucleic acid concentrations were quantitated using a spectrophotometer (Nanodrop ND-1000). DNA absorbance was measured at 260 nm.

2.2.16 Glycerol stocks

700 µl of bacterial culture was added to 300 µl 50% glycerol and mixed well. Stocks were stored at $-80°C$

2.3 Tissue Culture Methods

2.3.1 Preparation and maintenance of cells

Tissue culture reagents were pre-warmed to 37°C in a water-bath prior to cell culture work.

CHO–K1 (Sigma 85051005) and HeK (Sigma 85120602) cells were cultivated and maintained in growth medium (Section 2.1.15), incubated at 37°C with 5% CO_2 and split when $70 - 80\%$ confluent.

To prepare a new passage of cells for growth, 1 ml pre-warmed media (Section 2.1.15) was added to cells to defrost rapidly and the suspension added to 5 ml media (Section 2.1.15) in 15 ml falcons. The suspension was centrifuged at 180 xg for 5 mins and supernatant discarded to remove all dimethyl sulfoxide (DMSO). 5 ml media was added to the cells and the pellet re-suspended. The final suspension was added to 15 ml media in Nunc EasYFlasks (Sigma-Aldrich, UK) and incubated at 37°C with 5% CO_2.

2.3.2 Splitting Cells

Cells were cultured in Nunc EasYFlasks (Sigma) in either growth media or selection media (Section 2.1.15). Cells were split into new flasks when $70 - 80\%$ confluent. First, supernatant was discarded and cells washed in $2 - 20$ ml sterile PBS (depending on the size of the flask). 1 - 5 ml trypsin was added to coat the cells and excess poured off. The flasks were then incubated at 37°C for 5 mins to release cells. Cells were resuspended in $2 - 20$ ml of the correct media (depending on length of culture) and then added to a flask containing media to make up the maximum volume for the flask.

2.3.3 Transfection of cells

Cells were transfected using Fugene 6 transfection reagent (Promega), according to the manufacturer's instructions. After 24 hours, the medium was replaced with medium containing zeocin to select for successfully transfected cells.

2.3.4 Production of monoclonal population

Monoclonal populations were developed by diluting cells down to 1 x 10^2/ml and adding 100 µl to each well of a 96 well plate. When colonies formed from a single cells, these were expanded further and immunoblotting used to analyse protein secretion and select for the best expressing colonies.

2.3.5 Freezing cells for long-term storage

The medium was discarded and the cells washed in PBS then washed in trypsin/EDTA, and the flasks returned to 37°C to loosen the cells. Cells were resuspended in 5 ml selection medium (Section 2.1.15) and spun at 180 xg for 5 mins. The medium was discarded, the cells resuspended in 1 ml freezing media (10 %DMSO in FBS) and transferred to a cryovial. This was transferred to a freezing container (Thermo Scientific) and incubated at −80°C overnight, before being placed in liquid nitrogen for long-term storage.

2.4 Methods for Protein Purification and Analysis

2.4.1 Fast protein liquid chromatography FPLC

DBL-Fc recombinant protein was purified from supernatant using a protein G sepharose column using the liquid chromatography system ÄKTA-FPLC Frac-950 (Amersham Biosciences/GE Healthcare, Uppsala, Sweden). Filter-sterilised cell supernatant was either pure or diluted 1:1 in the appropriate binding buffer and then passed over the column in order to allow the protein to bind. The column was then washed in binding buffer to remove any unbound material. Protein was eluted from the column in 0.1 M Glycine pH 2.5 - 2.7 using a linear gradient and collected in 2 ml fractions. Fractions were neutralised using 1 M Tris HCl pH 9.

The Beer-Lambert equation was used for quantification of protein concentration from FPLCs:

$$\text{Concentration} = \frac{OD_{280nm}}{\epsilon * \text{pathlength}} \quad (1)$$

where path length is 0.2 cm. The molar absorptivity of the DBL-Fc fusion proteins were calculated as 103915 for DBL 3.4 and 101395 for DBL 3.8 (calculated using the ExPASy bioinformatics tool available at `http://web.expasy.org/protparam/`). Percent

solution extinction coefficients were calculated as 1.664 and 1.630 respectively.

2.4.2 Immunoblot Analysis

Immunoblot analysis was used to show the presence of protein. 5 µl of solution containing protein was coated on nitrocellulose and subsequently blocked with PBST plus 5% milk for 2 hours at room temperature. The membrane was then washed in PBST before adding the detecting antibody αhIgG-AP (1:5000) and incubating overnight at room temperature. The blot was developed using NBT/BCIP tablets (Table 2.12) according to manufacturer's instructions.

2.4.3 Ultrafiltration

Protein was concentrated using ultrafiltration spin columns Vivaspin 2, 5000 MWCO PES (Sartorius) or Ultra-4 centrifugal filter devices (Amicon). Samples were added to the concentrator at max volume of 2 ml and then spun at 4000 x g in a tabletop centrifuge until the desired concentration was achieved.

2.4.4 Size Exclusion Chromatography

Protein aggregation was analysed by gel filtration using the liquid chromatography system ÄKTA-FPLC Frac-950 with gel filtration column Superdex 200 10/300 GL (Amersham Biosciences) according to manufacturer's instructions.

The location of peaks was compared to the gel filtration standards in order to identify molecular weight (Figure 2.2).

2.4.5 Bio-Rad Protein Assay

The Bio-Rad protein assay was used in order to quantify protein concentrations. This assay, based on the Bradford method, detects a colour change in dye in response to various protein concentrations.

The standard procedure followed for proteins concentrations in the range 0.05 mg/ml to 0.5 mg/ml was as follows: 10 µl of each sample was added to duplicate wells of Nunc-maxisorp micrometer plates (Thermo/Fisher) and 200 µl diluted dye reagent (1 part dye reagent concentrate to 4 parts dH_2O, filtered to remove particulates) added to each well. Colour was left to develop at room temperature for 5 minutes and absorbance

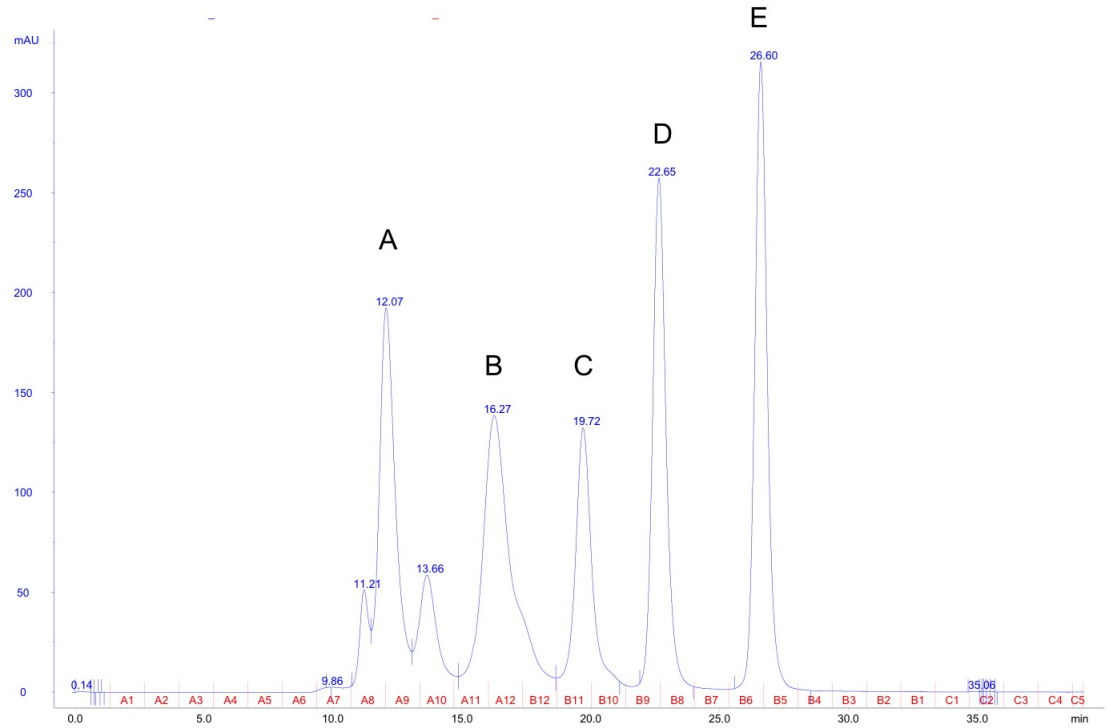


Figure 2.2: BioRad gel filtration standards. Peaks correspond to the molecular markers A) Tryroglobulin (bovine) - 670,000 Da, B) γ-globulin (bovine) - 158,000 Da, C) Ovalbumin (chicken) - 44,000 Da, D) Myoglobin (horse) - 17,000 Da and E) Vitamin B_{12} - 1,350 Da.

read at 595 nm on an Infinite F50 plate reader (Tecan). IVIG concentrations ranging from 0.05 mg/ml to 0.5 mg/ml were used as protein standards for comparison.

The micro assay procedure followed for protein concentrations in the range 8.0 µg/ml to 80 µg/ml was as follows: 160 µl of each sample was added to duplicate wells of Nunc-maxisorp micrometer plates (Thermo/Fisher) and 40 µl diluted dye reagent added to each well and mixed. Colour was left to develop at room temperature for 20 minutes and absorbance read at 595 nm on an Infinite F50 plate reader (Tecan). IVIG concentrations ranging from 8.0 µg/ml to 80 µg/ml were used as protein standards for comparison.

2.4.6 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were prepared with NuPAGE sodium dodecyl sulphate (SDS) sample buffer, and loaded either under reducing and non-reducing conditions, using dithiothreitol (DTT) as a reducing agent. The samples were centrifuged at 27,396 xg for 30 seconds and then heated at 95° C for 5 minutes before loading onto a NuPAGE Novex 4-12% Bis-Tris

pre-cast gel (Life Technologies, CA, USA). An XCell SureLock Mini-Cell electrophoresis system (Life Technologies) was used to run the gels in NuPAGE MOPS SDS Running Buffer (Life Technologies; 50 mL MOPS buffer in 950 mL of deionised water) at 200 V constant until the samples reached the end of the gel. SeeBlue Plus 2 Prestained Protein Marker (NEB) was used as a molecular weight marker (Figure 2.3). The proteins were either Coomassie stained (Section 2.4.7) or transferred to a membrane for Western blotting (Section 2.4.8).

2.4.7 Coomassie Blue staining

The gels were then removed from their housing cassettes and stained in 20 mL of Page Blue Protein Staining Solution (2.1.6) by microwaving at full power (750 watts) for 30 seconds. To optimize staining, the gels were then left to stain for a further 30 minutes on a platform shaker (Stovall Life Sciences Inc., Ringer, UK), before being left to destain for two hours in destaining solution. Protein bands were directly visualized on the destained gel and their respective molecular weights (kDa) were estimated by comparison with SeeBlue Plus2 prestained standards (Invitrogen).

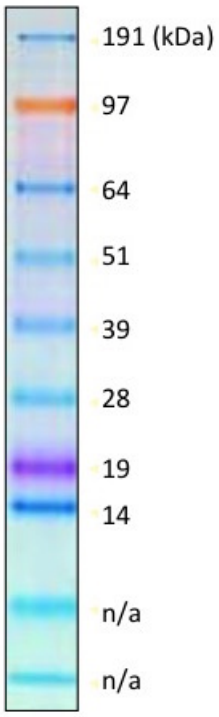


Figure 2.3: SeeBlue Plus2 prestained standard used as molecular weight markers to analyse the size of protein bands.

2.4.8 Western blot

SDS-PAGE gels were transferred to Protran nitrocellulose transfer membrane (GE Healthcare) in transfer buffer (Section 2.1.12) at 40V for one hour, using an XCell IITM blot module according to manufacturer's instructions. To block non-specific binding, the membrane was incubated in blocking solution for one hour, followed by three washes (PBST). The membrane was then incubated with primary antibody overnight on a rocker at room temperature, followed by washing and incubation with a secondary antibody conjugated to an appropriate substrate for three hours. On occasion when a primary antibody was not necessary, the membrane was directly coated with the conjugated secondary antibody overnight.

Proteins were visualised by developing horseradish peroxidase (HRP)-conjugated Abs with the chromogen 3',3'-Diaminobenzidine (DAB; DAKO), or alkaline phosphatase (AP)-conjugated antibodies with a developing buffer composed of 5-bromo-4-chloro-indolyl-phosphate/Nitroblue tetrazolium (BCIP/NBT, 100 mM Tris/Cl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$), according to manufacturer's instructions.

2.4.9 Enzyme-linked immunoabsorbent assay (ELISA)

Sandwich ELISA was used to measure the amount of protein between a capture and detection antibody. The capture antibody was diluted in 0.5 M carbonate/bicarbonate buffer (pH 9.6), coated down on a Nunc-maxisorp micrometer plates (Thermo/Fisher) and incubated at 4°C overnight in order to bind to the plate. Any uncoated material was then discarded and the plates were blocked with 200 µl blocking buffer (Section 2.1.8) and incubated for 1 hour at room temperature on a rocker. Wells were washed in PBST 5 times before the addition of protein or antibody in the appropriate dilution. The plate was incubated for 2 hours at room temperature on a rocker before washing all wells 5 times in PBST. This was repeated depending on the purpose of the ELISA. Wells were finally washed 5 times in PBST before the addition of the detection antibody (Alkaline Phosphate conjugate) which was incubated for 1 hour at room temperature. Plates were then washed as before and then developed using SIGMA$FAST^{TM}$ p-Nitrophenyl Phosphate tablets (20 ml solution containing 1.0 mg/ml pNPP, 0.2 M Tris buffer and 5 mM magnesium chloride). 100 µl of substrate buffer was added to each well and the plate was allowed to develop. The absorbance was read at 405 nm on an Expert plus plate reader (Biochrom, SLS).

Chapter 3

Production of Recombinant MSPDBL1 and MSPDBL2 DBL Domains

3.1 Background

MSP3 family

The MSP3 family of proteins are secreted polymorphic antigens associated with the merozoite and erythrocytic shizonts in the blood stages of *P. falciparum* infection and consists of six proteins: MSP3 (MSP 3.1), MSP6 (MSP 3.2), H101 (MSP 3.3), MSPDBL1 (MSP 3.4), H103 (MSP 3.7) and MSPDBL2 (MSP 3.8) (Singh et al., 2009). MSP3 proteins are synthesised simultaneously in the schizont stages of infection, after which they undergo proteolytic processing before being released by schizonts upon rupture (McColl and Anders, 1997). Most of the cleaved fragments are secreted, however some associate non-covalently to the merozoite surface (McColl and Anders, 1997). MSP3 proteins are highly conserved amongst isolates and are a target of naturally occurring antibodies which are able to cross-react across the gene family (Singh et al., 2009; Demanga et al., 2010). Anti-MSP3 antibodies have been associated with acquired clinical immunity (Roussilhon et al., 2007; Osier et al., 2007) and in phase-1 vaccine trials have elicited antibodies both *in vitro* and *in vivo* that mediate effective killing of *P. falciparum* (Druilhe et al., 2005). Anti-MSP3 antibodies therefore play an important role in protective immunity against malaria.

Members of the MPS3 family are characterised by sequence homology in their C-terminals and share a NLR(K/A)(A/G/N) signature motif (Figure 3.1). The C terminal region varies between each paralog but is highly conserved amongst isolates, which suggests that they may play a role in parasite survival. Each member also has a secreted polymorphic antigen associated with merozoites (SPAM) domain in the C terminal region, the role of which remains elusive but is thought to be involved in oligomerisation. MSPDBL1 and MSPDBL2 also contain an additional DBL domain which is found in *var* and *ebl* genes.

MSPDBL1 and MPSDBL2

MSPDBL1 and MSPDBL2 are the only MSPs to contain DBL domains which are well-known binding domains present in other *Plasmodium* protein families (Cowman et al., 2012). Both proteins appear to localise on the merozoite surface (Hodder et al., 2012; Wickramarachchi et al., 2009; Cowman et al., 2012) and have been identified as part of

the MSP1 complex, thought to be involved in erythrocyte invasion (Lin et al., 2014).

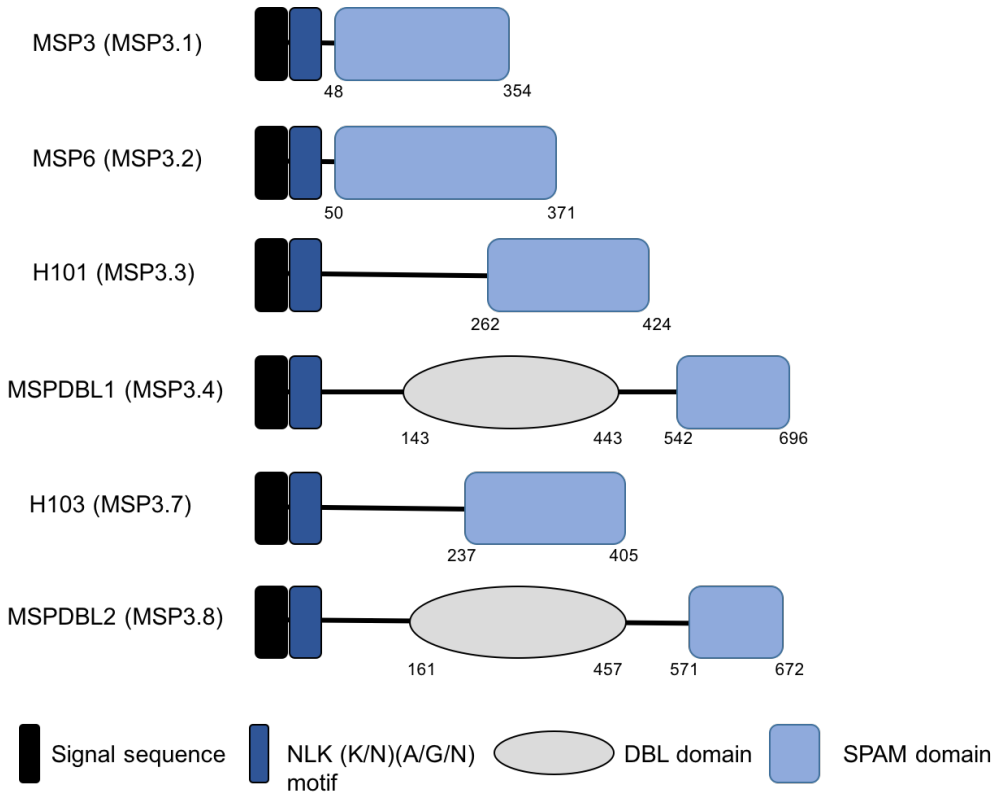


Figure 3.1: **Structure of the MPS3 family of merozoite surface proteins.** Schematic shows the structural characteristics of MSP3 family of proteins. Adapted from (Hodder et al., 2012).

It has been confirmed that the DBL domains are able to bind erythrocytes although the receptors have not yet been identified (Hodder et al., 2012; Wickramarachchi et al., 2009). MSPDBL1 and MSPDBL2 have also been observed to bind the Fc portion of natural IgM through their DBL domain (Crosnier et al., 2016). This characteristic has also been reported in DBL variants from PfEMP1 and is thought to protect the parasite from PfEMP1-specific IgG (Czajkowsky et al., 2010). In the case of PfEMP1, IgM-binding has been linked to several virulence-associated phenotypes including rosetting with uninfected RBC in severe childhood malaria as well as binding to CSA in placental malaria (Semblat et al., 2015). Due to their implied role in the erythrocyte invasion process as well as their interesting binding properties, MSPDBL1 and MSPDBL2 deserve attention as blood stage vaccine candidates (Chiu et al., 2015).

Structure of MSPDBL1 and MSPDBL2 DBL domains

The three-dimensional crystal structure of the DBL domain of MSPDBL2 contains a boomerang-like shaped α-helical core with three cysteine rich subdomains (Figure 3.2) and shows structural similarities with DBL domains from PfEMP1 such as VAR2CSA, DBL6e, DBL3X and PfEBA-175 (Hodder et al., 2012). The structure is stabilised by six inter-chain disulphide bonds which show a different linkage pattern than DBL domains in the PfEMP1 and EBL family. DBL domains from the EBL family do not typically have disulphide connectivity between subdomains 1 and 2 and although PfEMP1 DBL domains do, they typically have a modified linkage pattern in subdomain 3 (Hodder et al., 2012).

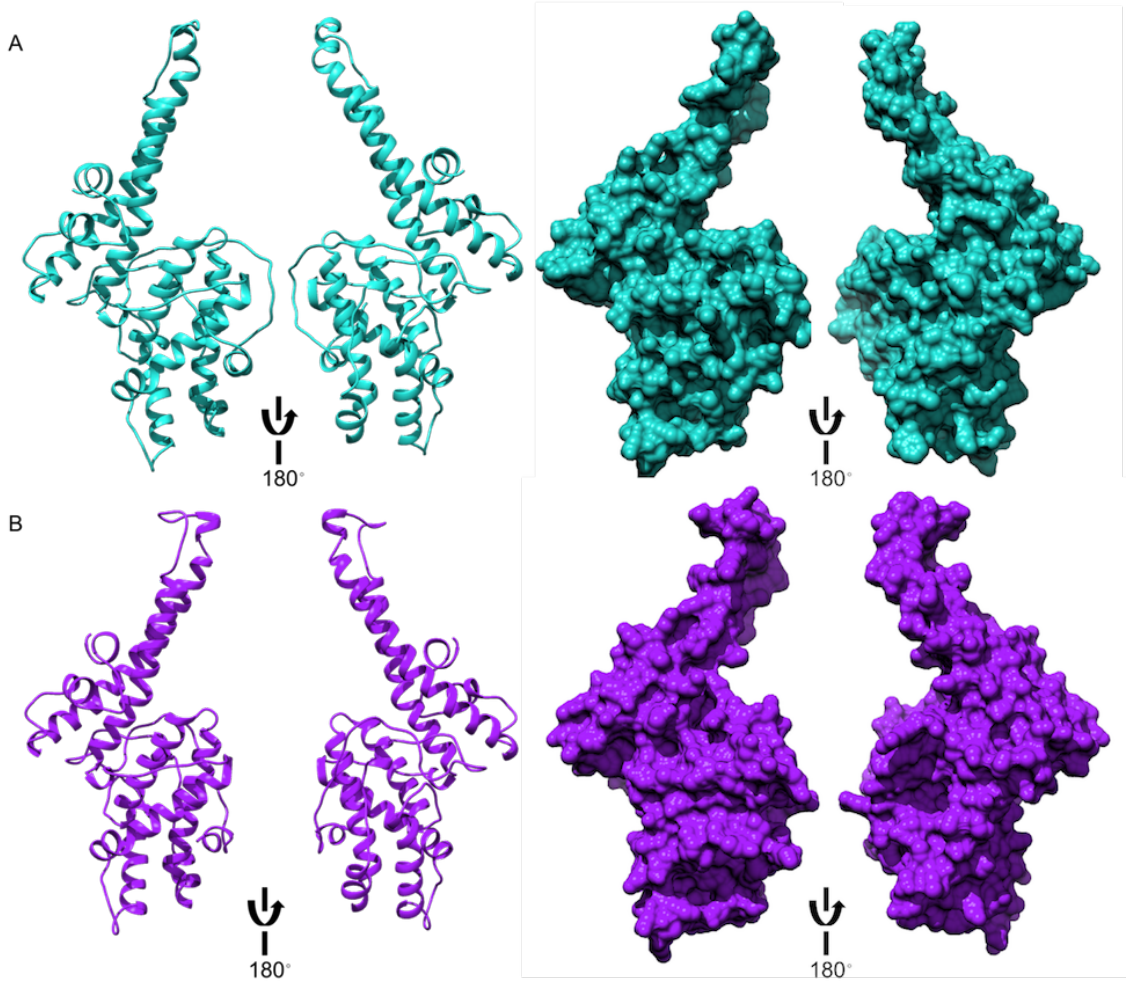


Figure 3.2: **Ribbon schematic and surface potential diagram of DBL domains from MSPDBL1/2.** (A) Ribbon schematic and surface potential diagrams showing the modelled structure of the MSPDBL1 DBL domain. This homology model is based on the crystal structure of MSPDBL2 and was modelled using Phyre 2. (B) Ribbon schematic and surface potential diagrams showing structure of the MSPDBL2 DBL domain including loops missing in the x-ray structure (PDB accession code 3VUU) which were modelled using Phyre 2 (adapted from (Hodder et al., 2012))

MSPDBL1 and MSPDBL2 bind IgM through DBL domains

The focus of this chapter is the ability of the DBL domains from MSPDBL1 and MSPDBL2 to bind the Cμ4 domain of human IgM. Interestingly this is the same region on the Ab that the FcμR is known to bind. In a recent study, reduced binding of the full-length proteins to immune IgG was observed when complexed with IgM (Crosnier et al., 2016). This suggests that binding of the MPSDBL1 and MSPDBL2 proteins to IgM is an evasion mechanism which acts to mask the merozoite from the immune system (Crosnier et al., 2016).

Expression of recombinant DBL-Fc fusion proteins

Fc-fusion proteins have shown a great deal of success as therapeutics, with nine currently in clinical use (reviewed in Section 1.2). The increased half-life from the interaction with the neonatal Fc receptor (FcRn) improves the therapeutic potential of the protein partner. Fc-fusion proteins used for drug delivery mainly target receptor-ligand interactions either working to block receptor binding or to directly stimulate receptor function and alter immune activity (Czajkowsky et al., 2012). Fc-fusion proteins also have potential for vaccine use as they are able to effectively deliver an antigen which can prime the immune response. Creating Fc-fusion proteins from malarial antigens (e.g. RTS,S) may improve on existing vaccine delivery by this improved priming effect.

DBL-Fc fusion proteins may be used as therapeutics for IgM-related autoimmune diseases due to their high affinity for IgM. Removal of IgE by the drug Omalizumab is highly effective for the treatment of asthma. Omalizumab reduces the clinical symptoms caused by hypersensitivity that is induced from overproduction of IgE (Strunk and Bloomberg, 2006). Anti-TNF therapy in rheumatoid arthritis and other inflammatory conditions such as ulcerative colitis blocks the action of the tumour necrosis factor (TNF) the overproduction of which is responsible for the inflammatory response (Feldmann, 2002; Prince et al., 2011). Removal of IgM from serum may reduce the symptoms of diseases such as Waldenstrőm's macroglobulinemia and Schnitzler's syndrome, peripheral neuropathies such as Anti-MAG demyelinating neuropathy, and systemic deposition of IgM (Simon et al., 2013; Camp and Magro, 2012; Audard et al., 2008; Baldini et al., 1994). Reduction of IgM and prevention of complement activation may also be useful for the treatment of diseases and infections associated with type II cryoglobulinemia (Terrier et al., 2012). The

treatment of such diseases with DBL-Fc fusion proteins would allow direct targeting of nIgM and would be more beneficial than mAbs-directed therapeutics because there would be no cross-linking of the IgM BCR.

Expressing the merozoite DBL domains as Fc-fusion proteins provides a proof of concept for the production of malarial DBL-based Fc-fusion proteins and the techniques used for production may be developed and applied to further therapeutic antigens. The expression of the DBL domains with the Fc tag also provides a useful purification tool. Protein G can be used for purification, streamlining protein production in order to focus on the binding characteristics of these interesting protein domains.

The DBL-Fc fusions themselves could also be used for IgM purification due to their high affinity for hIgM. IgM purification is a challenge due to its size and susceptibility to denaturation (Gautam and Loh, 2011) and cost-effective purification is not as readily available and reliable as it is for IgG. Protein-based affinity ligands are used for purification, such as mannose-binding protein (MBP) which currently is immobilised on agarose beads for commercial IgM purification. However, not all glycosylation variants of hIgM are purified by this method, leading to products which are only 90% pure. Other commercially available human IgM purification kits use ligands bound to sepharose but purities above 95% cannot be reached (Gautam and Loh, 2011).

The high A:T content of the *Plasmodium* genome results in low codon compatibility in heterologous expression systems and makes expression of recombinant protein difficult. Highly repetitive amino acid sequences and high molecular weights introduce further technical challenges. However, recombinant proteins have been successfully expressed in a diverse range of systems from bacteria, yeast and plants to mammalian systems (Birkholtz et al., 2008; Fernández-Robledo and Vasta, 2010; Stowers et al., 2001; Tsai et al., 2006; van Bemmelen et al., 2000; Ghosh et al., 2002; Gregory et al., 2012; VanBuskirk et al., 2004; Tolia et al., 2005; Tsuboi et al., 2008). *Escherichia coli* is currently the most popular expression system due to low cost and relatively high protein yield. However, producing protein under reducing conditions in the bacterial cytoplasm inhibits the formation of the disulphide bonds that are critical for structural stability. Extracellular proteins in particular contain structurally critical disulphide bonds. This means that when producing recombinant *Plasmodium* extracellular vaccine candidates in *E. coli* great care must be taken to create a properly folded disulphide-bonded protein, and the process can be

laborious and complex.

In eukaryotic cells, the endoplasmic reticulum (ER) provides an oxidative environment for disulphide bond formation, thus mammalian expression systems can make the the disulphide bond structure essential for *Plasmodium* extracellular proteins, however often produce low yields. Unnecessary N-linked glycosylation is often a problem in eukaryotic expression of recombinant *Plasmodium* surface proteins because, unlike most other eukaryotic extracellular protein, these surface and secreted proteins are not modified by N-linked glycans. Mammalian expression of the Fc-fusion proteins enables disulphide bond formation while potential glycosylation sites within the DBL domain have been removed.

The production of recombinant DBL domains from both MSPDBL1 and MSPDBL2 will allow functional and molecular characterisation of these exciting protein domains. A mammalian expression system can be used to produce functional protein with the correct disulphide bond structure, which is critical for the stability of the protein. The methods used to develop recombinant Fc-fusion DBL proteins can be applied to other malarial antigens, paving the way for the therapeutic use of malarial Fc-fusion proteins.

3.2 Objectives

The aim of this chapter is to produce recombinant DBL domains from MSPDBL1 and MSPDBL2 with an IgG-Fc fusion and to investigate the binding properties of the DBL domains. The specific objectives include:

- Produce novel Fc-fusion constructs for mammalian expression of the recombinant protein.
- Express recombinant DBL-Fc fusion protein.
- Show IgM binding of the DBL domains using the recombinant Fc-fusion proteins.
- Show that MSPDBL1 and MSPDBL2 bind to the Cμ4 domain of human IgM.

3.3 Methods

3.3.1 DNA preparation

The DBL domains from MSPDBL1 and MSPDBL2 will be referred to as DBL 3.4 and DBL 3.8 respectively.

```
- Glycans    KCPDENFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINK
DBL 3.4      KCPDENFCKGIKNVLSCPPKNSTGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINK
             **********************:************************************

- Glycans    VWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDY
DBL 3.4      VWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDY
             ************************************************************

- Glycans    QITKNINRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNM
DBL 3.4      QITKNINRALDKILRNETSNDKIKKRVDWWEANKSAFWDAFMCGYKVHIGNKPCPEHDNM
             *****************:****************:************************

- Glycans    DRIPQYLRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKH
DBL 3.4      DRIPQYLRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKH
             ************************************************************

- Glycans    YEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNTFKEFK
DBL 3.4      YEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNTFKEFK
             ************************************************************

- Glycans    D
DBL 3.4      D
             *
```

(a)

```
- Glycans    KCPTEEICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDF
DBL 3.8      KCPTEEICKDFSNLPQCRKNVHERNNWLGSSVKNFSSDNKGVLVPPRRQSLCLRITLQDF
             ********************************:************************

- Glycans    RTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTP
DBL 3.8      RTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTP
             ************************************************************

- Glycans    TSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNI
DBL 3.8      TSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNI
             ************************************************************

- Glycans    DDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKY
DBL 3.8      DDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKY
             ************************************************************

- Glycans    EEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQDVFELTFD
DBL 3.8      EEWYNKRKTEWTEQSIKYNNDKINYTDIKTLSPSEYLIEKCPECKCTKKNLQDVFELTFD
             ************************:***********************************
```

(b)

Figure 3.3: Alignment of the wild type amino acid sequences of (a) DBL 3.4 and (b) DBL 3.8 with the N-linked glycan sites removed (- Glycans). Glycan sites in the WT are highlighted in red.

Mutation of N-linked glycosylation sites

In order to prevent the inappropriate addition of large glycans that are absent in *P. falciparum*, potential N-linked glycosylation sites of the form N-X-S/T have been mutated with Serine/Threonine residues substituted by Alanine (Figure 3.3). Three glycosylation

sites have been removed from the 3.4 DBL domain and two from the 3.8 DBL domain. The removal of N-linked glycosylation sites in the mammalian expression of the *P. falciparum* protein RH5 has previously been shown to increase protein yield 5.5 fold (Crosnier et al., 2013).

Codon optimisation for mammalian expression

The high A:T content of *Plasmodium* gene sequences means that codon usage is significantly different to mammalian systems. In order to express DBL 3.4 and DBL 3.8 domains in mammalian expression systems the codon usage in the genes must be modified. Codon optimised DNA for mammalian expression was kindly provide by Gavin Wright (Wellcome Trust Sanger Institute, Hinxton, Cambridge).

Location of the DBL domains

When the work for this Chapter started, the exact DBL location was unknown. The DBL sequence published by Wickramarachchi et al. (2009) is 44 amino acids shorter than the sequence of the crystal structure published by Hodder et al. (2012). Prior to beginning the cloning work for this Chapter, the DBL domain location was discussed and on advice from Matt Higgins (University of Oxford) and Richard Pleass (Liverpool School of Tropical Medicine) an extended version of the sequence by Wickramarachchi et al. (2009) was used. The crystal structure was determined after the work in this Chapter was undertaken which identified the exact location of the DBL. This was thirteen amino acids longer than the DBL sequence used here, with six extra residues included at the N terminus and seven extra at the C terminus. The location of the DBL domain and the Secreted polymorphic antigen associated with merozoites (SPAM) domain is shown within the full length MSPDBL1 (Figure 3.4) and MSPDBL2 (Figure 3.5). The beginning and end of the shorter DBL sequence used in this Chapter is marked with |.

ATMEFQTQVLMSLLLCMSGAAANDLINYNDANLRNGLLNNALDLTNGLNNKDNSFID
SKIEEHENKAYQNKDNNIAIVGQDVPITSVYSSKIINANDLEGNSIDDTKGLSVTNS
GFDDGSAFGGGLPFSGYSPLQGNHNKCPDEN|FCKGIKNVLSCPPKNSAGRNGDWIS
VAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIKDEKNFKEEFVKVALGESNALMK
HYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDKILRNEASNDKIKK
RVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYK
NKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKF
EKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN|TFKEFKDNVALLKAVIDNKKN
QDSLTTTSLSTSINSVRDSSNLDQRGNIATSQGNSHRATVVQQVDQTNRLDNVNSVT
QRGNNNYNNNLERGLGSGALPGTNIITEEKYSLELIKLTSKDEEDIIKHNEDVREEI
EEQQEDIEEDEEELENEGEETKEEDDEEKNEANDAEDTDDTEDTEDIEEENKEKELS
NQQQSEKKSISKVDEDSYRILSVSYKDNNEVKNVAESIVKKLFSLFNDNNNLETIFK
GLTEDMTDLFQK

(a)

gccaccatggagtttcagacccaggtactcatgtccctgctgctctgcatgtctggtgcggccgccaac
gacctgatcaactacaacgacgccaacctgcggaacggcctgctgaacaacgccctggacctgaccaac
ggcctgaacaacaaggacaacagcttcatcgacagcaagatcgaggaacacgagaacaaggcctaccag
aacaaggataacaatatcgccatcgtgggccaggacgtgcccatcaccagcgtgtacagcagcaagatc
atcaacgccaacgatctggaaggcaacagcatcgacgacaccaagggcctgagcgtgaccaacagcggc
ttcgacgacggcagcgcctttggcggcggactgcctttcagcggctacagccccctgcagggcaaccac
aacaagtgccccgacgagaac|ttctgcaagggcatcaagaacgtgctgagctgccctcccaagaacag
cgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgct
ggtgccccccagacggaccaagctgtgcctgcggaacatcaacaaagtgtggcaccggatcaaggacga
gaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaa
agagaagaacctgaacgccctgaccgccattaagtacggcttcagcgacatgggcgacatcatcaaggg
caccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaagatcctgagaaacga
ggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgc
cttcatgtgcggctacaaggtgcacatcggcaacaagccctgccccgagcacgacaacatggaccggat
cccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcga
ggacgtgatcaagctgtgcaacatccagcagttcaccaaccaggacgacagccagctgctggaaatcag
caagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggcggaggcccgagtg
gaagggccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccga
gaagtacctgaaagaaatctgcagcgagtgcgactgcaagtacaaggacctggacaac|accttcaaag
aattcaaggacaacgtggccctgctgaaggccgtgatcgacaacaagaagaaccaggacagcctgacca
ccacaagcctgagcaccagcatcaacagcgtgcgggacagcagcaacctggaccagcggggcaacattg
ccaccagccagggcaacagccaccgggctaccgtggtgcagcaggtggaccagaccaacagactggaca
acgtgaacagcgtgacccagagaggcaacaacaactacaacaacaatctggaacggggcctgggctctg
gcgccctgcccggcaccaacatcatcaccgaggaaaagtacagcctggaactgattaagctgaccagca
aggacgaagaggatatcatcaagcacaacgaggacgtgcgcgaggaaattgaggaacagcaggaagata
ttgaagaggacgaggaagaactggaaaacgagggcgaggaaacaaaagaagaggacgacgaggaaaaga
acgaggccaacgacgccgaggacaccgacgacaccgaggatacagaggacatcgaagaggaaaacaaag
agaaagagctgagcaaccagcagcagagcgagaagaaatccatcagcaaggtggacgaggactcctacc
ggatcctgagcgtgtcctataaggacaacaacgaagtgaagaacgtggccgagagcatcgtgaagaagc
tgttcagcctgttcaacgataacaacaacctggaaaccatcttcaagggcctgaccgaggacatgaccg
acctgttccagaaggcgcgccgtcgacctccatcacggcctataagagtgaggggagtcagcggagt
tctccttcccactcaaccttggagaggaaagcctgcagggagagttgagatggaaggcagagaaggctc
cttcttcccagtcctggatcaccttctccctaaagaaccaaaaggtgtctgtgcagaagtctactagca
accccaagttccagctgtccgaaacgctcccactcacccttcagataccccaggtctcccttcagtttg
ctggttctggcaacctgaccctgactctggacagagggatactgtatcaggaagtgaacctggtggtga
tgaaagtgactcagcccgacagcaacactttgacctgtgaggtgatgggacccacctcacccaagatga
gactgatcttgaagcaggagaatcaggaggccagggtctccaggcaggagaaagtgattcaagtgcagg
cccctgaagcaggggtgtggcaatgtctactgagtgaaggtgaagaggtcaagatggactccaagatcc
aggttttatccaaagggttgaattccggatcactgcatcatattctggatgcacagaaaatggtgtgga
atcatcgttaa

(b)

Figure 3.4: (a) Location of the 3.4 DBL within the full length MSPDBL1 amino acid sequence. (b) Location of the 3.4 DBL within the full length MSPDBL1 nucleotide sequence. The DBL domain is shown in red and the SPAM domain in green. The DBL domain expressed in this Chapter is missing residues at either end compared to the literature, the start and end location of the DBL expressed is marked with |.

MIYILSIVFYIFFLHIDIYVNIYSTCFVVNEGNPNLRNNIINDDELKGKAYNNTIDA
NNQNIEYNKNLKHNVNSSHISKFSDIMDQEDKGDNENSHDIKFEEKKNINKSLDAES
NYGINEISITGNDSNSDNSNQNIFPDGSELAGGIPRSIYTINLGFNKCPTEE|ICKD
FSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTKKKKE
GDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKE
TITYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNID
DIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLK
KYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQD|V
FELTFDGKALLEKLKKEESPVSNSVNALPEPGQITLPDPSLKQTTQQENQPVVETPV
TTAVINEHQGQTEPNKGDNNNERENHESNVGSIQEVNQGSVSEESHSKTIDPSKIDD
RLELSSGSSSLEQHSKEDVKKGCALELVPLSLSDIEQIANESEDVLEEIEEEINTDG
EIEYITEEEIKEDIEEETEEDIEEETEEETEEETEEEADEETVKEIEDKPEQEIKNK
SLEEKQIDKNTDTSEKKGFNNSEKDEKARNLISKNYKNYNELDKNVHTLVNSIISLL
EEGNGSDSTLNSLSKDITNLFKN

(a)

gccaccatggagtttcagacccaggtactcatgtccctgctgctctgcatgtctggtgcggccgcctac
agcacctgtttcgtggtcaacgagggcaaccccaacctgcggaacaacatcatcaacgacgacgagctg
aagggcaaggcctacaacaacgccatcgacgccaacaaccagaacatcgagtacaacaagaacctgaag
cacaacgtgaacagcgcccacatcagcaagttcagcgacatcatggaccaggaagataagggcgacaac
gagaacagccacgacatcaagttcgaggaaaagaagaacatcaacaaggccctggacgccgagagcaac
tacggcatcaacgagatcagcatcaccggcaacgacgccaactccgacaacagcaaccagaatatcttc
cccgacggcagcgagctggctggcggcatcccccggtccatctacaccatcaacctgggcttcaacaag
tgccccaccgaggaa|atctgcaaggacttcagcaacctgccccagtgccggaagaacgtgcacgagcg
gaacaactggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccag
acggcagagcctgtgcctgagaatcaccctgcaggacttccggaccaagaagaagaaagagggcgactt
cgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacct
ggaaaaggcccaccaggccatccggtacagcttcgccgacatcggcaacatcatccggggcgacgacat
gatggacaccccaccagcaaagagacaatcacctatctggaaaaggtgctgaagatctacaatgagaa
caacgacaagcccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgat
gtgcggctaccagagcgcccagaaagacaaccagtgcaccggctacggcaacatcgacgatatccccca
gttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaataccct
gaaggccgtgtgcttccccaagcagcccagaaccgaggccaaccctgccctgaccgtgcacgagaacga
gatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccgagtggaccgagca
gagcatcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacct
gatcgagaagtgccccgagtgcaagtgcaccaagaaaaacctgcaggac|gtgttcgagctgaccttcg
acggcaaggccctgctggaaaagctgaagaaagaagagtcccccgtctccaacagcgtgaacgccctgc
ccgagcccggccagatcaccctgcccgatcccagcctgaagcagaccacccagcaggaaaaccagcccg
tcgtcgaaacccccgtgaccaccgccgtgatcaacgagcaccagggccagaccgagcccaacaaggggg
acaacaacaatgagagagagaaccacgagagcaacgtgggcagcatccaggaagtgaaccagggcagcg
tgtccgaggaaagccacagcaagaccatcgaccccagcaagatcgacgaccggctggaactgagcagcg
gcagcagcagcctggaacagcacagcaaagaagatgtgaagaagggctgcgccctggaactggtgcccc
tgagcctgagcgacatcgagcagatcgccaacgaggccgaggacgtgctggaagagatcgaggaagaga
tcaacaccgacggcgagatcgagtacatcaccgaagaggaaatcaaagaggacattgaggaagagacag
aagaggatattgaagaggaaaccgaggaagaaaccgaagaagagactgaggaagaagccgacgaggaaa
ccgtgaaagagattgaggacaagcccgagcaggaaatcaagaacaaagccctggaagagaagcagatcg
acaagaacaccgacaccagcgagaagaagggattcaacaacgccgagaaggacgagaaggcccggaacc
tgatcagcaagaactacaagaattacaacgagctggacaagaatgtgcacaccctggtcaacagcatca
tcagcctgctggaagaaggcaacggcgccgactccaccctgaacagcctgagcaaggacatcaccaatc
tgttcaagaatggcgcgccgtcgacctccatcacggcctataagagtgaggggagtcagcggagttct
ccttcccactcaaccttggagaggaaagcctgcagggagagttgagatggaaggcagagaaggctcctt
cttcccagtcctggatcaccttctccctaaagaaccaaaaggtgtctgtgcagaagtctactagcaacc
ccaagttccagctgtccgaaacgctcccactcacccttcagataccccaggtctcccttcagtttgctg
gttctggcaacctgaccctgactctggacagagggatactgtatcaggaagtgaacctggtggtgatga
aagtgactcagcccgacagcaacactttgacctgtgaggtgatgggacccacctcacccaagatgagac
tgatcttgaagcaggagaatcaggaggccagggtctccaggcaggagaaagtgattcaagtgcaggccc
ctgaagcaggggtgtggcaatgtctactgagtgaaggtgaagaggtcaagatggactccaagatccagg
ttttatccaaagggttgaattccggatcactgcatcatattctggatgcacagaaaatggtgtggaatc
atcgttaa

(b)

Figure 3.5: (a) Location of the 3.8 DBL within the full length MSPDBL2 amino acid sequence. (b) Location of the 3.8 DBL within the full length MSPDBL2 nucleotide sequence. The DBL domain is shown in red and the SPAM domain in green. The DBL domain expressed in this Chapter is missing residues at either end compared to the literature, the start and end location of the DBL expressed is marked with |.

3.3.2 Amplification of DNA

PCR primers for amplification of DBL 3.4 and DBL 3.8 from MSPDBL1 and MSPDBL2 digested products.

Primers were designed to remove the DBL domain from the full length gene and to introduce flanking restriction enzyme (RE) sites before and after the DBL domain by PCR for subcloning. For DBL 3.4, BglII and NcoI sites were chosen as they are unique RE sites present in the multiple cloning region of the pFUSE vector. For DBL 3.8, EcoRI and NcoI sites were chosen (see Table 3.1 for oligonucleotide sequences).

Table 3.1: PCR primers for amplification of DBL 3.4 and DBL 3.8 domains. Restriction sites are highlighted in red.

Primer	Sequence	Annealing Temperature (°C)
3.4 Forward	atc ggc cat ggt ttt ctg caa ggg cat caa	68.1
3.4 Reverse	gtt ttg tca gat ctg ttg tcc agg tcc	65.0
3.8 Forward	cgg aat tcg atc tgc aag gac ttc agc	61.0
3.8 Reverse	ttg tca gat cta acc atg gcg tcc tgc a	66.6

PCR reaction

The PCR solution used to amplify the MSPDBL domains and to add restriction sites was composed of:

- dNTP (2.5 mM) 20 µl
- Forward primer (10 µM) 10 µl
- Reverse primer (10 µM) 10 µl
- DNA 200 ng
- Taq buffer 20 µl
- Taq polymerase 10 µl
- dH_2O up to 200 µl

The following PCR program was used:

94° C 3 mins

25 cycles {
- 94° C 1 min
- 60° C 1 min
- 74° C 1 min

72° C 5 mins.

A small amount of the PCR product was run on an agarose gel to check that the band of interest had been amplified (Section 2.2.5). Once it was confirmed that the PCR reaction was successful, the rest of the PCR product was purified from solution (Section 2.2.3) prior to ligation.

3.3.3 TOPO Ligation to the 3.4 DBL and 3.8 DBL domains

Due to the use of Taq polymerase in the PCR reaction, which adds a single deoxyadenosine (A) to the 3' end of the PCR product, the PCR product was cloned by TOPO cloning. The PCR product was added directly into the Topoisomerase 1-activated pCR2.1-TOPO plasmid which contains a single 3' thymidine (T) overhang and has Topoisomerase 1 covalently bound. The reaction was done according to manufacturer's instructions.

The following reaction was set up on ice:

- PCR product 4 µl
- Salt solution 1 µl
- TOPO vector 1 µl.

The ligation reaction was incubated at room temperature for 5 minutes before being placed on ice and transformed into competent cells.

3.3.4 Transformation of competent cells with pCR2.1-TOPO-3.4 and pCR2.1-TOPO-3.8 constructs

Before transforming competent cells with the TOPO constructs, agarose plates containing 50 µg/ml ampicillin (Section 2.1.16) were prepared by spreading 40 µl of 40 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 40 µl of 100mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) on the plates and drying them at 37°C prior to use.

2 µl of the TOPO ligation reaction was added to a vial of OneShot (TOP10) competent cells which had been pre-thawed on ice, and the mixture was incubated on ice for 30 minutes. Cells were then heat shocked in a water bath at 42°C for 30 seconds before immediately being transferred to ice. 250 µl of pre-warmed S.O.C media was added to the cell mixture and the tube shaken horizontally for 1 hour at 225 rpm in a shaker incubator at 37°C. 10 - 50 µl of each transformation was spread onto the prepared selective plates and incubated overnight at 37°C. White or light blue colonies were selected for screening (Section 2.2.12).

Restriction digest analysis of positive colonies

White and light blue colonies selected for screening were grown in overnight cultures and DNA extracted by mini-prep (Section 2.2.11). DNA was then digested using restriction enzymes in order to identify correct orientation of the insert.

The digestion reaction consisted of the following solution:

- Restriction enzyme 1 — 1 µl
- Restriction enzyme 2 — 1 µl
- Compatible Buffer — 5 µl
- BSA (10x) — 5 µl
- DNA — 1 µg
- dH_2O — up to 50 µl.

For the pCR2.1-TOPO-3.4 construct, restriction enzymes Nco1 and BglII were used with NEB buffer 3.1. For the pCR2.1-TOPO-3.8 construct, restriction enzymes Nco1 and

EcoRI were used with NEB buffer 3.1. Restriction digests were incubated at 37°C for 2 hours and then the product checked by gel electrophoresis (Section 2.2.5) in order to confirm the presence of the insert.

DNA Sequencing

In order to confirm the pCR2.1-TOPO-3.4 and pCR2.1-TOPO-3.8 constructs, DNA from any positive colony which had the presence of the insert confirmed by restriction digest, was sent for sequencing (Section 2.2.13). The commercially available universal M13F and M13R sequencing primers were used (Table 2.2).

Maxi-prep

Once the pCR2.1-TOPO-3.4 and pCR2.1-TOPO-3.8 constructs were confirmed by sequencing, plasmid DNA was transformed (Section 2.2.10) directly into competent cells and a single colony transferred to a large scale culture in order to amplify the plasmid. Plasmid DNA was then purified using QIAprep Spin Maxiprep kit (Section 2.2.14) and sequenced again to confirm the construct.

3.3.5 DNA preparation for sub-cloning into pFuse construct

Restriction digest of TOPO constructs

The DNA inserts were cut out of the TOPO constructs by restriction digest. The digestion reaction consisted of the following solution:

- Restriction enzyme 1 — 1 µl
- Restriction enzyme 2 — 1 µl
- Compatible Buffer — 5 µl
- BSA (10x) — 5 µl
- pCR2.1-TOPO-DBL construct — 1 µg
- dH_2O — up to 50 µl

For the pCR2.1-TOPO-3.4 construct restriction enzymes NcoI and BglII were used with NEB buffer 3.1. For the pCR2.1-TOPO-3.8 construct restriction enzymes NcoI and

EcoRI were used with NEB buffer 3.1. Restriction digests were incubated at 37°C for 2 hours and the products separated by gel electrophoresis (Section 2.2.5). 3.4 DBL and 3.8 DBL inserts were identified, excised and gel purified (Section 2.2.6). DNA concentrations of the purified inserts were calculated using a nanodrop spectrophotometer (Section 2.2.15).

Restriction digest of pFUSE plasmid

The pFUSE plasmid was prepared for sub-cloning by restriction digest. Two preparations were made for sub-cloning in each of the DBL domains.

The digestion reaction consisted of the following solution:

- Restriction enzyme 1 — 1 µl
- Restriction enzyme 2 — 1 µl
- Compatible Buffer — 5 µl
- BSA (10x) — 5 µl
- pFUSE-hIgG1-Fc2 — 1 µg
- dH_2O — up to 50 µl

For preparation of sub-cloning with the 3.4 DBL domain, restriction enzymes NcoI and BglII were used with NEB buffer 3.1. For preparation of sub-cloning with the 3.8 DBL domain, restriction enzymes NcoI and EcoRI were used with NEB buffer 3.1. Restriction digests were incubated at 37°C for 2 hours and the products separated by gel electrophoresis (Section 2.2.5). Digested vectors were identified, excised and gel purified (Section 2.2.6). DNA concentrations of the purified vectors were calculated using a nanodrop spectrophotometer (Section 2.2.15).

pFUSE vector de-phosphorylation

In order to prevent self-ligation of the pFUSE vector, the digested vectors were de-phosphorylated using Shrimp Alkaline phosphatase (rSAP) prior to gel elution.

The following reaction was setup on ice:

- Vector 5µg
- Shrimp alkaline phosphatase 20 µl
- Cutsmart buffer 20 µl
- dH_2O up to 200 µl.

The reaction was held at 37°C in a water bath for 1 hour before placing on a heat block set to 65°C for 30 minutes to deactivate the phosphatase.

3.3.6 pFUSE ligation to the 3.4 DBL and 3.8 DBL domains

The following reaction with a molar ratio of 1:3 vector to insert was set-up on ice:

- 10x T4 DNA ligase buffer 2 µl
- Vector DNA (4.2 kb) 31 ng
- Insert DNA (860 bp) 19 ng
- T4 DNA ligase 1 µl
- dH_2O up to 20 µl.

The ligation reaction was held at 16° for 16 hours using a Progene thermocycler.

3.3.7 Transformation of competent cells with pFUSE-3.4 and pFUSE-3.8

A 50 µl vial of competent TOP10 cells was defrosted on ice. 5 µl of the ligated construct was added and the cells incubated on ice for 30 mins. The cells were given a heat shock in a water bath at 42°C for 30 seconds and then returned to ice for 2 mins. 250 µl of S.O.C media was added and then the cells shaken at 225 rpm, 37°C for 1 hour.

Agarose plates containing 25 µg/ml zeocin (Section 2.1.16) were prepared and pre-warmed to 37°C. 5–200 µl of the transformation reaction was spread onto the agarose plates and plates incubated at 37°C overnight for colonies to form.

Restriction digest analysis of DNA from positive colonies

Single colonies were grown overnight at 37°C in 5 ml cultures with vigorous shaking (225 rpm) and DNA purified by mini-prep (Section 2.2.11). DNA was then digested using restriction enzymes NcoI and BglII with NEB buffer 3.1 for the pFuse-3.4 construct, and NcoI and EcoRI with NEB buffer 3.1 for the pFuse-3.8 construct.

The digestion reaction consisted of the following solution:

- Restriction enzyme 1 1 µl
- Restriction enzyme 2 1 µl
- Compatible Buffer 5 µl
- BSA (10x) 5 µl
- pFUSE-DBL construct 1 µg
- dH_2O up to 50 µl.

Restriction digests were incubated at 37°C for 2 hours and then the product visualised by gel electrophoresis (Section 2.2.5). DNA preparations with the correct inserts were identified and sent for DNA sequencing to confirm (Section 2.2.13).

Maxi-prep

1 µl DNA containing the confirmed pFUSE-DBL constructs was transformed directly into TOP10 cells and colonies picked and grown for 8 hours at 37°C (225 rpm) in 5 ml cultures before being transferred to 250 ml cultures and grown overnight under the same conditions. The DNA was then extracted using QIAprep Spin Maxiprep kit (Section 2.2.14) and sequenced (Section 2.2.13).

Primers for sequencing of the pFUSE constructs

Primers for the sequencing of the pFUSE constructs were designed using Primer3Plus `http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/` (Table 2.2).

3.3.8 Expression of 3.4 and 3.8 DBL recombinant protein

Once the DBL-Fc fusion constructs had been made they were expressed in mammalian cells.

CHO-KI cell transfection

CHO-KI cells were prepared for transfection (Section 2.3.1) by growing cultures until the cells were $50 - 80\%$ confluent. Cells were transfected using the FuGENE 6 Transfection Reagent. In a sterile tube containing 3 µg of plasmid DNA, 100 µl of medium was added and 9 µl of FuGENE 6 was added directly to the mixture. The tube was vortexed gently to mix and incubated at room temperature for 15 mins.

5 x 10^5 cells were added to 10 ml DMEM and, following incubation, the DNA/FuGENE was added to the CHO-KI cells, mixed and added to tissue-culture plates. 24 hours after transfection, the medium was replaced with DMEM selection media (Section 2.1.15).

Production of polyclonal population

After 7 - 9 days, transfected cells were passaged to expand the polyclonal population. Cells were washed in 10 ml sterile PBS before coating with trypsin and incubating for 5 mins at 37°C. Cells were resuspended in DMEM selection medium and expanded in small then medium-sized flasks. Polyclonal populations were frozen down for long-term storage when 80% confluent (Section 2.3.5).

Production of monoclonal population

Polyclonal populations were diluted to 10 cells/ml and 100 µl plated in each well of 96 well plate. 10 µl of supernatant from wells where single cells could be seen to have formed a colony was harvested after 9 days, and protein detected by immunoblot analysis (Section 2.4.2). The colonies found to be producing the most protein were expanded further. A single colony was transferred to a single well of a 24-well plate before being expanded to a small flask, and subsequently into medium flasks to expand the clonal cell line. Finally, clonal cell lines were cultured in 4 large tissue-culture flasks for 10 days to produce 200 ml supernatant containing protein from the monoclonal population. Timescales for expanding the single colony up to medium flasks was flexible depending on the rate of cell growth. However, all cultures seeded into large flasks were cultured for a period of exactly 10 days

to produce protein.

FPLC purification of recombinant DBL protein

Cell supernatants containing protein were filter sterilised by vacuum using Steritop GP Sterlisation Unit 0.22 μm PES membrane (Millipore). Sterile cell supernatants were purified using the liquid chromatography system ÄKTA-FPLC Frac-950 (Amersham Biosciences/GE Healthcare, Uppsala, Sweden). A 1 ml protein-G sepharose HiTrap column (GE Healthcare) was used and equilibrated by washing with a minimum 20 ml of HBSS buffer. Supernatant was loaded onto the column at a flow rate < 0.5 ml/min. Once fully loaded, the column was thoroughly washed with a minimum 30 ml of HBSS buffer to remove any unbound sample. Protein was eluted using an increasing gradient of 0.1M glycine-HCl, pH 2.5 and collected in 2 ml fractions. Eluates were neutralised with 1M Tris-HCl pH 9 and stored with 0.2% sodium azide at 4°C.

Immunoblot to test for the presence of recombinant DBL protein

FPLC fractions were tested for the presence of recombinant DBL protein using immunoblot analysis. 5 μl of each fraction was blotted onto nitrocellulose and blocked with PBS-tween plus 5% milk for 2 hours at 37°C on a rocker. The nitrocellulose was washed in PBST and anti-human IgG (Fc-specific) Alkaline Phosphatase antibody (Sigma A9544) was added at a dilution of 1:5,000 and incubated overnight at 37°C on a rocker. The blot was developed using NBT/BCIP tablets (Table 2.12) according to manufacturer's instructions. Fractions containing protein were combined and concentrated by ultrafiltration (Section 2.4.3). The flow-through liquid was also tested for protein; if protein was still present in the flow-through it was passed over the column again in order to extract all the sample.

3.3.9 ELISA for Fc detection

Duplicate wells of Nunc-maxisorp micrometer plates were coated with 100 μl of αhIgG-Fc (Life Technologies H1000) at a dilution of 1:1000 in 0.5 M carbonate/bicarbonate buffer (pH 9.6) and incubated at 4°C overnight. Any uncoated material was then discarded and the plates were blocked with 200 μl blocking buffer (Section 2.1.8) and incubated for 1 hour at room temperature on a rocker. Wells were washed in PBST (PBS + 0.05% Tween 20) 5 times before the addition of tissue culture supernatants and IVIG in duplicate wells.

Supernatants were coated in pure, 1:10 and 1:100 dilutions in media and intravenous immunoglobulin (IVIG) was coated in doubling dilutions starting at 50 µg/ml and reducing to 0.19 µg/ml. The plate was incubated at room temperature for 2 hours on a platform shaker. Wells were washed in PBST 5 times before the addition of 100 µl per well of detecting antibody αhIgG(Fc specific)-HRP (A0170, Sigma) at a dilution of 1:500 in PBS. The plate was incubated for 1 hour at room temperature on a platform shaker. Plates were washed as before and then developed using 3'3'5'5'tetramethylbenzidine dihydrochloride (one tablet in 10 ml phosphate citrate buffer). 100 µl of substrate buffer was added to each well and the plate was allowed to develop for 10 - 20 mins. 50 µl H_2SO_4 was added per well to stop the reaction. The absorbance was read at 450 nm on an Expert plus plate reader (Biochrom, SLS).

Protein concentrations were calculated by plotting absorbance on the IVIG standard curve and adjusting for dilution.

3.3.10 IgM-binding ELISA

Duplicate wells of Nunc-maxisorp micrometer plates were coated with 100 µl of unconjugated αhIgM (Sigma I6385) at a dilution of 1:5000 in 0.5 M carbonate/bicarbonate buffer (pH 9.6) and incubated at 4°C overnight. Any uncoated material was then discarded and the plates blocked with 200 µl blocking buffer (Section 2.1.8) and incubated for 1 hour at room temperature on a rocker. Wells were washed in PBST 5 times before the addition of hIgM at a concentration of 10 µg/ml in PBS. The plate was incubated for 2 hours at room temperature on a rocker before washing all wells 5 times in PBST. DBL domain proteins were added at a range of dilutions in 0.1 M Glycine Tris-HCl pH 7.0. The plate was further incubated for 2 hours at room temperature on a rocker. Wells were washed 5 times in PBST before the addition of αhIgG-AP (Sigma A9544) at a dilution of 1:5000 in PBS, and incubated for 1 hour at room temperature. Plates were then washed as before and then developed using SIGMA$FAST^{TM}$ p-Nitrophenyl Phosphate tablets (20ml solution containing 1.0 mg/ml pNPP, 0.2 M Tris buffer and 5 mM magnesium chloride). 100 µl of substrate buffer was added to each well and the plate was allowed to develop. The absorbance was read at 405 nm.

Protein concentration was plotted against absorbance for each DBL domain. This enabled binding to be compared directly between different DBL domains.

3.3.11 Western blot

Proteins were run on a NuPAGE Novex 4-12% Bis-Tris pre-cast gel (Section 2.4.6) and transferred to nitrocellulose for Western blotting (Section 2.4.8). After blocking for 2 hours the membrane was washed three times.

To detect the presence of the IgG-Fc, the membrane was incubated with αhIgG-Fc-AP (Sigma A9544) at a dilution 1:5000 in PBST overnight at 37°C on a rocker. The membrane was again washed three times before proteins were visualised by developing BCIP/NBT tablets (Sigma) according to manufacturer's instructions.

3.3.12 Gel filtration of recombinant DBL domains

Recombinant DBL proteins were separated by gel filtration using the Superdex 200 10/300 GL column (GE Healthcare) on the liquid chromatography system ÄKTA-FPLC Frac-950. The column was equilibrated by washing with a minimum of 50 ml of 20 mM sodium phosphate pH 7.0. Lyophilised gel filtration standards (Bio-Rad) were prepared to manufacturer's instructions and 0.1 - 0.5 ml injected at a flow rate of < 0.85 ml/min. Fractions were eluted and collected, and peaks corresponding to the molecular weight markers were noted (Figure 2.2). The column was washed again with 50 ml of 20 mM sodium phosphate pH 7.0 to remove any trace of the markers. A volume of between 0.1 - 0.5 ml DBL-Fc construct were then added to the column and proteins separated and collected using the same flow rate as the markers.

3.4 Results

3.4.1 Production of pCR2.1-TOPO-DBL constructs

DBL 3.4 and DBL 3.8 were first cloned into the PCR2.1-TOPO vector (Invitrogen, UK). TOPO TA cloning was chosen so that the DNA could be quickly inserted into the plasmid. Amplification with Taq polymerase adds a single Adenosine (A) to the 3' end. The vector PCR2.1-TOPO contains single 3'-thymidine (T) overhangs and is activated with Topoisomerase I so that the Taq-amplified DNA can be inserted directly. This produced a stock of cloning plasmid quickly, reliably and at a low cost which was used for further sub-cloning.

The DBL domains were amplified from the full-length gene and flanking restriction sites introduced by PCR to enable sub-cloning into the pFUSE vector (Section 3.3.2 for methods). For DBL 3.4, BglII and NcoI sites were used and for DBL 3.8, EcoRI and NcoI sites were used. These restriction sites were also used for checking the constructs.

3.4.1.1 Restriction Digest

Restriction digest of DNA extracted from white colonies after transformation of competent cells confirms that the DBL domains have been inserted in the correct orientation into the pCR2.1-TOPO plasmid. The constructs produced have been named pCR2.1-TOPO-DBL3.4 and pCR2.1-TOPO-DBL3.8 (See Appendix A.2 and A.3 for plasmid maps). When cut by restriction digest (NcoI and BglII for DBL 3.4 and EcoRI and NcoI for DBL 3.8), the plasmids have fragments of size 596bp, 872bp, 994bp, 2347bp and 869bp, 1578bp, 2334bp respectively (Figure 3.6) which is as predicted.

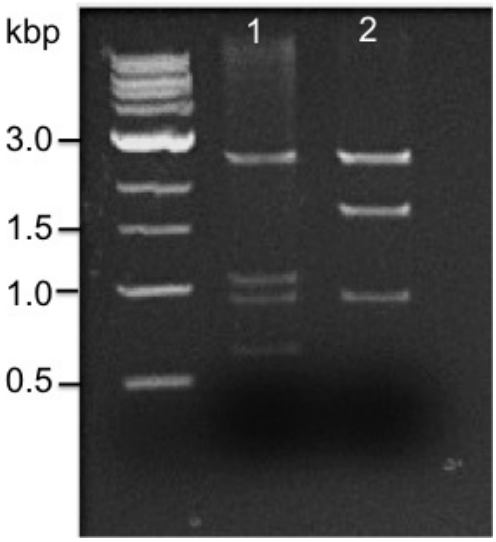


Figure 3.6: **Lane 1:** pCR2.1-TOPO-DBL3.4 digested with restriction sites NcoI and BglII. The restriction digest releases fragments of 596bp, 872bp, 994bp and 2347bp from the 4809bp plasmid. **Lane 2:** pCR2.1-TOPO-DBL3.8 digested with restriction sites EcoRI and NcoI. The restriction digest releases fragments of 869bp, 1578bp and 2335bp, from the 4782bp plasmid.

3.4.1.2 Sequencing Results

pCR2.1-TOPO-DBL3.4 sequencing results show the nucleotide sequence of the positive colony selected from transformation of the construct into *E.coli* cells and identified by restriction digest. The sequence is aligned against the 3.4 DBL domain nucleotide sequence (Figure 3.7). Restriction sites NcoI and BglII are shown in red. The sequences have 100% sequence identity.

```
3.4 DBL         ----------------------------------------TTCTGCAAGGGCATCAA
3.4 +VE         AACGGCCGCCAGTGTGCTGGAATTCGGCTTATCGGCCATGGTTTTCTGCAAGGGCATCAA
                                                        ******************

3.4 DBL         GAACGTGCTGAGCTGCCCTCCCAAGAACAGCGCCGGCAGAAACGGCGACTGGATCAGCGT
3.4 +VE         GAACGTGCTGAGCTGCCCTCCCAAGAACAGCGCCGGCAGAAACGGCGACTGGATCAGCGT
                ************************************************************

3.4 DBL         GGCCGTGAAAGAGAGCAGCACCACCAACAAGGGCGTGCTGGTGCCCCCCAGACGGACCAA
3.4 +VE         GGCCGTGAAAGAGAGCAGCACCACCAACAAGGGCGTGCTGGTGCCCCCCAGACGGACCAA
                ************************************************************

3.4 DBL         GCTGTGCCTGCGGAACATCAACAAAGTGTGGCACCGGATCAAGGACGAGAAGAACTTCAA
3.4 +VE         GCTGTGCCTGCGGAACATCAACAAAGTGTGGCACCGGATCAAGGACGAGAAGAACTTCAA
                ************************************************************

3.4 DBL         AGAGGAATTCGTCAAGGTCGCACTGGGCGAGAGCAACGCCCTGATGAAGCACTACAAAGA
3.4 +VE         AGAGGAATTCGTCAAGGTCGCACTGGGCGAGAGCAACGCCCTGATGAAGCACTACAAAGA
                ************************************************************

3.4 DBL         GAAGAACCTGAACGCCCTGACCGCCATTAAGTACGGCTTCAGCGACATGGGCGACATCAT
3.4 +VE         GAAGAACCTGAACGCCCTGACCGCCATTAAGTACGGCTTCAGCGACATGGGCGACATCAT
                ************************************************************

3.4 DBL         CAAGGGCACCGACCTGATCGATTACCAGATCACCAAGAACATCAACCGGGCCCTGGACAA
3.4 +VE         CAAGGGCACCGACCTGATCGATTACCAGATCACCAAGAACATCAACCGGGCCCTGGACAA
                ************************************************************

3.4 DBL         GATCCTGAGAAACGAGGCCAGCAACGACAAGATCAAGAAACGGGTGGACTGGTGGGAGGC
3.4 +VE         GATCCTGAGAAACGAGGCCAGCAACGACAAGATCAAGAAACGGGTGGACTGGTGGGAGGC
                ************************************************************

3.4 DBL         CAACAAGGCCGCCTTCTGGGACGCCTTCATGTGCGGCTACAAGGTGCACATCGGCAACAA
3.4 +VE         CAACAAGGCCGCCTTCTGGGACGCCTTCATGTGCGGCTACAAGGTGCACATCGGCAACAA
                ************************************************************

3.4 DBL         GCCCTGCCCCGAGCACGACAACATGGACCGGATCCCCCAGTACCTGCGGTGGTTTCGCGA
3.4 +VE         GCCCTGCCCCGAGCACGACAACATGGACCGGATCCCCCAGTACCTGCGGTGGTTTCGCGA
                ************************************************************

3.4 DBL         GTGGGGCACCTACGTGTGCAGCGAGTACAAGAACAAGTTCGAGGACGTGATCAAGCTGTG
3.4 +VE         GTGGGGCACCTACGTGTGCAGCGAGTACAAGAACAAGTTCGAGGACGTGATCAAGCTGTG
                ************************************************************

3.4 DBL         CAACATCCAGCAGTTCACCAACCAGGACGACAGCCAGCTGCTGGAAATCAGCAAGAAAGA
3.4 +VE         CAACATCCAGCAGTTCACCAACCAGGACGACAGCCAGCTGCTGGAAATCAGCAAGAAAGA
                ************************************************************

3.4 DBL         CAAGTGCAAAGAGGCCCTGAAACACTACGAGGAATGGGTCAACCGGCGGAGGCCCGAGTG
3.4 +VE         CAAGTGCAAAGAGGCCCTGAAACACTACGAGGAATGGGTCAACCGGCGGAGGCCCGAGTG
                ************************************************************

3.4 DBL         GAAGGGCCAGTGCGATAAGTTCGAGAAAGAGAAGTCTAAGTACGAGGACACCAAGAGCAT
3.4 +VE         GAAGGGCCAGTGCGATAAGTTCGAGAAAGAGAAGTCTAAGTACGAGGACACCAAGAGCAT
                ************************************************************

3.4 DBL         CACCGCCGAGAAGTACCTGAAAGAAATCTGCAGCGAGTGCGACTGCAAGTACAAGGACCT
3.4 +VE         CACCGCCGAGAAGTACCTGAAAGAAATCTGCAGCGAGTGCGACTGCAAGTACAAGGACCT
                ************************************************************

3.4 DBL         GGACAAC-------------------------------------------------
3.4 +VE         GGACAACAGATCTGACAAAACAAGCCGAATTCTGCAGATATCCATCACACTGGCGGCCGC
                *******
```

Figure 3.7: Alignment of pCR2.1-TOPO-DBL3.4 nucleotides with 3.4 DBL shows 100% sequence identity.

```
3.8 DBL       ------------------------------------------------------------
3.8 +VE       TAGTAACGGCCGCCAGTGTGCTGGAATTCGGATCCACTAGTAACGGCCGCCAGTGTGCTG


3.8 DBL       ---------------------ATCTGCAAGGACTTCAGCAACCTGCCCCAGTGCCGGAAG
3.8 +VE       GAATTCGCCCTTCGGAATTCGATCTGCAAGGACTTCAGCAACCTGCCCCAGTGCCGGAAG
                                    ************************************

3.8 DBL       AACGTGCACGAGCGGAACAACTGGCTGGGCAGCAGCGTGAAGAACTTCGCCAGCGACAAC
3.8 +VE       AACGTGCACGAGCGGAACAACTGGCTGGGCAGCAGCGTGAAGAACTTCGCCAGCGACAAC
              ************************************************************

3.8 DBL       AAGGGCGTGCTGGTGCCCCCCAGACGGCAGAGCCTGTGCCTGAGAATCACCCTGCAGGAC
3.8 +VE       AAGGGCGTGCTGGTGCCCCCCAGACGGCAGAGCCTGTGCCTGAGAATCACCCTGCAGGAC
              ************************************************************

3.8 DBL       TTCCGGACCAAGAAGAAGAAAGAGGGCGACTTCGAGAAGTTCATCTACAGCTACGCCAGC
3.8 +VE       TTCCGGACCAAGAAGAAGAAAGAGGGCGACTTCGAGAAGTTCATCTACAGCTACGCCAGC
              ************************************************************

3.8 DBL       AGCGAGGCCCGGAAGCTGCGGACCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCC
3.8 +VE       AGCGAGGCCCGGAAGCTGCGGACCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCC
              ************************************************************

3.8 DBL       ATCCGGTACAGCTTCGCCGACATCGGCAACATCATCCGGGGCGACGACATGATGGACACC
3.8 +VE       ATCCGGTACAGCTTCGCCGACATCGGCAACATCATCCGGGGCGACGACATGATGGACACC
              ************************************************************

3.8 DBL       CCCACCAGCAAAGAGACAATCACCTATCTGGAAAAGGTGCTGAAGATCTACAATGAGAAC
3.8 +VE       CCCACCAGCAAAGAGACAATCACCTATCTGGAAAAGGTGCTGAAGATCTACAATGAGAAC
              ************************************************************

3.8 DBL       AACGACAAGCCCAAGGACGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTGGGAG
3.8 +VE       AACGACAAGCCCAAGGACGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTGGGAG
              ************************************************************

3.8 DBL       GCCATGATGTGCGGCTACCAGAGCGCCCAGAAAGACAACCAGTGCACCGGCTACGGCAAC
3.8 +VE       GCCATGATGTGCGGCTACCAGAGCGCCCAGAAAGACAACCAGTGCACCGGCTACGGCAAC
              ************************************************************

3.8 DBL       ATCGACGATATCCCCCAGTTCCTGCGGTGGTTCCGCGAGTGGGGCACCTACGTGTGCGAA
3.8 +VE       ATCGACGATATCCCCCAGTTCCTGCGGTGGTTCCGCGAGTGGGGCACCTACGTGTGCGAA
              ************************************************************

3.8 DBL       GAGTCCGAGAAGAACATGAATACCCTGAAGGCCGTGTGCTTCCCCAAGCAGCCCAGAACC
3.8 +VE       GAGTCCGAGAAGAACATGAATACCCTGAAGGCCGTGTGCTTCCCCAAGCAGCCCAGAACC
              ************************************************************

3.8 DBL       GAGGCCAACCCTGCCCTGACCGTGCACGAGAACGAGATGTGCAGCAGCACCCTGAAGAAA
3.8 +VE       GAGGCCAACCCTGCCCTGACCGTGCACGAGAACGAGATGTGCAGCAGCACCCTGAAGAAA
              ************************************************************

3.8 DBL       TACGAGGAATGGTACAACAAGCGCAAGACCGAGTGGACCGAGCAGAGCATCAAGTATAAC
3.8 +VE       TACGAGGAATGGTACAACAAGCGCAAGACCGAGTGGACCGAGCAGAGCATCAAGTATAAC
              ************************************************************

3.8 DBL       AACGATAAGATCAACTACGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCGAG
3.8 +VE       AACGATAAGATCAACTACGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCGAG
              ************************************************************

3.8 DBL       AAGTGCCCCGAGTGCAAGTGCACCAAGAAAAACCTGCAGGAC------------------
3.8 +VE       AAGTGCCCCGAGTGCAAGTGCACCAAGAAAAACCTGCAGGACGCCATGGTTAGATCTGAC
              ******************************************

3.8 DBL       ---------------------------------
3.8 +VE       AAAGCCGAATTCTGCAGATATCCATCACACTGG
```

Figure 3.8: Alignment of pCR2.1-TOPO-DBL3.8 nucleotides with 3.8 DBL shows 100% sequence identity.

pCR2.1-TOPO-DBL3.8 sequencing results show the sequenced nucleotide sequence of the positive colony selected from transformation of the construct into *E.coli* cells and identified by restriction digest. The sequence is aligned against the confirmed DBL domain

nucleotide sequence (Figure 3.8). Restriction sites EcoRI and NcoI are shown in red. The sequences have 100% sequence identity.

Alignment of the protein sequences translated from sequencing results with the known DBL sequences confirms the correct insertion of the DBL domain inserts into the pCR2.1-TOPO plasmid (Figure 3.9). The pCR2.1-TOPO-3.4 and pCR2.1-TOPO-3.8 constructs have successfully been produced and are ready for sub-cloning.

```
3.4 TOPO    FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK
            FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK
3.4 DBL     FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK

3.4 TOPO    DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI
            DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI
3.4 DBL     DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI

3.4 TOPO    NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY
            NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY
3.4 DBL     NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY

3.4 TOPO    LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN
            LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN
3.4 DBL     LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN

3.4 TOPO    RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN
            RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN
3.4 DBL     RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN
```

(a)

```
3.8 TOPO    ICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTKKKK
            ICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTKKKK
3.8 DBL     ICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTKKKK

3.8 TOPO    EGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETI
            EGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETI
3.8 DBL     EGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETI

3.8 TOPO    TYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQF
            TYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQF
3.8 DBL     TYLEKVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQF

3.8 TOPO    LRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNK
            LRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNK
3.8 DBL     LRWFREWGTYVCEESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNK

3.8 TOPO    RKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQD
            RKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQD
3.8 DBL     RKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQD
```

(b)

Figure 3.9: Alignment of (a) pCR2.1-TOPO-DBL3.4 and (b) pCR2.1-TOPO-DBL3.8 amino acids with the known DBL amino acid sequences.

3.4.2 Production of pFUSE constructs

The pCR2.1-TOPO-DBL3.4 and pCR2.1-TOPO-DBL3.8 constructs were used to sub-clone DBL 3.4 and DBL 3.8 into the pFuse-hIgG1-Fc2 plasmid to create the expression vectors pFuse-hIgG1-Fc2-DBL3.4 and pFuse-hIgG1-Fc2-DBL3.8.

Firstly, the DBL domain was cut out of the pCR2.1-TOPO-DBL3.4 plasmid using restriction enzymes NcoI and BglII while the pFuse-hIgG1-Fc2 plasmid was prepared by digestion using the same restriction enzymes. Similarly, the DBL domain was cut out of the pCR2.1-TOPO-DBL3.8 plasmid using restriction enzymes NcoI and EcoRI while the pFuse-hIgG1-Fc2 plasmid was also prepared by digestion using NcoI and EcoRI (Section 3.3.5). Digested inserts were separated from TOPO plasmid by gel electrophoresis and purified (Section 2.2.6) while digested pFUSE-hIgG1-Fc2 plasmid was also purified from agarose. The cut vector was de-phosphorylated to prevent self-ligation (Section 3.3.5).

Prepared pFUSE-hIgG1-Fc2 plasmid was ligated to the prepared DBL inserts (Section 3.3.6).

3.4.2.1 Screening of positive colonies

The ligation product was transformed into TOP10 competent cells (Section 2.2.10) and colonies screened for the insert using restriction digest (Figure 3.10). Most colonies were positive as the inserts of 864 bp or 861 bp were seen by restriction digest. Of these, three colonies were selected and sequenced to confirm the result.

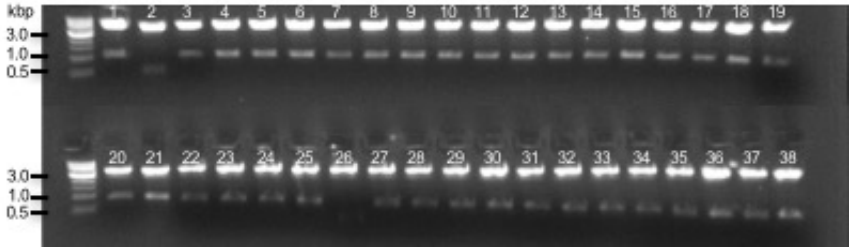


Figure 3.10: **Restriction digest to screen colonies grown from transformation of pFUSE constructs.** Lanes 1 - 19: pFUSE 3.4 colonies. Lanes 20-38: pFUSE 3.8 colonies. Colonies are positive if they contain the DBL insert (864 bp or 861 bp). Lanes 1 and 3-19 contain positive 3.4 colonies and lanes 20-21 and 23-38 contain positive 3.8 colonies.

3.4.2.2 Sequencing Results

Sequencing results show the pFUSE-hIgG1-Fc2-DBL3.4 nucleotide sequence aligned against the pFUSE-hIgG1-Fc2-DBL3.4 master nucleotide sequence (Figure 3.11) and the pFUSE-hIgG1-Fc2-DBL3.8 nucleotide sequence aligned against the pFUSE-hIgG1-Fc2-DBL3.8 master nucleotide sequence (Figure 3.12). In both cases the sequences have 100% sequence identity.

```
+ve colony   ACCATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAAT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   ACCATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAAT  169

+ve colony   TCGATATCGGCCATGGTTTTCTGCAAGGGCATCAAGAACGTGCTGAGCTGCCCTCCCAAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   TCGATATCGGCCATGGTTTTCTGCAAGGGCATCAAGAACGTGCTGAGCTGCCCTCCCAAG  229

+ve colony   AACAGCGCCGGCAGAAACGGCGACTGGATCAGCGTGGCCGTGAAAGAGAGCAGCACCACC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   AACAGCGCCGGCAGAAACGGCGACTGGATCAGCGTGGCCGTGAAAGAGAGCAGCACCACC  289

+ve colony   AACAAGGGCGTGCTGGTGCCCCCCAGACGGACCAAGCTGTGCCTGCGGAACATCAACAAA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   AACAAGGGCGTGCTGGTGCCCCCCAGACGGACCAAGCTGTGCCTGCGGAACATCAACAAA  349

+ve colony   GTGTGGCACCGGATCAAGGACGAGAAGAACTTCAAAGAGGAATTCGTCAAGGTCGCACTG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   GTGTGGCACCGGATCAAGGACGAGAAGAACTTCAAAGAGGAATTCGTCAAGGTCGCACTG  409

+ve colony   GGCGAGAGCAACGCCCTGATGAAGCACTACAAAGAGAAGAACCTGAACGCCCTGACCGCC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   GGCGAGAGCAACGCCCTGATGAAGCACTACAAAGAGAAGAACCTGAACGCCCTGACCGCC  469

+ve colony   ATTAAGTACGGCTTCAGCGACATGGGCGACATCATCAAGGGCACCGACCTGATCGATTAC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   ATTAAGTACGGCTTCAGCGACATGGGCGACATCATCAAGGGCACCGACCTGATCGATTAC  529

+ve colony   CAGATCACCAAGAACATCAACCGGGCCCTGGACAAGATCCTGAGAAACGAGGCCAGCAAC  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   CAGATCACCAAGAACATCAACCGGGCCCTGGACAAGATCCTGAGAAACGAGGCCAGCAAC  589

+ve colony   GACAAGATCAAGAAACGGGTGGACTGGTGGGAGGCCAACAAGGCCGCCTTCTGGGACGCC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   GACAAGATCAAGAAACGGGTGGACTGGTGGGAGGCCAACAAGGCCGCCTTCTGGGACGCC  649

+ve colony   TTCATGTGCGGCTACAAGGTGCACATCGGCAACAAGCCCTGCCCCGAGCACGACAACATG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   TTCATGTGCGGCTACAAGGTGCACATCGGCAACAAGCCCTGCCCCGAGCACGACAACATG  709

+ve colony   GACCGGATCCCCCAGTACCTGCGGTGGTTTCGCGAGTGGGGCACCTACGTGTGCAGCGAG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   GACCGGATCCCCCAGTACCTGCGGTGGTTTCGCGAGTGGGGCACCTACGTGTGCAGCGAG  769

+ve colony   TACAAGAACAAGTTCGAGGACGTGATCAAGCTGTGCAACATCCAGCAGTTCACCAACCAG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   TACAAGAACAAGTTCGAGGACGTGATCAAGCTGTGCAACATCCAGCAGTTCACCAACCAG  829

+ve colony   GACGACAGCCAGCTGCTGGAAATCAGCAAGAAAGACAAGTGCAAAGAGGCCCTGAAACAC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   GACGACAGCCAGCTGCTGGAAATCAGCAAGAAAGACAAGTGCAAAGAGGCCCTGAAACAC  889

+ve colony   TACGAGGAATGGGTCAACCGGCGGAGGCCCGAGTGGAAGGGCCAGTGCGATAAGTTCGAG  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   TACGAGGAATGGGTCAACCGGCGGAGGCCCGAGTGGAAGGGCCAGTGCGATAAGTTCGAG  949

+ve colony   AAAGAGAAGTCTAAGTACGAGGACACCAAGAGCATCACCGCCGAGAAGTACCTGAAAGAA  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   AAAGAGAAGTCTAAGTACGAGGACACCAAGAGCATCACCGCCGAGAAGTACCTGAAAGAA  1009

+ve colony   ATCTGCAGCGAGTGCGACTGCAAGTACAAGGACCTGGACAACAGATCTG  949
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.4 master   ATCTGCAGCGAGTGCGACTGCAAGTACAAGGACCTGGACAACAGATCTG  1058
```

Figure 3.11: **pFUSE-hIgG1-Fc2-DBL3.4 sequencing results.** The sequencing results of DNA extracted from the positive colony had 100% sequence identity with the pFUSE-hIgG1-Fc2-DBL3.4 master sequence. The leader sequence is shown in black, the DBL sequence is purple and restriction sites (NcoI and BglII) shown in blue.

```
+ve colony   CACCATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CACCATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAA  168

+ve colony   TTCGATCTGCAAGGACTTCAGCAACCTGCCCCAGTGCCGGAAGAACGTGCACGAGCGGAA  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   TTCGATCTGCAAGGACTTCAGCAACCTGCCCCAGTGCCGGAAGAACGTGCACGAGCGGAA  228

+ve colony   CAACTGGCTGGGCAGCAGCGTGAAGAACTTCGCCAGCGACAACAAGGGCGTGCTGGTGCC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CAACTGGCTGGGCAGCAGCGTGAAGAACTTCGCCAGCGACAACAAGGGCGTGCTGGTGCC  288

+ve colony   CCCCAGACGGCAGAGCCTGTGCCTGAGAATCACCCTGCAGGACTTCCGGACCAAGAAGAA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CCCCAGACGGCAGAGCCTGTGCCTGAGAATCACCCTGCAGGACTTCCGGACCAAGAAGAA  348

+ve colony   GAAAGAGGGCGACTTCGAGAAGTTCATCTACAGCTACGCCAGCAGCGAGGCCCGGAAGCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   GAAAGAGGGCGACTTCGAGAAGTTCATCTACAGCTACGCCAGCAGCGAGGCCCGGAAGCT  408

+ve colony   GCGGACCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCCATCCGGTACAGCTTCGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   GCGGACCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCCATCCGGTACAGCTTCGC  468

+ve colony   CGACATCGGCAACATCATCCGGGGCGACGACATGATGGACACCCCCACCAGCAAAGAGAC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CGACATCGGCAACATCATCCGGGGCGACGACATGATGGACACCCCCACCAGCAAAGAGAC  528

+ve colony   AATCACCTATCTGGAAAAGGTGCTGAAGATCTACAATGAGAACAACGACAAGCCCAAGGA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   AATCACCTATCTGGAAAAGGTGCTGAAGATCTACAATGAGAACAACGACAAGCCCAAGGA  588

+ve colony   CGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTGGGAGGCCATGATGTGCGGCTA  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTGGGAGGCCATGATGTGCGGCTA  648

+ve colony   CCAGAGCGCCCAGAAAGACAACCAGTGCACCGGCTACGGCAACATCGACGATATCCCCCA  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CCAGAGCGCCCAGAAAGACAACCAGTGCACCGGCTACGGCAACATCGACGATATCCCCCA  708

+ve colony   GTTCCTGCGGTGGTTCCGCGAGTGGGGCACCTACGTGTGCGAAGAGTCCGAGAAGAACAT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   GTTCCTGCGGTGGTTCCGCGAGTGGGGCACCTACGTGTGCGAAGAGTCCGAGAAGAACAT  768

+ve colony   GAATACCCTGAAGGCCGTGTGCTTCCCCAAGCAGCCCAGAACCGAGGCCAACCCTGCCCT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   GAATACCCTGAAGGCCGTGTGCTTCCCCAAGCAGCCCAGAACCGAGGCCAACCCTGCCCT  828

+ve colony   GACCGTGCACGAGAACGAGATGTGCAGCAGCACCCTGAAGAAATACGAGGAATGGTACAA  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   GACCGTGCACGAGAACGAGATGTGCAGCAGCACCCTGAAGAAATACGAGGAATGGTACAA  888

+ve colony   CAAGCGCAAGACCGAGTGGACCGAGCAGAGCATCAAGTATAACAACGATAAGATCAACTA  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CAAGCGCAAGACCGAGTGGACCGAGCAGAGCATCAAGTATAACAACGATAAGATCAACTA  948

+ve colony   CGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCGAGAAGTGCCCCGAGTGCAA  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3.8 master   CGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCGAGAAGTGCCCCGAGTGCAA  1008

+ve colony   GTGCACCAAGAAAAACCTGCAGGACGCCATGG  932
             ||||||||||||||||||||||||||||||
3.8 master   GTGCACCAAGAAAAACCTGCAGGACGCCATGG  1040
```

Figure 3.12: **pFUSE-hIgG1-Fc2-DBL3.8 sequencing results.** The sequencing results of DNA extracted from the positive colony had 100% sequence identity with the pFUSE-hIgG1-Fc2-DBL3.8 master sequence. The leader sequence is shown in black, the DBL sequence is red and restriction sites (EcoRI and NcoI) shown in blue.

The translated amino acid sequence from the positive colonies is as expected from the master sequence (Figure 3.13).

MYRMQLLSCIALSLALVTNSISAMVFCKGIKNVLSCPPKNSAGRNGDWISVAV
KESSTTNKGVLVPPRRTKLCLRNINKVWHRIKDEKNFKEEFVKVALGESNALM
KHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDKILRNEASN
DKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREW
GTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN
RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKT
HTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFN
WYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP
APIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWE
SNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNH
YTQKSLSLSPGK*VLA

(a)

MYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFASDN
KGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNN
LEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKD
AKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVC
EESEKNMNTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEW
TEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQDAMVRSDKTHTC
PPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV
DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPI
EKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG
QPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQ
KSLSLSPGK*VLA

(b)

Figure 3.13: Translated amino acid sequences of DNA extracted from the positive colonies. The lead sequence of the pFUSE-hIgG1-Fc2 plasmid is in black. (a) The 3.4 DBL domain is in purple and the IgG-Fc sequence is in red. (b) The 3.8 DBL domain is in brown and the IgG-Fc sequence is in red.

3.4.2.3 Restriction Digest

Restriction digest confirms the construction of the pFUSE-hIgG1-Fc2-DBL3.4 and pFUSE-hIgG1-Fc2-DBL3.8 plasmids which, when cut by restriction digest, have inserts of size 864bp and 861bp respectively (Figure 3.14). This is the expected size from the plasmids produced.

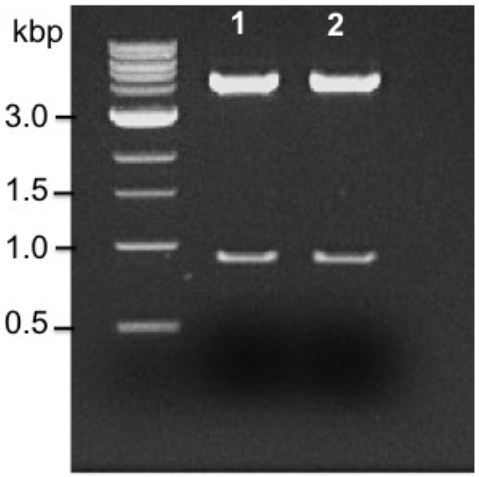


Figure 3.14: Bands observed from restriction digest are as expected for the plasmids produced. **Lane 1:** pFUSE-hIgG1-Fc2-DBL3.4 digested with restriction sites NcoI and BglII. The restriction digest releases a fragment of 864 bps from the 4192 bp plasmid. **Lane 2:** pFUSE-hIgG1-Fc2-DBL3.8 digested with restriction sites EcoRI and NcoI. The restriction digest releases a fragment of 861 bps from the 4196 bp plasmid.

3.4.3 Expression of recombinant MSPDBL1 and MSPDBL2 DBL domains

CHO–K1 cells were transfected with pFUSE-hIgG1-Fc2-DBL3.4 and pFUSE-higG1-Fc2-DBL3.8 plasmids using the FuGENE transfection reagent. Zeocin was used to select for successfully transfected cells and a monoclonal population of high expression cells produced (Section 3.3.8).

3.4.3.1 Immunoblotting

Cell lines grown from single cells were selected for high protein expression by immunoblotting (Figure 3.15) and further expanded. Cell supernatant was removed from each culture and filter sterilised after 10 days of selection.

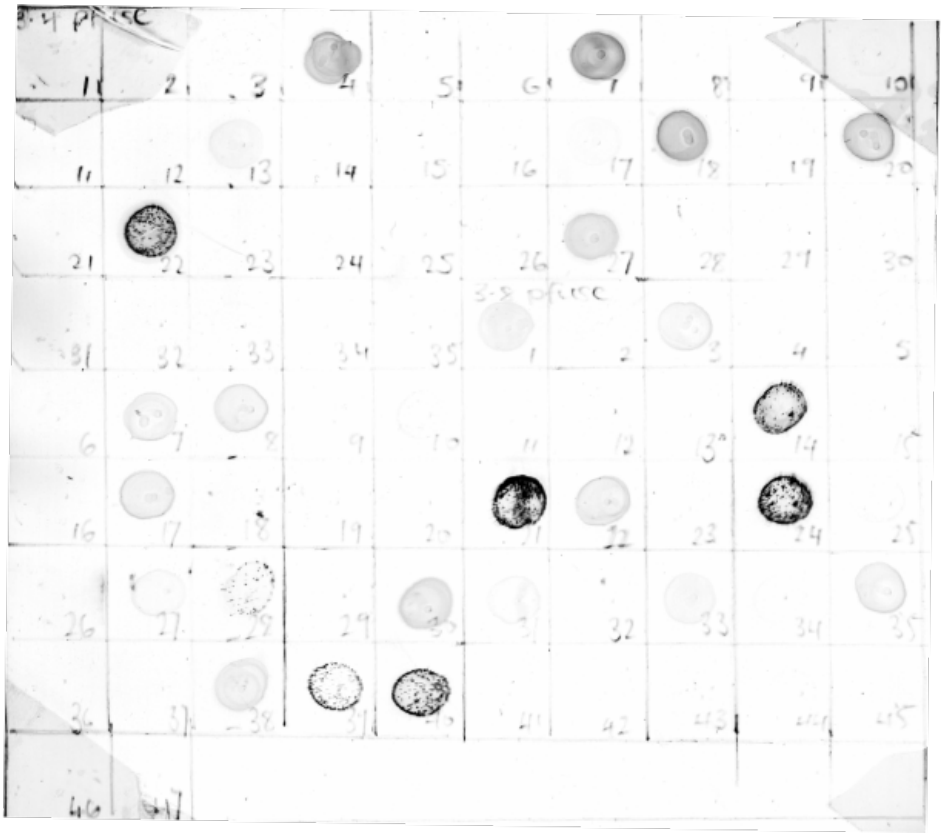
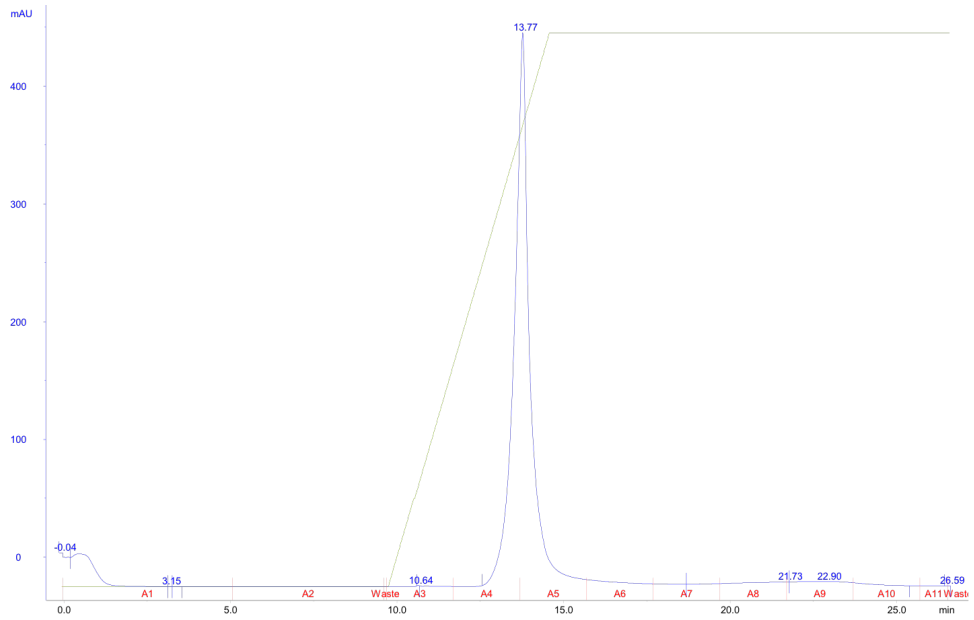


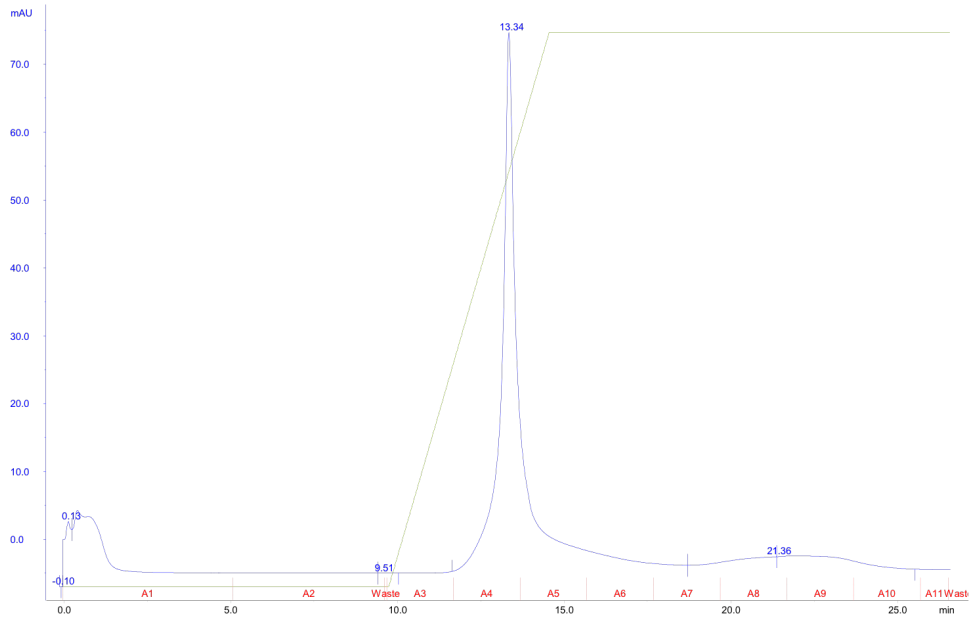
Figure 3.15: Single cells were added to a 96-well plate and after 9 days, colonies could be observed in some wells. 10 µl of cell supernatant was tested for protein expression by immunoblotting. For the pFUSE-3.4 transfection, 35 cell lines were produced and the highest two expressing protein were selected. Cell lines 7 and 22 were expanded to large cultures and cell supernatant with protein collected after 10 days. For the pFUSE-3.8 transfection, 47 cell lines were produced and the highest two expressing protein were selected. Cell lines 21 and 24 were expanded to large cultures and cell supernatant with protein collected.

3.4.3.2 FPLC protein purification

DBL-Fc recombinant protein was purified from supernatant by affinity chromatography, using a protein G sepharose column (Section 3.3.8). A linear gradient of 0.1M glycine-HCl, pH 2.7 was used to elute the protein from the column (Figure 3.16).



(a)



(b)

Figure 3.16: FPLC elution profile of (a) 3.4 DBL recombinant protein and (b) 3.8 DBL recombinant protein. UV profile of the eluted protein is shown in blue and linear glycine gradient is green. The recombinant protein comes off the sepharose beads as the glycine gradient increases. Peaks were collected in fractions which are shown in red.

Fractions were collected and immunoblotted with αhIgG-AP to show that the protein was present in the fraction corresponding to the FPLC peak (Figure 3.17) (Section 3.3.8). Fractions containing protein were pooled and concentrated by Vivaspin 2 ultrafiltration columns (Section 2.4.3).

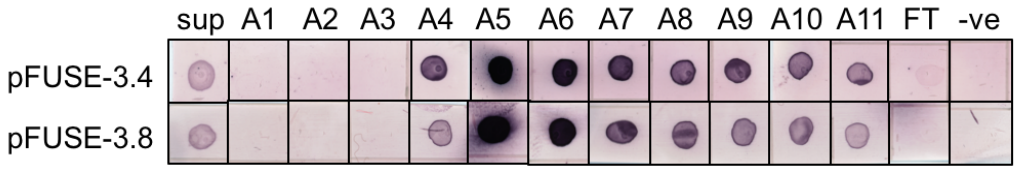


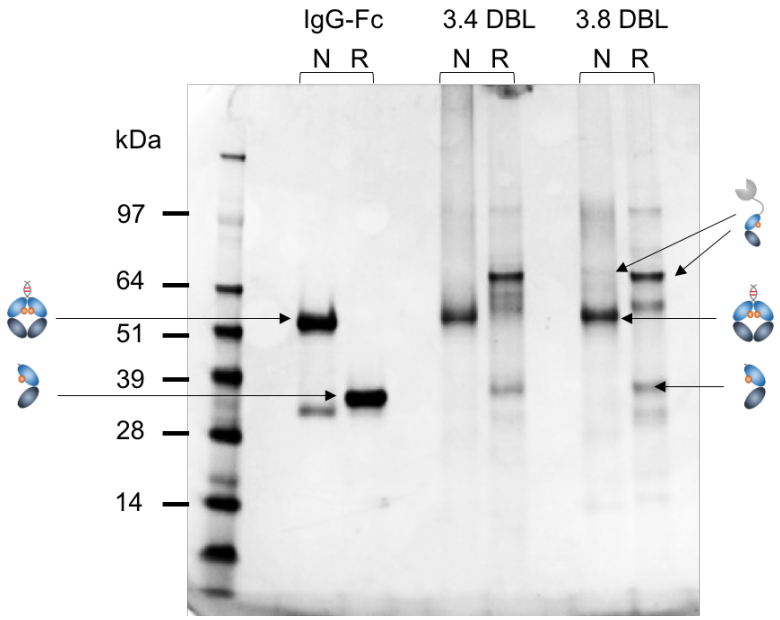
Figure 3.17: Fractions A1-A11 from the FPLC elution were tested by immunoblotting for the presence of protein. These were compared to a blot of the cell supernatant (sup) prior to purification as well as the flow-through (FT) to check that all protein had been removed as well as a negative control (-ve) containing only media.

Binding and elution conditions were maximised and it was found that diluting the cell supernatant 1:1 with binding buffer prior to loading improved binding. The optimal binding buffer was determined as 20 mM sodium phosphate pH 7.0. The pH of the elution buffer was also optimised and found to give sharper peaks at pH 2.5.

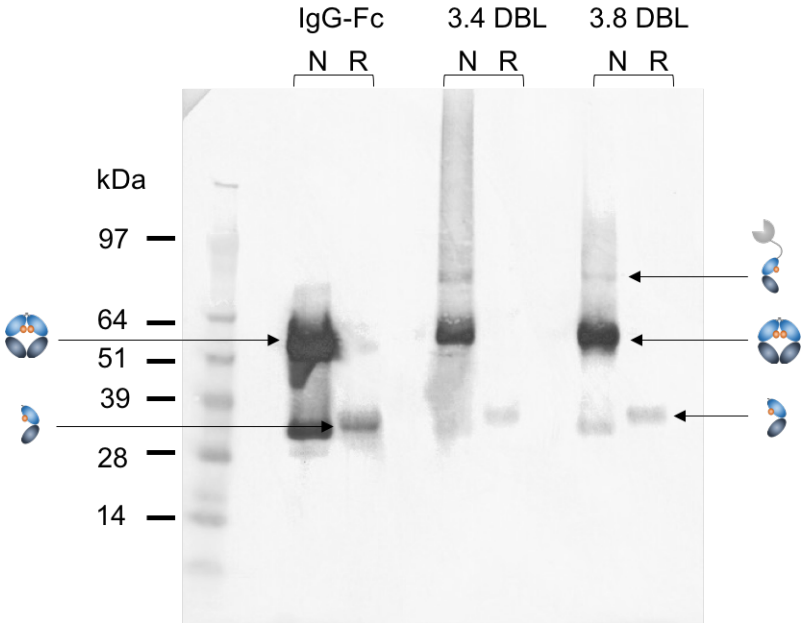
3.4.4 Characterisation of recombinant protein

In order to characterise the DBL-Fc fusion proteins, reduced and non-reduced samples were separated by SDS-page electrophoresis and stained with Coomassie Blue (Section 2.4.6) or transferred to a membrane and IgG Fc detected using αhIgG-Fc-AP (Sigma) (Figure 3.18). The proteins had difficulty folding with the addition of the Fc tag and were unable to form dimers. However, an $\sim$60 kDa band visible in the Western blot suggests that there is a small amount of monomeric DBL-Fc present.

Due to the disruption to oligomerisation by SDS, in order to investigate native structure of the fusion proteins the recombinant protein was run by gel filtration on a Superdex 200 10/300 GL column (Section 3.3.12). Each fraction collected from the gel filtration was immunoblotted to test for the presence of IgG Fc (Figure 3.19). There was IgG Fc detected in all fractions between A8 and B5 from the gel filtration experiment (corresponding to 700 - 1kDa). This suggests that protein aggregates are being formed. Large peaks might explain the smears at the top of the protein gel and Western blot (Figure 3.18).

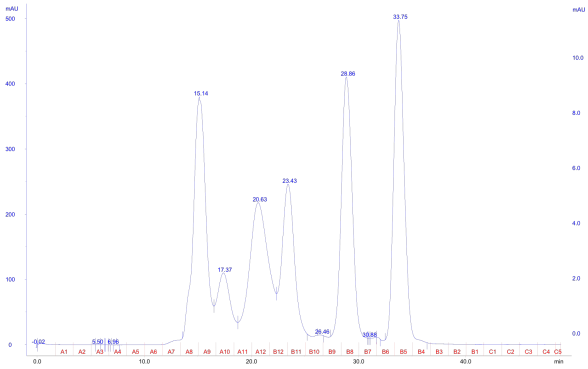


(a) Coomassie stain

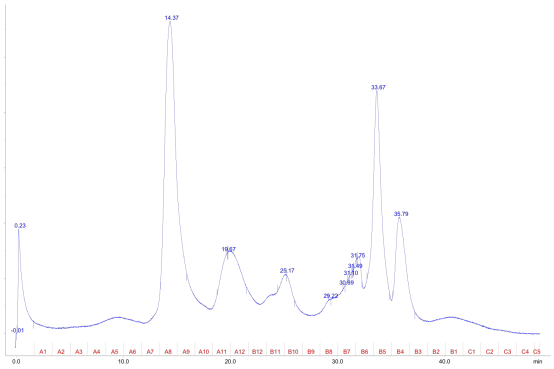


(b) Western blot

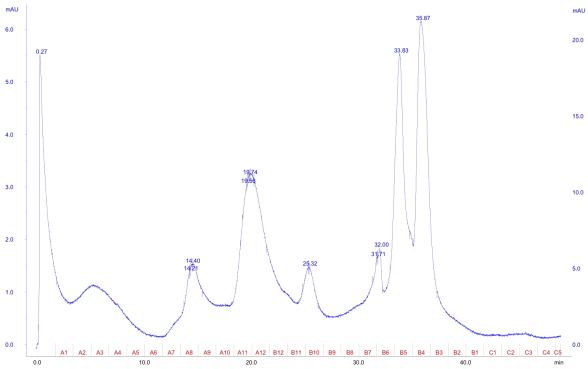
Figure 3.18: **DBL-Fc fusion proteins have difficulty folding.** SDS-PAGE of recombinant (a) IgG Fc control visualised with Coomassie blue stain; (b) the Fc detected using an αhIgG-Fc alkaline phosphotase conjugated antibody. Protein samples were either reduced (R) or non-reduced (N). The bands observed for the IgG Fc control agree with predictions (50 kDa non-reduced and 25 kDa reduced). Bands were predicted for non-reduced DBL-Fc fusion proteins at 120 kDa. The lack of these bands suggests that the proteins have difficulty forming homodimers. However, a band at ~60 kDa suggests the presence of a monomeric DBL-Fc fusion.



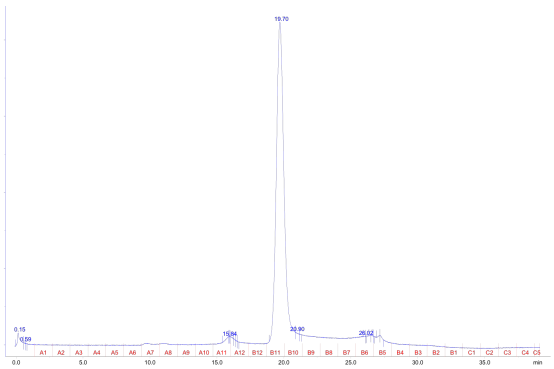
(a) Size exclusion standards



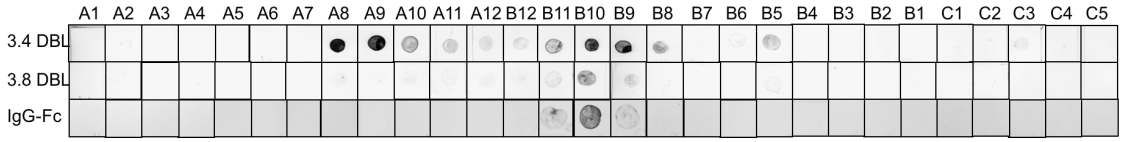
(b) 3.4 DBL size exclusion peaks



(c) 3.8 DBL size exclusion peaks



(d) IgG-Fc size exclusion peaks



(e) Immunoblot of size exclusion fractions

Figure 3.19: **Size exclusion profiles of the 3.4 DBL and 3.8 DBL recombinant proteins.** (a) shows the peaks for 500 µl of the protein standards, (b) and (c) show the 3.4 DBL and 3.8 DBL size exclusion profiles respectively. (d) Immunoblot of the 3.4 and 3.8 DBL size exclusion fractions shows fractions where IgG Fc is detected.

3.4.5 Functional analysis of DBL domains

It has recently been shown that the DBL domains from MSPDBL1 and MSPDBL2 bind hIgM (and not IgE, IgG or IgA) and that binding occurs irrespective of whether the IgM was bound to antigen or not (Crosnier et al., 2016). It was shown by SPR analysis that the binding interaction occurs at high affinity (Kd = 0.3 nM for the MSPDBL1 DBL domain and Kd = 1.1 nM for the MSPDBL2 DBL domain) and that no binding occurs via the SPAM domain. A series of IgM/A domain swap antibodies showed that binding occurs in the Cμ4 domains of human IgM (Crosnier et al., 2016). Interestingly, the Cμ4 domain is the same domain that is known to bind hFcμRs (Kubagawa et al., 2009). We aimed to show that the recombinant DBL-Fc protein produced binds hIgM and that the merozoite DBL proteins bind the Cμ4 domain of hIgM.

3.4.5.1 The recombinant DBL-Fc domains bind human IgM

IgM-binding of the DBL-Fc fusion proteins was investigated by sandwich ELISA (Section 3.3.10). IgM was captured by αhIgM, and the purified DBL proteins added and allowed to bind before αhIgG-Fc-AP was added as the detection antibody. Media was used as a negative control and a t-test performed which showed that the positive MSPDBL3.4 and MSPDBL3.8 binding results were statistically significant.

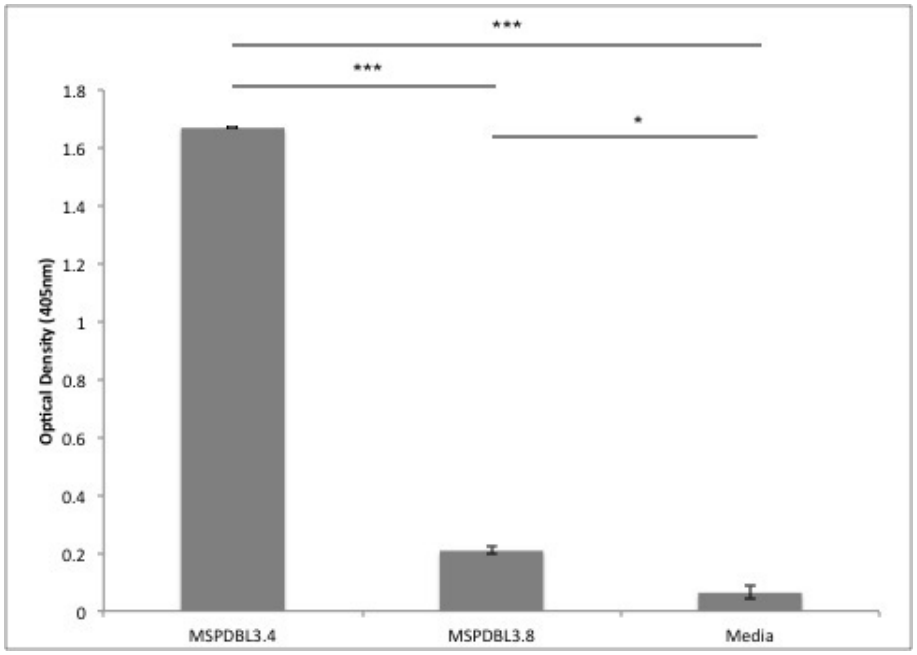


Figure 3.20: **Recombinant DBL domains bind human IgM.** (a) IgM-binding was detected by sandwich ELISA. IgM was captured by αhIgM coated down on the plate and DBL 3.4 and DBL 3.8 detected by αhIgG-Fc-AP. Results are shown from two repeat experiments. *$p<0.05$, ***$p<0.001$. IgM-binding of the IgG Fc was not detected in a separate experiment using the same sandwich ELISA protocol and a recombinant IgG Fc control (Appendix B).

3.4.5.2 Malarial MSPDBL1 and MSPDBL2 bind the Cμ4 domain of human IgM

The Cμ4 domain on IgM has been shown to bind FcμRs (Kubagawa et al., 2009) and malarial MSPDBL1 and MSPDBL2 have recently been shown to bind Cμ4 (Crosnier et al., 2016). We aimed to show that the merozoite DBL proteins bind the Cμ4 domain. As IgM and IgA are closely related secretory Abs containing tailpieces, in order to investigate the binding region we used IgG/M domain swap mutants.

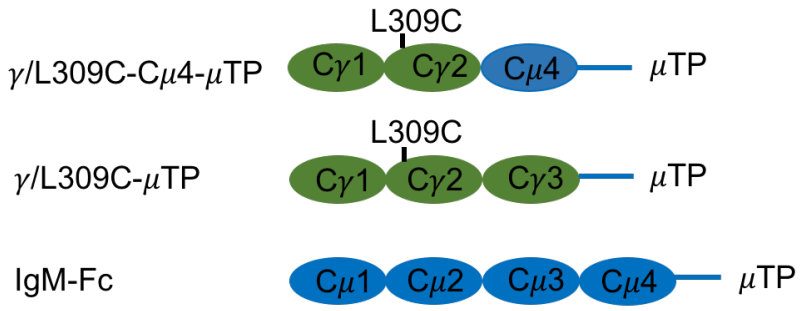
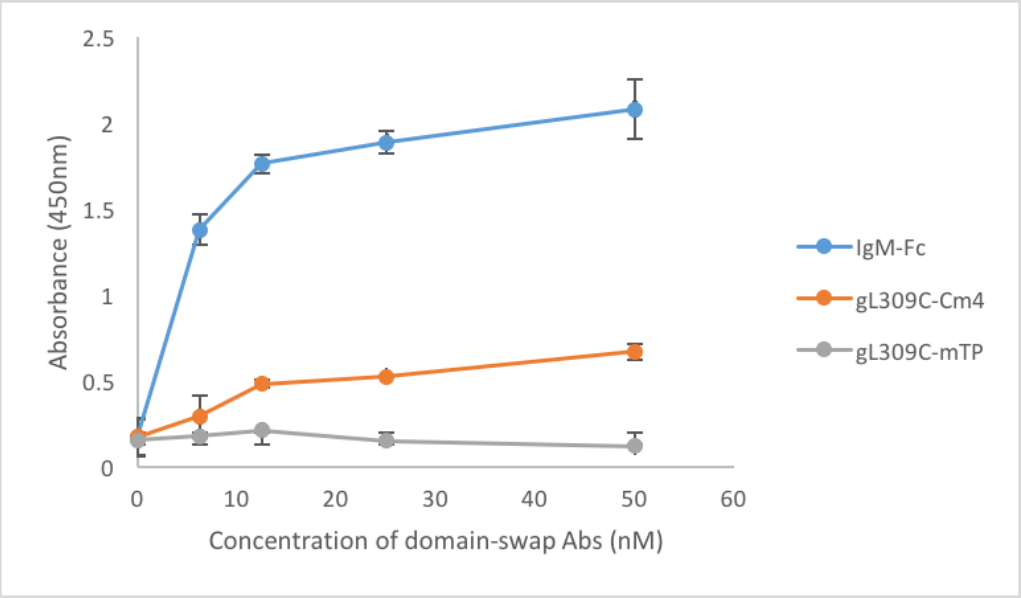
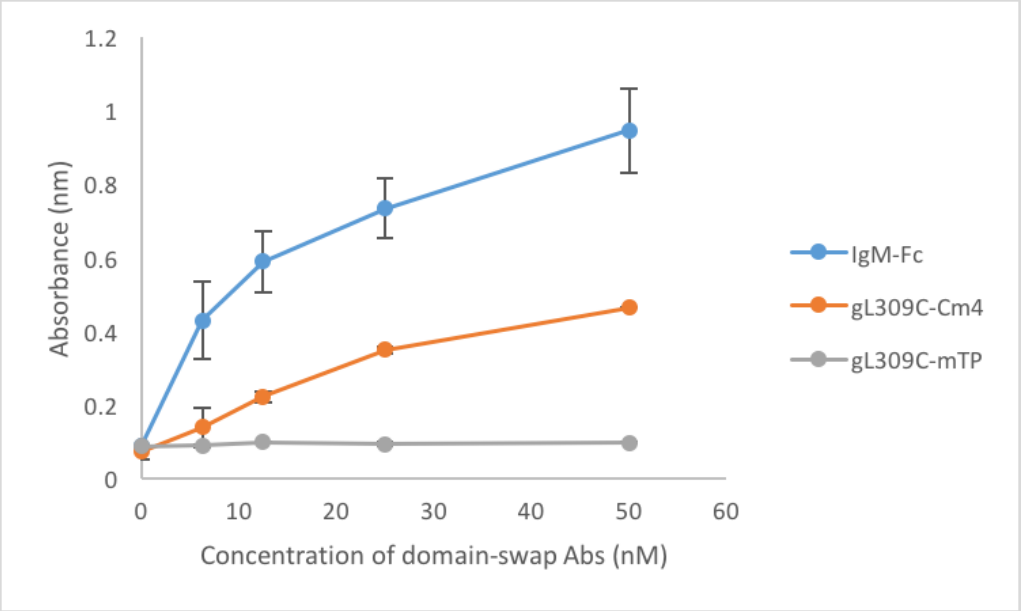


Figure 3.21: **IgG/IgM domain swap mutants.** Domain-swap mutants described by Ghumra et al. (2008) were used to investigate the role of the Cμ4 domain in MSP DBL-IgM binding. Two controls were used, IgM-Fc and L309C-μTP, which contains the Cγ1-3 regions with an IgM tailpiece. The L309C-Cμ4 mutant was used to investigate the role of the Cμ4 domain in IgM-binding. L309C is a single Leucine/Cysteine mutation in the IgG domain (homologous to the cysteine in the corresponding region of IgM); this mutant has been shown to have a greater ability to polymerise (Sørensen et al., 1999).

A panel of previously generated domain-swap Abs (Ghumra et al., 2008) was used to characterise the interaction of the full-length MSPDBL1 and MSPDBL2 with hIgM by ELISA (Figure 3.21). These experiments were performed by Professor Pleass (Liverpool School of Tropical Medicine) using full-length recombinant biotinylated MSPDBL1 and MSPDBL2 (kindly provided by Cecile Crosnier, Sanger Institute). The experiment showed that the Cμ4 domain was critical for binding of MSPDBL1 and MSPDBL2 to domain-swap antibodies. No binding was detected in the L309C-μTP mutant which lacked the Cμ4 domain. Binding was, however, detected in the γL309C Cμ4 mutant where the Cγ3 domain was swapped with Cμ4 from IgM. These results showed the ability of recombinant MSPDBL1 and MSPDBL2 to bind the Cμ4 domains of hIgM.



(a) MSPDBL1



(b) MSPDBL2

Figure 3.22: **MSPDBL1 and MSPDBL2 bind to the Cμ4 domain of human IgM.** An indirect ELISA was used to show binding of recombinant full-length biotinylated MSPDBL1 and MSPDBL2 to hIgM. Binding of the complete Fc section of hIgM (IgG-Fc) was compared to two domain-swap Ab mutants. Binding was shown in γL309C Cμ4 which contains the Cμ4 domain and was not present in the same mutant where the Cμ4 domain was replaced by a Cγ3 domain. The mean ± SD are shown for two experiments.

3.5 Summary

MSPDBL1 and MSPDBL2 are two new malaria vaccine candidates which have been identified recently with interesting binding properties. The only two merozoite surface proteins to contain a DBL domain, thought to be a virulence factor in *P. falciparum* infection, they have recently been shown to bind the Cμ4 domain of human IgG through this domain (Crosnier et al., 2016). In this chapter DBL domains from MSPDBL1 and MSPDBL2 were expressed as Fc-fusion proteins in a mammalian expression system. There was some difficulty expressing the Fc-fusion as a homodimer as the recombinant protein had some difficulty folding correctly. Monomeric DBL-Fc produced was shown to be fully functional although the yield was low. There was evidence of IgM binding and it was shown that the MSPDBL proteins bind the Cμ4 domain of human IgM.

Previously, DBL domains have been expressed more frequently in *E.coli* due to high protein yield and low costs, however recombinant proteins are either expressed as truncated forms or precipitate in insoluble inclusion bodies in the bacterial cells (Flick et al., 2004). Methods to obtain correctly folded proteins have been developed but are complex and laborious (Crosnier et al., 2013). We set out to create novel Fc-fusion constructs to express the DBL domains in a mammalian expression system. An Fc-fusion protein has the benefit of extending half-life of the recombinant protein as well as adding a convenient tag for protein purification, thus streamlining production. TOPO-TA cloning was used to create a parking plasmid containing the DNA and suitable restriction sites for sub-cloning. DNAs were sub-cloned into the pFUSE-hIgG1-Fc plasmid to create a DBL-Fc fusion construct. The Fc-fusion constructs were expressed in CHO-KI cells and purified using the Fc-tag on a protein G sepharose column.

The cloning was challenging, particularly cloning the 3.8 DBL as there were limited transformants and amplification of plasmids was difficult due to slow growth of bacterial cultures. Various *E. coli* strains were tried which exhibit a range of transformation efficiency as well as strains which stabilise the insert but growth was not improved. This suggests that the DNA may be toxic to *E.coli* cells.

Protein expression of both constructs was also particularly challenging. Although purification methods were optimised extensively, yields were still low. The constructs were efficiently transfected into CHO-K1 cells (compared to HeK cells in which transfection proved difficult), however, the cells produced low protein yield. Monoclonal populations

were produced from cells selected for high protein expression. This produced better results than the polyclonal population, however yields were still low suggesting that the cells have difficulty producing the protein. Protein purification was optimised and efficient purification techniques were developed to extract all protein produced by the cells.

Characterisation of the protein produced suggests that the DBL domain has difficulty folding as an Fc-fusion protein. Typical Fc-fusion proteins consist of two effector molecules because each molecule attaches to one chain of the Fc dimer. The DBL-Fc fusion proteins produced contain a single DBL domain attached to a single chain of the Fc which is unable to form a homodimer, possibly because the DBL domain is too large and disulphide bonds in the hinge region are unable to form between two chains of the Fc. The monomeric DBL-Fc fusion was, however, shown to be functional and was therefore used for further analysis.

The recombinant protein produced was tested for binding to human IgM. Binding was observed by both DBL domains, using sandwich ELISA, and was significantly higher for the 3.4 DBL domain than the 3.8, which is in agreement with another recent study where affinity was reportedly 10-fold higher in the full MSPDBL1 compared to MSPDBL2 (Crosnier et al., 2016). In this study, the Cμ4 domain was identified as the binding site for the DBL-IgM interaction which is the same domain where the FcμR binds to the Ab. In this Chapter it was shown that the Cμ4 domain was necessary for binding to the MSPDBL1 and MSPDBL2 proteins. In an ELISA, binding was knocked out in a domain-swap Ab mutant where the Cμ4 domain had been replaced with a Cγ3 domain.

The main obstacle in the work presented here was the generation of adequate quantities of correctly folded recombinant protein. Although the DBL-Fc fusion had difficulty folding into a homodimer, some functional monomeric protein was produced and IgM-binding was observed using this recombinant protein. The protein produced was functional with respect to both IgM and protein G binding, however, the cells have great difficulty in producing an adequate amount of protein for further functional investigations. This problem of folding was addressed in Chapter 5 while the protein produced here has been used for further investigations in Chapter 4, due to the IgM-binding observed in the monomeric DBL-Fc fusion.

Chapter 4

Analysis of Known IgM Binding DBL domains

4.1 Background

In Chapter 3 it was shown that the DBL domains from MSPDBL1 and -2 bind human IgM. The Cμ4 domain (Figure 4.1) appears to be important for the interaction, in agreement with a recent study by Crosnier et al. (2016). The Cμ4 domain has also been identified as part of the binding region for FcμRs on the surface of B and T lymphocytes as well as NK cells (Kubagawa et al., 2009).

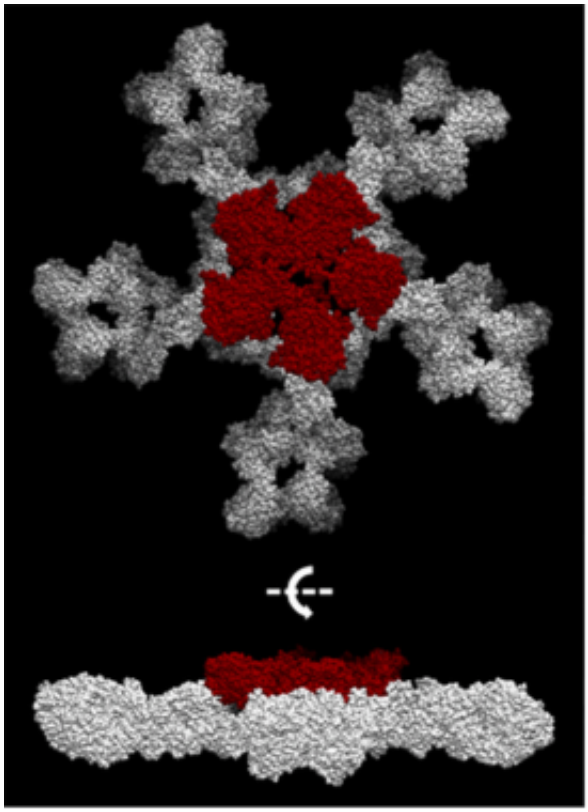


Figure 4.1: **Model of IgM**. The Cμ4 domain is important in binding to DBL domains and is shown on the model of IgM in red (adapted from Pleass et al. (2015)).

MSPDBL1 and MSPDBL2 are the only merozoite surface proteins to contain a DBL domain. This domain is also a feature of PfEMP1 proteins located on the erythrocyte surface, encoded by the *var* multigene family, and is thought to be responsible for their cytoadherent and subsequently virulent nature. These multi-domain proteins contain up to five DBL domains along with cysteine-rich inter-domain regions (CIDR) and are responsible for mediating adhesion of IEs to a multitude of host receptors. The DBL domains belonging to PfEMP1 are classified into six classes based on amino acid sequence (DBL α, β, γ, δ, ξ, ε) and within the DBL domains ten semi-conserved regions, or homology

blocks, have been categorised (Rask et al., 2010).

DBL classes have been associated with specific pathogenesis. For example, the DBLβ class has affinity for ICAM-1 and DBLδ for platelet-endothelial cell adhesion molecule 1 (PECAM-1) (Springer et al., 2004; Berger et al., 2013). DBLα is associated with heparin sulphate (HS) binding, complement receptor 1 (CR1) and blood group A antigen (Carlson et al., 1990; Mayor et al., 2005; Barragan et al., 2000b). CR1 binding is associated with coating of IE with uninfected erytrocytes, also known as rosetting (Rask et al., 2010). However, sequence diversity within each DBL class is high. It is therefore difficult to identify functional regions which mediate binding based on sequence analysis alone (Higgins and Carrington, 2014; Lau et al., 2015).

PfEMP1 and MSPDBL1 show structurally similar domain architecture, as do DBL domains from the erythrocyte binding-like (EBL) family of proteins such as EBA-175 which binds glycophorin A, the major glycoprotein found on human erythrocytes, during erythrocyte invasion. Interestingly, although DBL structures are relatively conserved, there is great variation in the host receptors with which they interact (Hodder et al., 2012). DBL domains consist of three α-helical, cysteine-rich subdomains. A study by Hodder et al. (2012) which compared the known structure of eight DBL domains found in *Plasmodium spp.* showed that the size and location of major helices are highly conserved. However, the extent of the connecting loops varies within each structure and the disulphide bond architecture, which stabilises each structure, appears to be unique to each protein family.

MSPDBL1 and MSPDBL2 contain an interesting leucine zipper-like region at the C terminus which is part of the SPAM domain, a domain present in most other members of the MSP3 family (Singh et al., 2009). This region has been shown to be responsible for oligomerisation of MSP3. It has been suggested that its presence enables MSP3 to form a highly extended and oligomeric structure on the merozoite surface allowing it to interact with erythrocytes at a relatively long distance (Burgess et al., 2005; Hodder et al., 2012). Since MSPDBL1 and MSPDBL2 also contain this region, it is likely that they are also long extended molecules with far-reaching interactions (Hodder et al., 2012).

4.1.1 IgM-binding DBL domains

The *P. falciparum* genome contains ∼60 *var* genes which encode for PfEMP1. Several PfEMP1 DBL domains are known to bind the Fc region of human IgM (Table 4.1).The function of DBL-Fcμ binding is not fully understood but the expression of these Fcμ binding proteins is linked to severe malaria. This phenotype appears to be driven by parasite evasion of human immune effector mechanisms (Pleass et al., 2015) (Figure 4.2).

Table 4.1: Known hIgM-binding and non-binding *Plasmodium falciparum* proteins. Adapted from Pleass et al. (2015).

Binds to hIgM	Protein	DBL domain	Reference
Yes	FCR3var1csa	DBLε7	(Semblat et al., 2006)
	FCR3var2csa	DBLε6	(Semblat et al., 2006)
	HB3var06	DBLξ2	(Stevenson et al., 2015a)
	MAL6P1.4	DBLε2	(Jeppesen et al., 2015)
	MAL5P1.4	DBLε3	(Jeppesen et al., 2015)
	PFL0020w	DBLε4	(Jeppesen et al., 2015)
	PFL0030c	DBLεPAM5	(Jeppesen et al., 2015)
	TM274var1	DBLξ4	(Ghumra et al., 2008; Semblat et al., 2015)
No	FCR3var1csa	DBLε4	(Semblat et al., 2006)
	FCR3var1csa	DBLξ6	(Semblat et al., 2006)
	FCR3var2csa	DBLε4	(Semblat et al., 2006)
	FCR3var2csa	DBLε5	(Semblat et al., 2006)
	HB3var06	DBLε13	(Stevenson et al., 2015a)
	HB3var06	DBLε14	(Stevenson et al., 2015a)
	ITvar9/R29var1	DBLε3	(Stevenson et al., 2015a)
	MAL6P1.4	DBLε7	(Jeppesen et al., 2015)
	PF07_0139	DBLε4	(Jeppesen et al., 2015)
	PFL0020w	DBLξ5	(Jeppesen et al., 2015)
	PFL0030c	DBLεPAM4	(Jeppesen et al., 2015)
	PFL0030c	DBLξPAM10	(Jeppesen et al., 2015)
	TM284var1	DBLε3	(Ghumra et al., 2008)
	TM284var1	DBLε5	(Ghumra et al., 2008)
	Var0	DBLξ	(Vigan-Womas et al., 2012)

CSA-binding IEs which are involved in the pathogenesis of placental malaria, are known to bind IgM, mediated by a type of PfEMP1 called Var2csa (Semblat et al., 2006). This interaction appears to be protective as it prevents phagocytic clearance of IgG-opsonised IEs without compromising the CSA-adhesive function of the antigen (Jeppesen et al., 2015; Pleass et al., 2015).

Another important type of PfEMP1 known to bind IgM is one which also mediates the

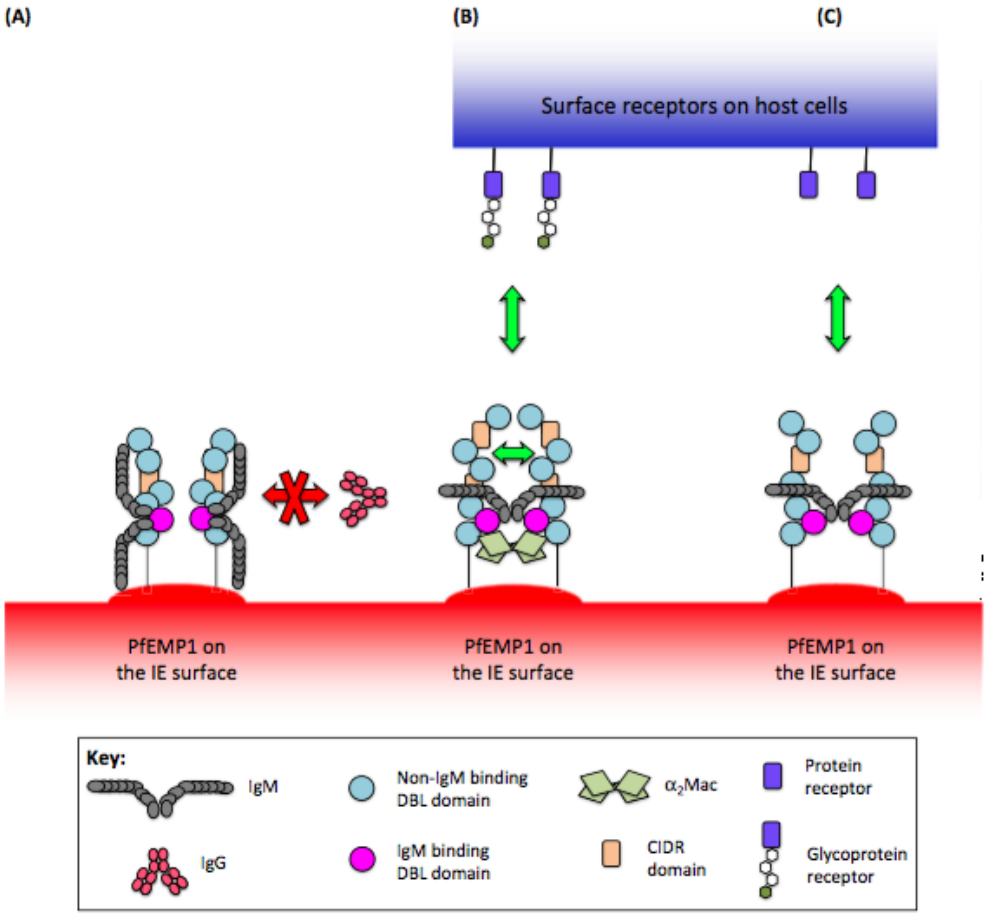


Figure 4.2: **Binding of IgM to *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) (Pleass et al., 2015).** (A) In CSA-binding isolates, IgM protects the antigen from destruction by immune IgG. (B) In the case of rosetting isolates, IgM binding tethers PfEMP1 for stronger interactions with host surface receptor carbohydrates. (C) IgM also binds to PfEMP1 on infected erythrocytes that neither bind CSA or rosette (Pleass et al., 2015).

formation of rosettes, which is the masking of an IE by several uninfected erythrocytes. It has been suggested that IgM augments low-affinity interactions with the erythrocyte carbohydrate receptor involved in rosetting (Pleass et al., 2015). Pentameric IgM is able to cross-link PfEMP1 proteins as each molecule can bind two PfEMP1s (Jeppesen et al., 2015). Interestingly, another abundant serum protein, α_2-macroglobulin (α_2M) appears to serve a similar function, binding the same DBL domain, and has a higher potential for cross-linking individual molecules of PfEMP1 (Stevenson et al., 2015b; Pleass et al., 2015).

Other PfEMP1 variants which bind IgM, other than those types which bind CSA or mediate rosetting, have been identified. The addition of these variants may act to expand

the repertoire of sequestration-supporting host receptors, thereby reinforcing avoidance of clearance by the spleen (Pleass et al., 2015).

In PfEMP1, the IgM binding domains seem to be situated near the infected erythrocyte membrane, whereas the single DBL of the MSP suggests that binding is more exposed on the merozoite surface and possibly allows bridging to host receptors.

As yet, no obvious binding motifs or critical amino acids have been identified, therefore production of a library of MSP DBL mutants aims to provide insight into the IgM binding site of these domains. A comprehensive structural analysis of the mutants along with other known IgM binding and non-binding DBL domains will shed light on the structural features important for IgM-binding.

4.2 Objectives

The aim of this Chapter is to analyse known IgM-binding DBL domains and to identify shared structural properties as well as sequence motifs. The specific objectives include:

- Produce a library of recombinant merozoite DBL mutants to investigate residues critical for IgM-binding.
- Predict potential binding sites based on the crystal structure and model of IgM binding.
- Analyse sequences of known IgM binders and non-binders and identify common features for binding.
- Analyse structures of known IgM binders and non-binders and identify common features for binding.

4.3 Methods

4.3.1 PCR for random mutagenesis

The Genemorph II Random Mutagenesis kit (Agilent Technologies) uses Mutazyme II DNA polymerase. This error-prone polymerase has less mutational bias than other Taq DNA polymerase mutants. The amount of initial target DNA was calculated in order to produce a low error frequency of 1 - 2 amino acids. Low frequency errors occur by increasing the initial template amount, as the more times the target is replicated, the more errors it accumulates. Targets amplified from high amounts of target DNA undergo fewer duplications than targets amplified from low concentrations. 600 ng of plasmid DNA ($\sim$ 4.7 kbp) gives 110 ng of target DNA ($\sim$ 860 bp) which, according to manufacturer's instructions, should produce approximately nine mutations/kb and eight mutations within the DBL target. This corresponds to mutations of approximately four amino acids and is a common mutational frequency used in directed evolution studies (Cherry et al., 1999; You and Arnold, 1996).

The PCR solution used for low frequency random mutagenesis was composed of the following:

- Mutazyme II polymerase 1 µl
- Mutazyme II reaction buffer 5 µl
- dNTPs (40 mM) 1 µl
- Forward primer (10 µM) 3 µl
- Reverse primer (10 µM) 3 µl
- pFuse-DBL plasmid 600 ng
- dH_2O up to 50 µl

The same primers that were used to amplify the DNA in Chapter 3 were used for the mutagenesis reaction (Table 2.1) except that from the forward primer for the 3.8 DBL reaction which was redesigned adding extra bases in front of the EcoRI site (red) for optimal cleavage.

3.8-F New: 5' tcgcggcg**gaattc**gatctgcaaggac 3'

PCR reaction

A Progene thermocycler (Techne, Cambridge, UK) was used for the PCR reaction. The following PCR program was used:

94° C 3 mins

35 cycles
- 94° C 1 min
- 60° C 1 min
- 72° C 1 min

72° C 10 mins.

Purification from PCR product

PCR products were purified from solution using High Pure PCR cleanup micro kit (Section 2.2.3) and DNA concentration quantified by nanodrop spectro-photometer (Section 2.2.15).

4.3.2 DNA preparation for sub-cloning of mutated pFuse constructs

Restriction digest of mutated DBL inserts

The amplified mutated DBL insert was cut out by restriction digest. The restriction digest solution consisted of the following:

- Restriction enzyme 1 — 1 µl
- Restriction enzyme 2 — 1 µl
- Compatible Buffer — 5 µl
- BSA (10x) — 5 µl
- Mutated DNA — 1 µg
- dH_2O — up to 50 µl

Whenever less than 1 µg of DNA was available, a volume of DNA up to 30 µl was added and the dH_2O quantity adjusted accordingly. For the 3.4 DBL construct, restriction enzymes Nco1 and BglII were used with NEB buffer 3.1. For the 3.8 DBL construct

restriction enzymes Nco1 and EcoRI were used with NEB buffer 3.1. The restriction digest reaction was incubated at 37°C for 2 hours and then the product run by gel electrophoresis where bands correlating to the correct insert size were excised and purified (Section 2.2.6). Recovered DNA was quantified by nanodrop spectro-photometer (Section 2.2.15).

4.3.3 Restriction digest of pFUSE plasmid

The pFUSE-hIgG1-Fc2 plasmid was prepared for the insertion of the mutated DBL inserts in the same way as in Chapter 3 for sub-cloning of the pFUSE constructs. Plasmid was digested using the appropriate restriction enzymes, de-phosphorylated to prevent self-ligation, run by gel electrophoresis, excised and purified (Section 3.3.5). Recovered DNA was quantified by nanodrop spectro-photometer (Section 2.2.15).

4.3.4 pFUSE ligation to the mutated 3.4 DBL and 3.8 DBL domains

The following reaction with a molar ratio of 1:3 vector to insert was set-up on ice:

- 10x T4 DNA ligase buffer 2 µl
- Vector DNA (4.2 kb) 50 ng
- Insert DNA (860 bp) 30 ng
- T4 DNA ligase 1 µl
- dH_2O up to 20 µl.

The ligation reaction was held at 16°C for 16 hours using a Progene thermocycler.

4.3.5 Transformation of competent cells with mutated pFUSE-hIgG1-Fc2-3.4-DBL and pFUSE-hIgG1-Fc2-3.8-DBL

Cells were transformed using the same method as in Chapter 3 for the transformation of the original pFUSE constructs. Single colonies were grown and DNA purified by mini-prep which was then screened to check that it contained the insert by restriction digest (Section 3.3.7).

DNA sequencing of mutant DBL-Fc fusion constructs

Plasmids containing insert were identified and DNA sent for sequencing. Several plasmids had mutations which introduced stop codons and many mutations were silent. Plasmid sequences were checked and mutants were identified for expression.

4.3.6 Expression and purification of mutant 3.4 and 3.8 DBL-Fc fusion recombinant protein library

Recombinant mutant proteins were expressed in CHO-KI cells using the same methods as were used in Chapter 3 (Section 3.3.8), however, HAMs F10 media was used instead of DMEM. CHO-KI cells were transfected using FuGENE 6 transfection reagent and selected using zeocin selection media. Monoclonal high-expressing colonies were selected by immunoblot analysis and subsequently cell supernatant was purified using a protein-G sepharose HiTrap column on the ÄKTA-FPLC. Supernatants were diluted in a new binding buffer in order to optimise binding conditions. Methods were used as in Chapter 3 but with 20 mM sodium phosphate pH 7.0 as binding buffer. Protein was eluted using 0.1 M glycine-HCl, pH 2.7. Immunoblot analysis was used to test for the presence of recombinant protein in FPLC fractions which were then pooled and concentrated by ultrafiltration.

4.3.7 IgM-binding ELISA

The library of mutants was screened for IgM-binding by ELISA. Methods were used as in Chapter 3 (Section 3.3.10).

4.3.8 Western blot analysis

After proteins were run on a NuPAGE Novex 4-12% Bis-Tris pre-cast gel by SDS-PAGE (Section 2.4.6) they were transferred to nitrocellulose for Western blotting (Section 2.4.8).

Detection of MSP 3.4 and 3.8

To detect the presence of the MSP DBL domains αMSPDBL1 and αMSPDBL2, polyclonal antibodies (kindly provided by Cecile Crosnier, Sanger Institute) were used. Polyclonal antibodies were diluted to 0.5 μg/ml in PBST and the nitrocellulose blot incubated overnight at 37°C on a rocker. The membrane was washed three times before the addition of an αrabbit-IgG-AP (Sigma A9919) at a dilution of 1:5,000 in PBST which was incubated

for 2 hours at 37°C. The membrane was again washed three times before proteins were visualised by BCIP/NBT (Sigma) according to the manufacturer's instructions.

Detection of IgG-Fc

To detect the presence of IgG-Fc an αhIgG-Fc-AP was used as in Chapter 3 (Section 3.3.11).

IgM binding

To investigate IgM binding in DBL mutants, 5 μg of each DBL mutant was incubated with 5 μg of hIgM for 4 hours at 4°C on a shaker. 50 μl of αhIgM agarose beads were washed three times and resuspended in 200 μl PBS. The bead suspension was added to the DBL-IgM complex and incubated overnight at 4°C on a shaker in order to pull out any DBL-IgM in complex. The beads were washed three times with PBS, SDS loading buffer added and the samples held at 95°C for 5 mins. Samples were run on SDS-Page before transferring to nitrocellulose (Sections 2.4.6 and 2.4.8). After the membrane was blocked and washed, αhIgG-Fc-AP was used to visualise proteins as in Chapter 3 (3.3.11).

4.3.9 Gel filtration of recombinant mutant DBL domains

The library of mutants was characterised by gel filtration using the same methods as in Chapter 3 (Section 3.3.12).

4.3.10 Sequence analysis

Sequence alignments

Multiple sequence alignments were produced using the EMBL-EBI Clustal Omega multiple sequence alignment web form available at `http://www.ebi.ac.uk/Tools/msa/clustalo/` (Sievers et al., 2011).

Local nucleotide and amino acid sequence similarities were analysed using the NCBI BLAST: Basic Local Alignment Search Tool available at `http://blast.ncbi.nlm.nih.gov/Blast.cgi`. BLAST was also used for sequence alignment.

Homology-extended multiple sequence alignments were producing using the multiple sequence alignment toolbox, PRALINE available on the IBIVU web server at `http://www.ibi.vu.nl/programs/pralinewww/` (Simossis and Heringa, 2005).

Homology block analysis

Frequent homology block patterns within known IgM-binding and non-binding DBL domains were data mined using the release version of SPMF Open Source Data Mining Library and the Apriori Algorithm which is available at `http://www.philippe-fournier-viger.com/spmf/` (Fournier-Viger et al., 2014).

Homology blocks were analysed using the VarDom 1.0 Server `http://www.cbs.dtu.dk/services/VarDom/|` (Rask et al., 2010).

Phylogenetic analysis

Maximum likelihood phylograms were constructed using MEGA 6.0 available from `http://www.megasoftware.net` (Tamura et al., 2013).

4.3.11 Structural analysis

DBL-IgM binding simulations

DBL-IgM binding simulations were performed in collaboration with Dr Daniel Czajkowsky (Shanghai Jiao Tong University). The DBL homology model of MSPDBL1 was developed based on the known structure of the MSPDBL2 DBL domain. Molecular dynamic simulations were based on the previously modelled DBL-IgM interaction (Czajkowsky and Shao, 2009; Czajkowsky et al., 2010).

Molecular graphics and analyses were performed with PyMOL (Molecular Graphics System, v1.7.4.5 Edu Enhanced for Mac OS X, Schrodinger, LLC) and the UCSF Chimera package (Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco. Supported by NIGMS P41-GM103311) (Pettersen et al., 2004).

3D structure prediction

3D structure prediction was performed using the Phyre 2 protein fold recognition server available at `http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index` (Kelley et al., 2015) and the SWISS-MODEL protein structure homology-modelling server available at `http://swissmodel.expasy.org` (Biasini et al., 2014).

Stability analysis

Stability of point mutations was analysed using the SDM server at `http://mordred.bioc.cam.ac.uk/~sdm/sdm.php` which analyses the local structure of the wild-type and mutant and calculates a stability score for the point mutation (analogous to free energy difference) (Worth et al., 2011). Multiple mutations in a single mutant were analysed separately as point mutations.

4.3.12 Protein-protein interaction site prediction

Protein-protein interaction site prediction was performed using the meta-PPISP web server available at `http://pipe.scs.fsu.edu/meta-ppisp.html` (Qin and Zhou, 2007).

4.4 Results

4.4.1 Generation of Library of MSPDBL mutants

4.4.1.1 Random mutagenesis

A Genemorph II Random Mutagenesis kit (Agilent Technologies) was used for random mutagenesis, using the Mutazyme II error-prone DNA polymerase. The PCR product was run on a 1% gel and bands corresponding to the DBL inserts were identified (Figure 4.3).

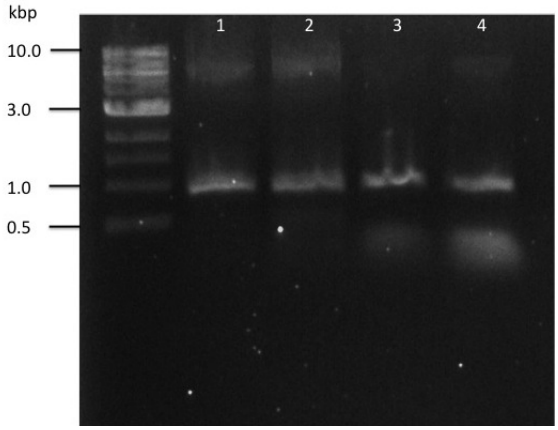


Figure 4.3: PCR amplified products of the MSP3.4 and MSP3.8 DBL domains on 1% agarose gel. **Lanes: 1** MSP3.4 DBL amplified using Phusion High Fidelity DNA Polymerase, **2** MSP3.4 DBL amplified using Mutazyme II polymerase (Genemorph II kit), **3** MSP3.8 DBL amplified using Phusion High Fidelity DNA Polymerase, **4** MSP3.8 DBL amplified using Mutazyme II polymerase (Genemorph II Kit)

The PCR products were then purified (Section 2.2.3) and cut by restriction digest alongside the pFUSE-hIgG1-Fc2-DBL3.4 and pFUSE-hIgG1-Fc2-DBL3.8 plasmids (Section 4.3.2).

Digested PCR product

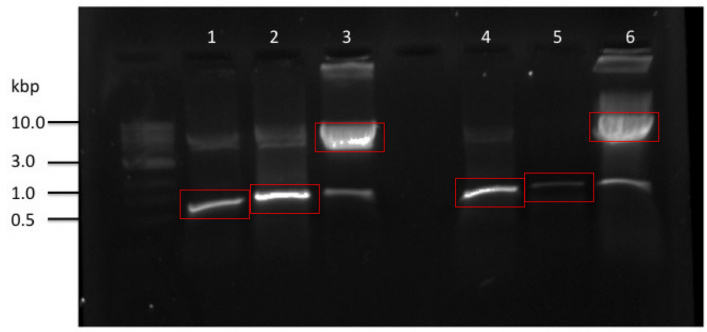


Figure 4.4: Restriction digest of 3.4 DBL and 3.8 DBL domains amplified by both Phusion and Mutazyme II DNA polymerase as well as pFUSE digested plasmid on 1% agarose gel. Bands highlighted in red were cut out and purified (Section 2.2.6). **Lanes: 1** 3.4 DBL amplified with Phusion and cut with restriction sites NcoI and BglII, **2** 3.4 DBL amplified with Mutazyme II and cut with restriction sites NcoI and BglII, **3** pFUSE-hIgG1-Fc2-DBL3.4 plasmid digested with NcoI and BglII, **4** 3.8 DBL amplified with Phusion and cut with restriction sites EcoRI and NcoI, **5** 3.8 DBL amplified with Mutazyme II and cut with restriction sites EcoRI and NcoI, **6** pFUSE-hIgG1-Fc2-DBL3.8 plasmid digested with EcoRI and NcoI.

Purified PCR products were ligated into the relevant pFUSE plasmid (Section 2.2.8) and 5 µl of the ligation reaction transformed into TOP10 cells (Section 2.2.10). Colonies were grown overnight in ampicillin selection media (Section 2.1.16) before using mini-prep to isolate the DNA (Section 2.2.11) from each colony.

Screening of positive colonies

Each colony was digested using NcoI/BglII for DBL 3.4 and EcoRI/NcoI for DBL 3.8 (Section 2.2.12) in order to identify colonies which contained the insert (Figure 4.5).

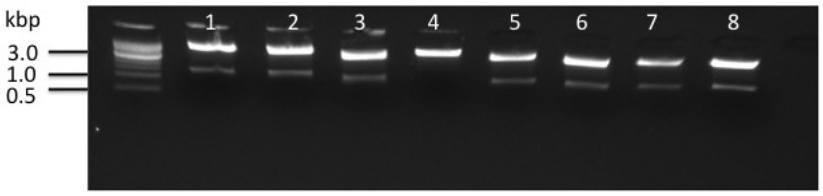


Figure 4.5: Typical restriction digest of colonies grown from transformation. Colonies are positive if they contain the DBL insert. **Lanes: 1-3, 5-8** Positive colonies, **4** Negative colony.

Mini-prep DNA from colonies identified as positive was checked by sequencing (Source

Bioscience, Section 2.2.13). From mutagenesis of the 3.4 DBL construct, DNA was extracted from 116 colonies and 103 were identified as positive by restriction digest screening. From mutagenesis of the 3.8 DBL construct, DNA was extracted from 80 colonies. Of these, 41 were identified as positive by restriction digest screening.

A script was written in Python to help analyse the sequences. The 5' sequence was aligned with the reverse complement 3' sequence and any mis-matches identified and corrected using the ab1 file until both sequences were identical. The 5' sequence was then aligned with the amino acid sequence from the DBL domain and any mismatches confirmed as mutations.

Analysis of sequencing results

Sequencing results from all colonies were analysed (Table 4.2). Of the 103 3.4 DBL colonies, reads could not be obtained for five. Of the remaining 98, only 18 contained mutations which did not introduce a stop codon, or lose/gain a nucleotide thus introducing a frame shift. The mutagenesis reaction had a low success rate, as 71.4% of the colonies contained wild-type DBL inserts. Wild-type inserts occurred either because any mutations changed only the codon and not the corresponding amino acid, or no mutations occurred.

Of the 41 3.8 DBL colonies sent for sequencing, reads could not be obtained for two. Of the remaining 39, 30 contained mutations which did not introduce a stop codon, or introduce a frame shift. This mutagenesis reaction had a much higher success rate as only 10% of the colonies sequenced contained wild-type DBL inserts.

Table 4.2: **Analysis of mutant sequencing results.** Of the total colonies identified as positive, a proportion of the colonies sequenced were found to contain wild-type inserts (WT) with no mutations. A proportion were found to contain mutated inserts with potential for expression (Mut). Sometimes, a mutation introduced a stop codon (stop) or introduced a frame shift (FS).

	Total colonies	# stop	% stop	# FS	% FS	# WT	% WT	# Mut	% Mut
3.4 DBL	98	8	8.2	2	2	70	71.4	18	18.4
3.8 DBL	39	4	10	2	5	4	10	29	75

The observed difference in success rates of the mutagenesis reaction is interesting because a higher percentage of colonies selected from the 3.4 DBL mutagenesis reaction contained an insert when screened by restriction digest. However, upon further analysis of sequencing results, these inserts were more likely to be wild-type. It was difficult to produce colonies for the 3.8 DBL mutagenesis reaction, and colonies selected were less likely to contain an insert when screened by restriction digest. However, upon sequencing the insert was much more likely to contain a mutation.

4.4.2 Selection of mutants for expression

A molecular simulation of DBL-IgM binding developed by Czajkowsky et al. (2010) localises the IgM-binding region of both 3.4 DBL and 3.8 DBL to an exposed area of subdomain 2. This consists of the end of helix 5 (h5) with the loop between helices h5 and h6, and the end of helix 2 (h2) and the start of helix 3 (h3) as well as the loop joining h2 and h3 (Figure 4.6).

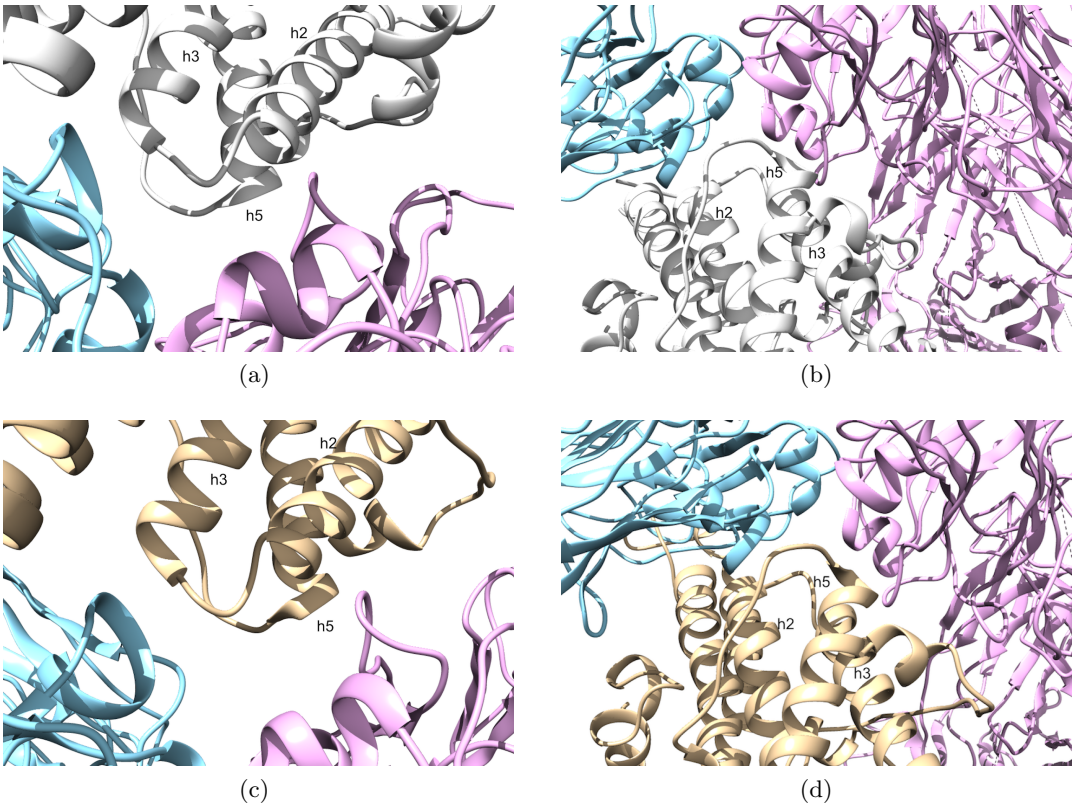


Figure 4.6: **Model of DBL-IgM binding.** (a & b) Model of 3.4 DBL (white) binding between two monomers of IgM (blue and pink). (c & d) Model of 3.8 DBL (brown) binding between two monomers of IgM. Both models highlight the end of helix 5 (h5) and the loop between helix 5 and 6 as well as helices 2 (h2) and 3 (h3) as regions involved in IgM-binding.

Ten mutants for each of the MSP DBL domains were selected for protein expression (3.4: Table 4.3 3.8: Table 4.4). These mutants were selected based on the two regions identified by the IgM-binding model. Where multiple mutants contained mutations in these areas, mutants were selected based on changes in charged residues. Those containing mutations involving cysteine residues were also selected due to their key role in stabilising the core structure. A complete list of all mutants produced along with the location of their mutations can be found in Appendix C.

Table 4.3: **Summary of 3.4 DBL mutants.** Mutants expressed and their corresponding amino acid mutations. Sequence location of each mutation is described along with its helical position and any changes in charge noted. Recombinant protein yield for each mutant has been summarised with ++ representing equivalent yield to wild-type while + represents a lower yield and +++ improved yield. IgM-binding is represented by ✓which is equivalent to wild-type binding and ✓✓represents improved IgM-binding compared to wild-type. Reduced IgM-binding was not observed for any mutants.

DBL mutant	Mutation	Change in charge	Position	Protein yield	IgM-bindng
3.4 #10	G169S		h1	++	✓
	K194N	Y	h1		
	N278D	Y	h4		
	C427R	Y	h9		
3.4 #26	K194T	Y	h1	++	✓
	Q265H	Y	h4		
	E279G	Y	h4		
	I285F		loop (h4-h5)		
	N388S		h7		
3.4 # 36	D323V	Y	loop (h5-h6)	+	✓
3.4 # 47	I266N		h4	+++	✓
3.4 # 48	S281G		loop (h4-h5)	++	✓
	A294V		h5		
	W300R	Y	h5		
3.4 # 56	I243V		h3	+++	✓
3.4 # 66	I200V		loop (h1-h2a)	++	✓
	H310R	Y	h5		
	E318V	Y	loop (h5-h6)		
3.4 # 74	D170E		h1	+	✓
	L228P		h2		
	N321Y		loop (h5-h6)		
3.4 # 77	A175T		loop (h1-h2a)	+	✓
	N269D	Y	h4		
	C340S		h6		
	K407M	Y	h7		
3.4 # 87	N201Y		h2a	++	✓✓

Table 4.4: **Summary of 3.8 DBL mutants.** Mutants expressed and their corresponding amino acid mutations. Sequence location of each mutation is described along with its helical position and any changes in charge noted. Recombinant protein yield for each mutant has been summarised with ++ representing equivalent yield to wild-type while + represents a lower yield and +++ improved yield. IgM-binding is represented by ✓which is equivalent to wild-type binding and ✓✓represents improved IgM-binding compared to wild-type. Reduced IgM-binding was not observed in any mutants.

DBL mutant	Mutation	Change in charge	Position	Protein yield	IgM-binding
3.8 #4	S210G		h1	+	✓
	S395G		h6		
3.8 #12	L203V		loop (h1a-h1)	+	✓
	R243L	Y	h5		
	T396A		h7		
	N424K	Y	h7		
3.8 #17	V202M		loop (h1a-h1)	+	✓
	P205S		loop (h1a-h1)		
	L211P		loop (h1-h2a)		
	K255R		h3		
	T335S		loop (h5-h6)		
3.8 #23	K200N	Y	loop (h1a-h1)	+	✓
	D229G	Y	h2		
	Y235H	Y	h2		
3.8 #34	G228D	Y	loop (h2a-h2)	+	✓
	K330E	Y	loop (h5-h6)		
3.8 #37	W309R	Y	h5	+	✓
	M321L		h5		
	D421V	Y	h7		
3.8 #38	N300K	Y	loop (h4-h5)	+	✓
	K304E	Y	loop (h4-h5)		
3.8 #44	N184S	Y	loop (h1a-h1)	++	✓✓
	N339D	Y	loop (h5-h6)		
3.8 #65	H257Q	Y	h3	++	✓✓
	T396A		h7		
3.8 #70	N199Y	Y	loop (h1a-h1)	+	✓
	Q376H	Y	loop (h6-h7)		

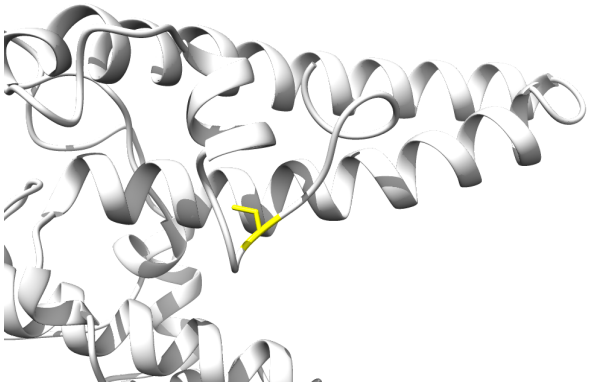
Stability analysis of mutations

The SDM server was used to predict the effect of a mutation on the DBL structure. This calculates the free energy difference between the wild-type and mutant which is represented in a stability score (Worth et al., 2011). The aim was to identify mutations which have an impact on protein structure and may therefore have an impact on function. Multiple highly destabilising mutations were predicted in various DBL mutants. 3.4 DBL mutants #48 and #77 contain such multiple highly destabilising mutations (S281G, W300R and C340S) and 3.8 DBL mutant #34 contains exclusively destabilising mutations, one of which is predicted to have a highly destabilising effect (G228D) (full analysis can be found in Appendix F). The combined destabilising effect of multiple mutations is unclear from this analysis as the free energy difference is calculated only from point mutations. Analysis was therefore not used in selecting mutants for expression, however, mutants identified from this analysis were investigated further if loss of functionality was experimentally verified.

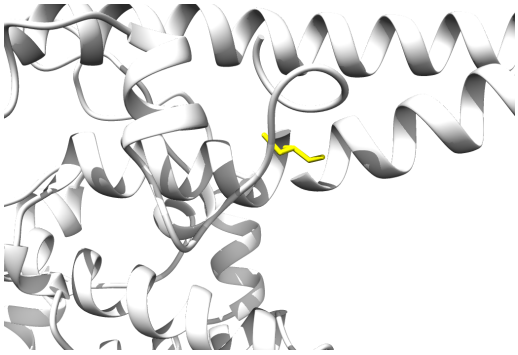
Cysteine mutants

While many cysteine residues are conserved, the overall disulphide bond structure of a DBL domain has been shown to be unique to this protein family (Hodder et al., 2012). Merozoite DBL domains have a distinct disulphide bond structure which links together subdomains 1 and 2, as well as stabilising the long alpha-helical subdomain 3. The location of cysteine residues is shown in the homology model of 3.4 DBL (Figure 4.7). Mutant 3.4 #10 contains the mutation C427R which corresponds to C12a, a cysteine residue located at the C-terminal and found only in the merozoite DBL domains and PfEMP1 Var2CSA DBL6ε. This is therefore an interesting mutant to express. 3.4 #77 contains the mutation C340S which corresponds to C7, a linking cysteine in subdomain 3. While present in both merozoite DBL domains, this cysteine is also a common feature of PfEMP1 DBL cysteine architecture. It is seen in some members of the EBA family and could therefore be critical for stability of the structure.

(a)



(b)



(c)

Figure 4.7: **Mutated cysteine residues in 3.4 DBL mutants.** (a) Homology model of 3.4 DBL showing the location of cysteine residues. (b) Homology model showing the location of C427R in 3.4 DBL mutant 10. (c) Homology model showing the location of C340S in 3.4 DBL mutant 77.

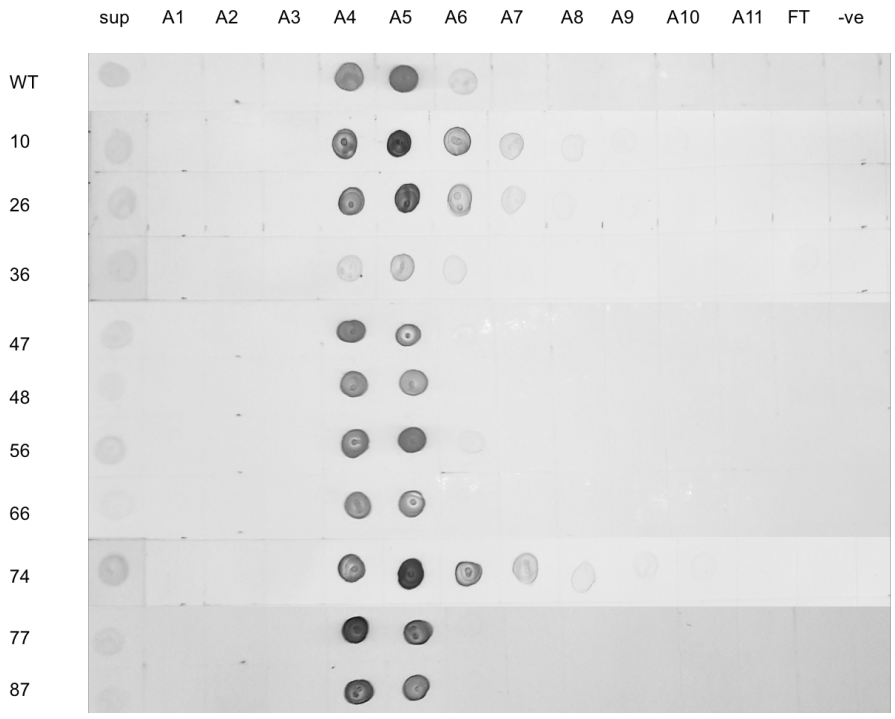
4.4.2.1 Expression of recombinant DBL mutants

DBL mutants were expressed using HAMs F10 media and the same methods as Chapter 3. Colonies grown from high-expressing single cells were expanded to large cultures and following 10 days of growth, the cell supernatant harvested and purified using a protein G column on an AKTA FPLC. Purification methods were developed to optimise binding conditions. HBSS was replaced with 20 mM sodium phosphate which was found to be the optimal binding buffer. Cell supernatant was diluted 1:1 with binding buffer prior to loading. This ensured optimal ionic strength and pH. Elution conditions were optimised and the pH for elution was increased to pH 2.7, this still enabled effective elution and prevented protein denaturing.

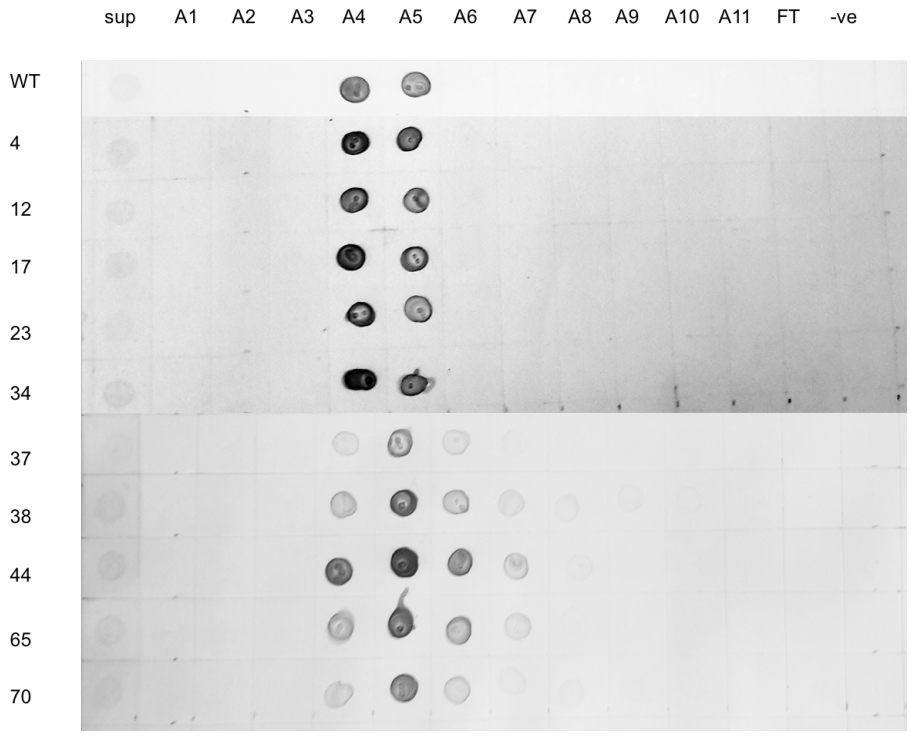
Once optimised, the protocol and conditions were repeated exactly for each mutant. The size of the peaks for most mutants was comparable to that of the wild-type, however 3.4 mutants 36, 77, and 3.8 mutants 12 and 34 had notably lower yields. High expressing mutants were 3.4 mutants 47 and 56 as well as 3.8 mutant 44. Protein yields are summarised in Tables (4.3) and (4.4) and FPLC peaks are in Appendix (J).

Immunoblot of FPLC fractions

Fractions were collected and immunoblotted with αhIgG-Ap to show that the protein was present in the fraction corresponding to the FPLC peak (Figure 4.8). Fractions containing protein were pooled and concentrated by Vivaspin 2 ultrafiltration columns.



(a) Immunoblot of 3.4 DBL-Fc mutant FPLC fractions.



(b) Immunoblot of 3.8 DBL-Fc mutant FPLC fractions.

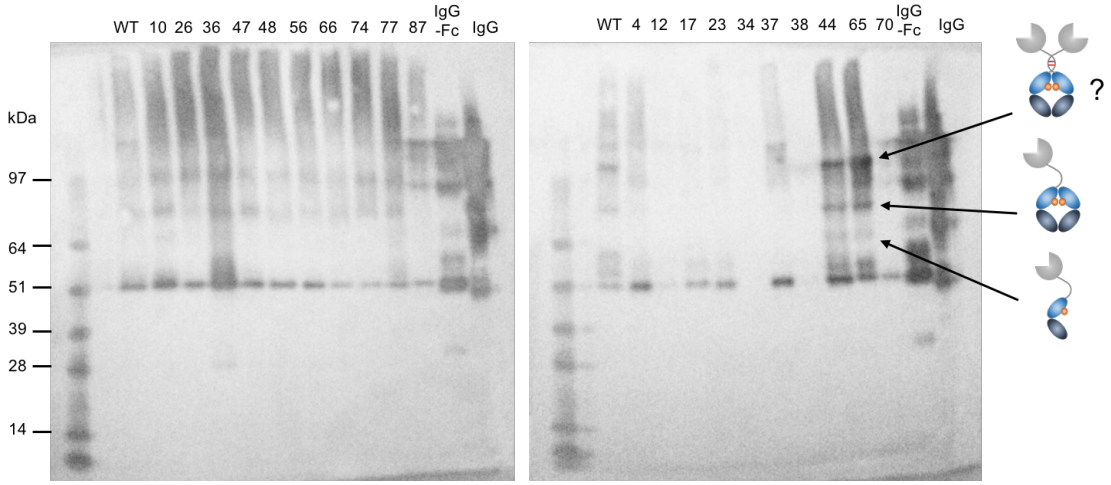
Figure 4.8: **Immunoblot of DBL mutant FPLC fractions.** Fractions A1-A11 from the FPLC elution were tested by immunoblotting for the presence of protein. These were compared to a blot of the supernatant (sup) prior to purification as well as the flow-through (FT) to check that all the protein had been removed. Media was used as a negative control (-ve). Immunoblot of (a) 3.4 DBL and (b) 3.8 DBL mutants fractions after FPLC purification.

4.4.2.2 Characterisation of recombinant DBL mutants

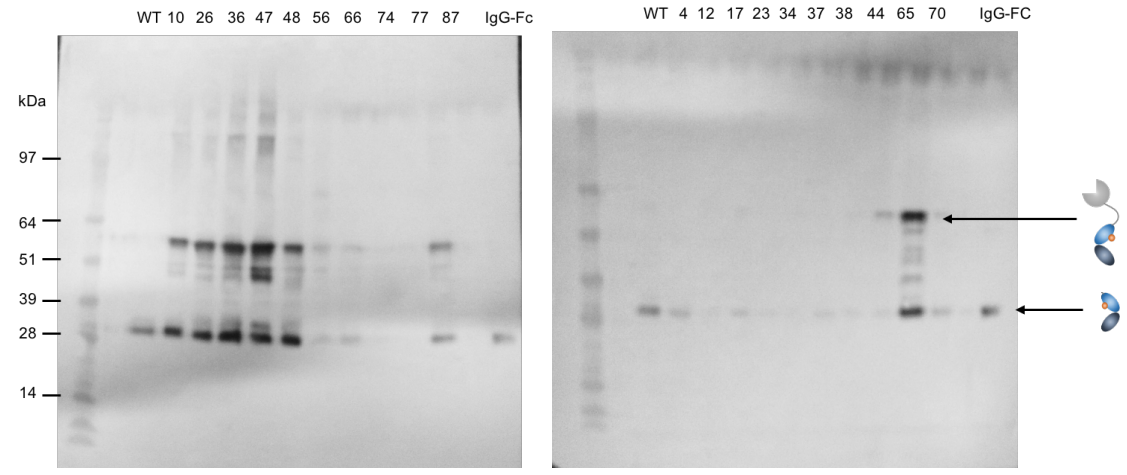
Variations were observed in protein expression for the recombinant DBL mutants. In order to check whether there were differences in protein folding caused by mutations, Western blot analysis was performed.

αIgG-Fc Western blot

Western blots using an αIgG-Fc to detect Fc-fusion proteins showed that protein yield was low (particularly for 3.8 DBL mutants). There was limited formation of homodimeric Fc-fusion proteins, and mostly monomeric DBL fusions were produced (Figure 4.9).



(a) Non-reduced Western blot of DBL 3.4 mutants. (b) Non-reduced Western blot of DBL 3.8 mutants.



(c) Reduced Western blot of DBL 3.4 mutants. (d) Reduced Western blot of DBL 3.8 mutants.

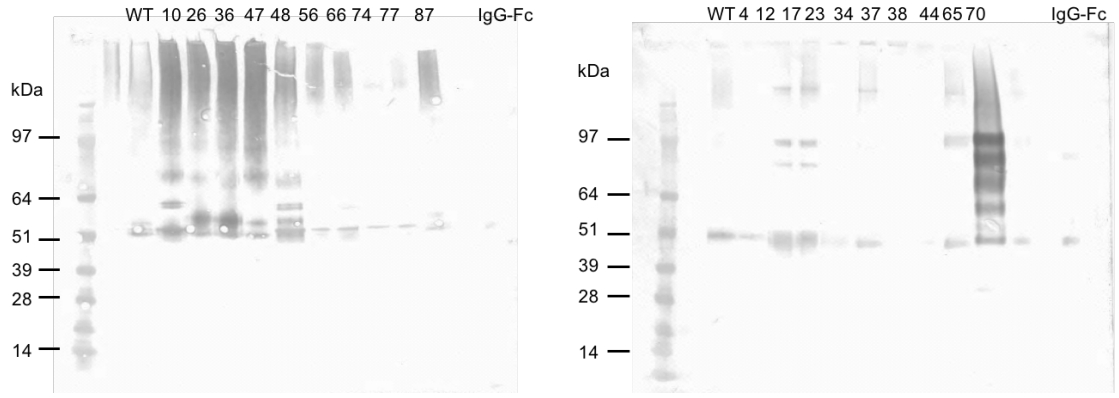
Figure 4.9: **Non-reduced and reduced Western blots of 3.4 and 3.8 DBL mutants detected using αIgG-Fc.** Protein expression and folding varied across the mutants. Monomeric DBL fusions were produced by 3.4 DBL mutants and some 3.8 DBL mutants, but protein yield was low.

All 3.4 DBL mutants produced monomeric recombinant protein. Protein production appears to be particularly difficult for 3.8 DBL compared to the 3.4 DBL domain as yields are low. Mutants 3.8 DBL mutants 44 and 65 produced more monomeric Fc-fusion protein compared to other 3.8 DBL mutants. It is possible that a band visible at ∼120 kDa is homodimer. This band is also present in 3.4 DBL mutant 87 as well as a faint band for the wild type DBL. This shows an improvement from protein produced in Chapter 3 and shows successful optimisation of the methods.

3.8 DBL mutant 44, which contains two asparagine substitutions (located in the loops between helices h1a-h1 and h5-h6) has previously been identified as a high-expressing mutant. It is likely that the lack of monomeric Fc produced for the other 3.8 DBL mutants is due to low yield.

αMSPDBL Western blot

To further characterise the protein produced, an αMSPDBL Western blot was used with detection from αMSPDBL1 and αMSPDBL2 polyclonal antibodies (kindly provided by Cecile Crosnier, Sanger Institute). This provided evidence of some homodimer formation as a band was seen at 120 kDa for some mutants (Figure 4.10). Bands seen at 50 kDa which were detected in αIgG-Fc Western blot analysis suggest that there is more improper folding of the Fc-fusion protein than initially observed. This band could be a monomeric fusion protein with improper folding of the Fc.

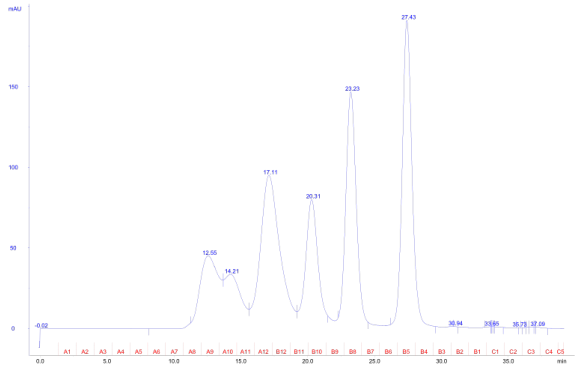


(a) Non-reduced Western blot of 3.4 DBL mutants (b) Non-reduced Western blot of 3.8 DBL mutants

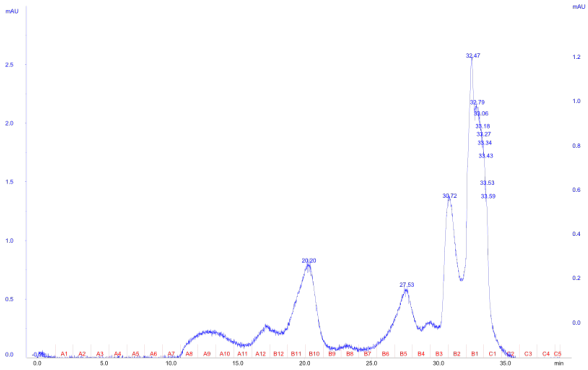
Figure 4.10: **Non-reduced Western blots of 3.4 and 3.8 DBL mutants detected using αMSPDBL polyclonal antibodies.** Western blot analysis provided evidence of some homodimer formation as a band was detected in some mutants at ∼ 120 kDa.

Size exclusion

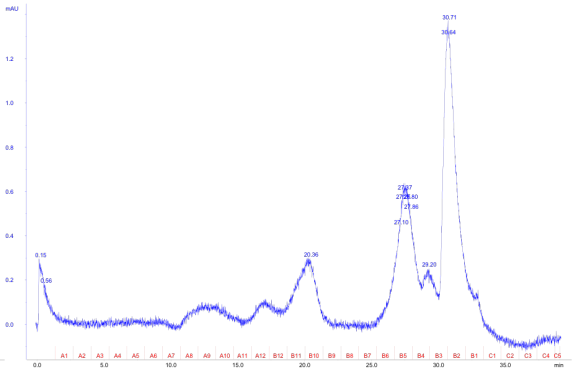
In order to investigate the native structure of the fusion proteins, the recombinant protein was run by gel filtration (Section 3.3.12). Peaks were spread across fractions A8 to B5 (corresponding to 700 - 1 kDa) as seen in the initial wild-type DBL-fusions produced in Chapter 3 (Figure 4.11). This again suggests that protein aggregates are being formed. Two high-expressing mutants 3.4 DBL 56 and 3.8 DBL 44 (Figures 4.11c and 4.11e respectively) were investigated and their size exclusion profiles found to be similar to the wild types (Figures 4.11b and 4.11b).



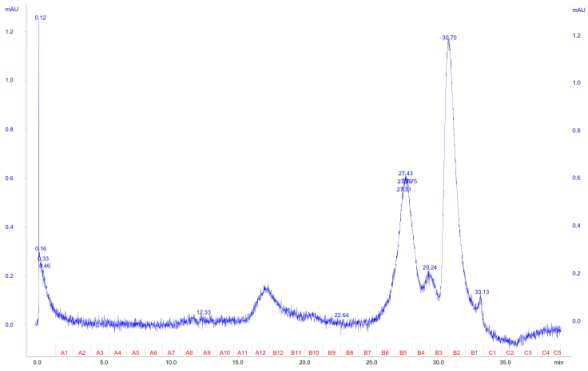
(a) BioRad size exclusion standards



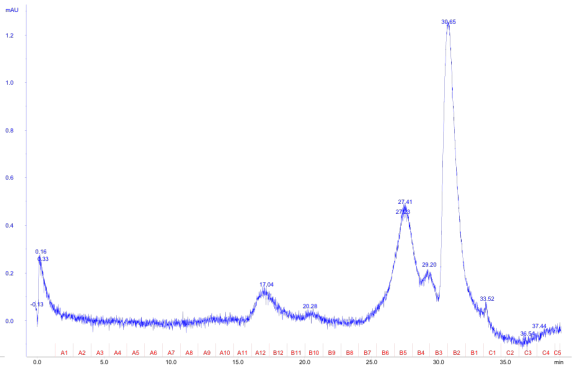
(b) 3.4 DBL WT



(c) 3.4 DBL mutant #56



(d) 3.8 DBL WT



(e) 3.8 DBL mutant #44

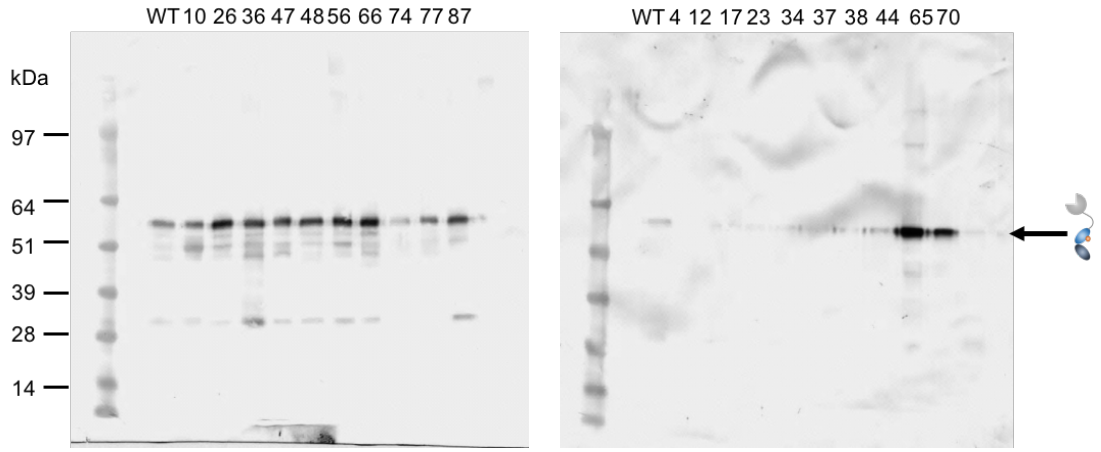
Figure 4.11: **Size exclusion profiles from DBL mutants.** (a) Peaks for 100 µl protein standards. (b) 3.4 DBL WT size exclusion profile and (c) 3.4 DBL mutant 56 size exclusion profile. (d) 3.8 DBL WT size exclusion profile and (e) 3.8 DBL mutant 44 size exclusion profile.

4.4.2.3 IgM-binding analysis of MSPDBL Mutants

In Chapter 3 it was shown that the recombinant DBL Fc-fusion proteins bind human IgM by ELISA. Increased binding was observed for 3.4 DBL compared to 3.8 DBL which is in agreement with the literature where 3.8 DBL binds with ten-fold affinity (Crosnier et al., 2016). A panel of recombinant mutant DBL Fc-fusion proteins has been created which were further investigated to identify whether any mutations had an impact on IgM-binding.

αIgM Western blot

IgM-binding of the DBL mutants was first investigated using Western blot analysis (Figure 4.12). DBL domains were pre-incubated with hIgM and then pulled out using αIgM agarose beads. These were reduced and added to SDS-PAGE before transferring to nitrocellulose and blotting with αIgG-Fc (Section 4.3.8). Protein detected was monomeric, although a lower band which is also present in αIgG-Fc Western blot analysis (Figure 4.9) suggests that the fusion proteins have difficulty folding. This band may correspond to an incorrectly folded monomeric Fc-fusion.



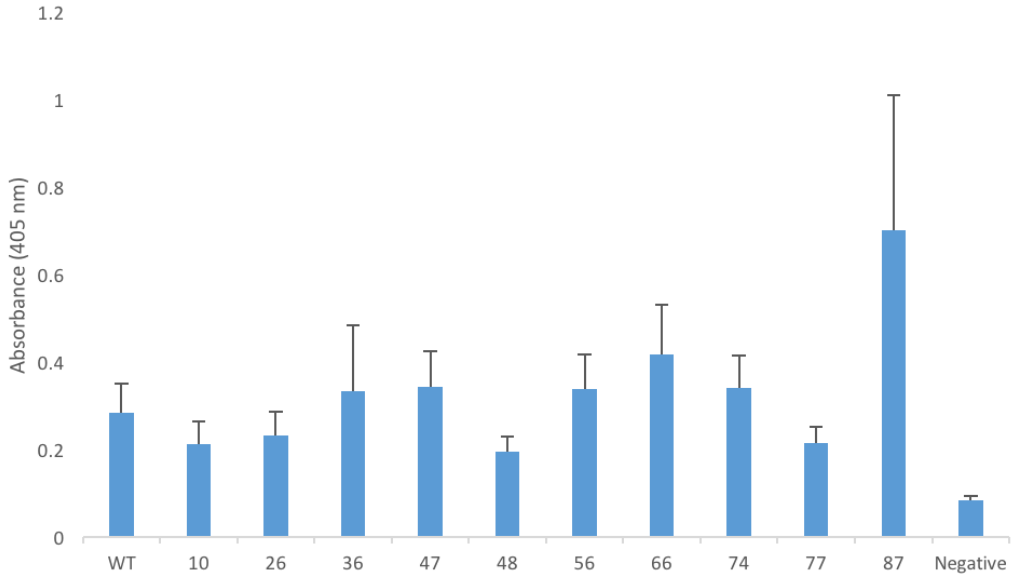
(a) 3.4 DBL mutants (b) 3.8 DBL mutants

Figure 4.12: **Reduced Western blots of 3.4 and 3.8 DBL Fc-fusion mutants bound to hIgM detected using αIgG-Fc.** DBL Fc-fusion proteins were pre-incubated with hIgM and pulled out with αhIgM-agarose before being added under reducing conditions to SDS-PAGE and transferred to nitrocellulose. Protein was detected using αIgG-Fc.

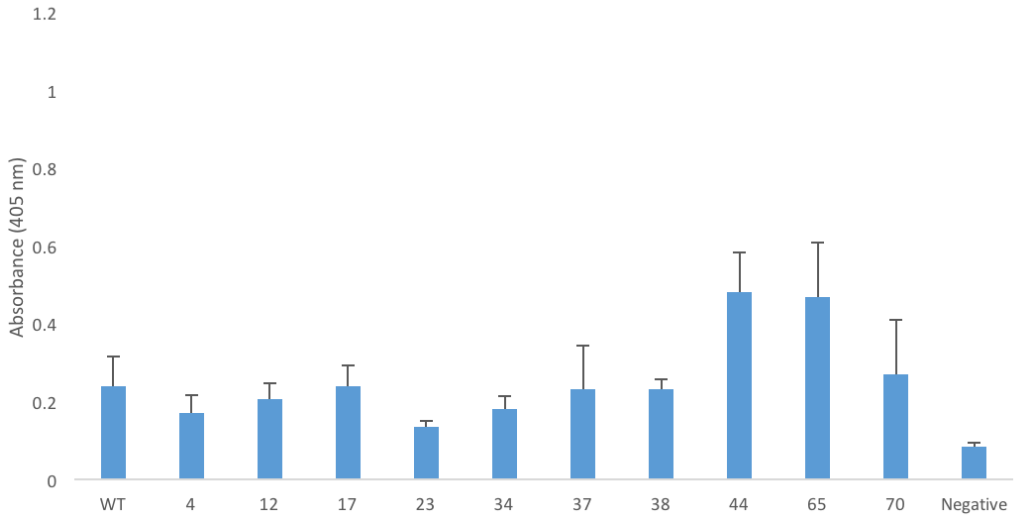
IgM binding ELISA

The IgM-binding sandwich ELISA used in Chapter 3 was repeated for the DBL mutants in order to compare their IgM-binding function compared to the wild-type proteins (Figure 4.13). Interestingly, there was no loss of binding and there appeared to be improved binding for 3.4 DBL mutant 87 which contains a single mutation N201Y located in helix 2a of the second subdomain. 3.8 DBL mutants 44 and 65 also appeared to improve binding. This could be due to the presence of homodimers as detected by Western blot analysis.

However, due to high variation in experimental repeats, possibly due the unquantifiable protein fragments present, these results were not statistically significant (data not shown).



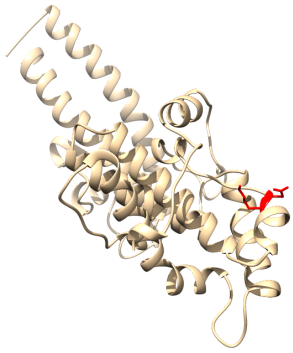
(a) IgM-binding of recombinant 3.4 DBL mutant Fc-fusion proteins.



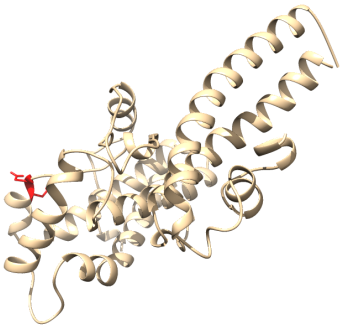
(b) IgM-binding of recombinant 3.8 DBL mutant Fc-fusion proteins.

Figure 4.13: **IgM-binding analysis of (a) 3.4 DBL and (b) 3.8 DBL mutants.** A sandwich ELISA was used to characterise IgM-binding of recombinant mutant DBL Fc-fusion proteins compared to the wild-type DBL fusion proteins. The mean ± SD are shown for four experiments.

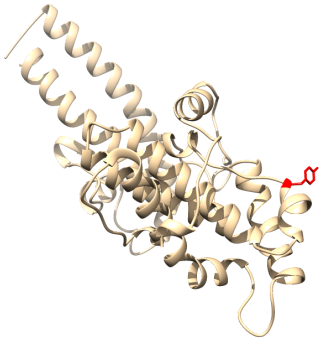
The location of the N201Y mutation (helix 2a) is not an area predicted by the binding model, as helix 2a is thought to be buried in the DBL structure (Figure 4.13). Structural analysis of DBL mutants predicted that mutations E318V in 3.4 DBL mutant 66 and N321Y in 3.4 DBL mutant 74 might have an impact on IgM binding, as well as T335S in 3.8 DBL mutant 17, K330E in 3.8 DBL mutant 34 and N339D in 3.8 DBL mutant 44 (Appendix D). However, this study was not successful in experimentally investigating this.



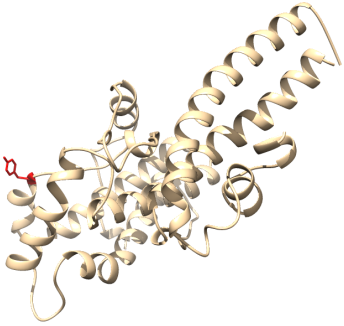
(a) 3.4 DBL WT asparagine at position 201



(b) 3.4 DBL WT



(c) 3.4 DBL mutant 87, tyrosine at position 201



(d) 3.4 DBL mutant 87

Figure 4.14: **3.4 DBL mutant 87 (N201Y).** 3.4 mutant 87 has an Asparagine mutation in helix 2a which may improve IgM binding.

The experimental results were disappointing, therefore protein interface prediction was used next to provide structural insight into the IgM-binding region.

4.4.3 Protein-Protein Interaction Prediction (PPIP) analysis

There are a range of web-based servers available for protein-interface prediction, by identifying conserved residues and the nature of these residues, as well as entropy. Surface residues are then scored for their properties and binding interfaces are defined as clusters

of high scoring residues.

Meta-PPISP was used to predict binding interfaces in the merozoite DBL domains (Appendix G). This server combines the results from three further servers cons-PPISP (`http://pipe.scs.fsu.edu/ppisp.html`), Promate (`http://bioportal.weizmann.ac.il/promate`), and PINUP (`http://sparks.informatics.iupui.edu/PINUP/`) in a linear regression analysis and shows improved performance compared to the individual methods (Qin and Zhou, 2007).

Meta-PPISP predicted two main binding interfaces in both MSPDBL1 and MSPDBL2 (loops h1a-h1 and h1-h2a in MSPDB2 DBL domain and h2a, h4 in MSPDBL1 DBL domain). Interestingly, helix 2a is predicted as a common binding region for both proteins and may therefore correspond to the IgM binding site.

4.4.4 Sequence analysis of IgM-binding domains

Since the experimental analysis of the library of DBL mutants proved to be inconclusive and hampered by limited protein yield and lack of proper folding, we next turned to sequence analysis for further insight into residues critical for IgM-binding. Sequence alignments for all mutants expressed can be found in Appendix E.

Eight *P. falciparum* DBL domains have been shown to bind IgM in addition to MSPDBL2 and MSPDBL2 (Table 4.1). In a recent study by Crosnier et al. (2016), IgM binding was investigated in eleven MSPDBL1 isolates and nine were found to bind, whereas isolates 7G8 and 028 did not bind. By sequence alignment it was observed that a Glu residue at position 310 was common to both non-binding isolates. However when this residue was substituted for the Gln residue present in isolate 384 (known to bind IgM) by mutagenesis, this mutation was not sufficient to restore IgM-binding (Crosnier et al., 2016).

Phylogenetic analysis of all known DBLε IgM-binding and non IgM-binding DBL domains showed a lack of clustering (Figure 4.15). Sequence alignments of fifteen known non-binders (Figure 4.16) and ten IgM-binding domains (Figure 4.17) is also consistent with this result. Analysis of conserved regions revealed that these regions correspond to homology blocks which make up the core structure of all DBL domains (Figures 4.18 and 4.19). This suggests that the quaternary structure of these DBL domains is highly important for IgM binding in the context of the native protein. A more comprehensive analysis awaits the publication of thousands of *P. falciparum* genomes that are currently being analysed (Pleass et al., 2015).

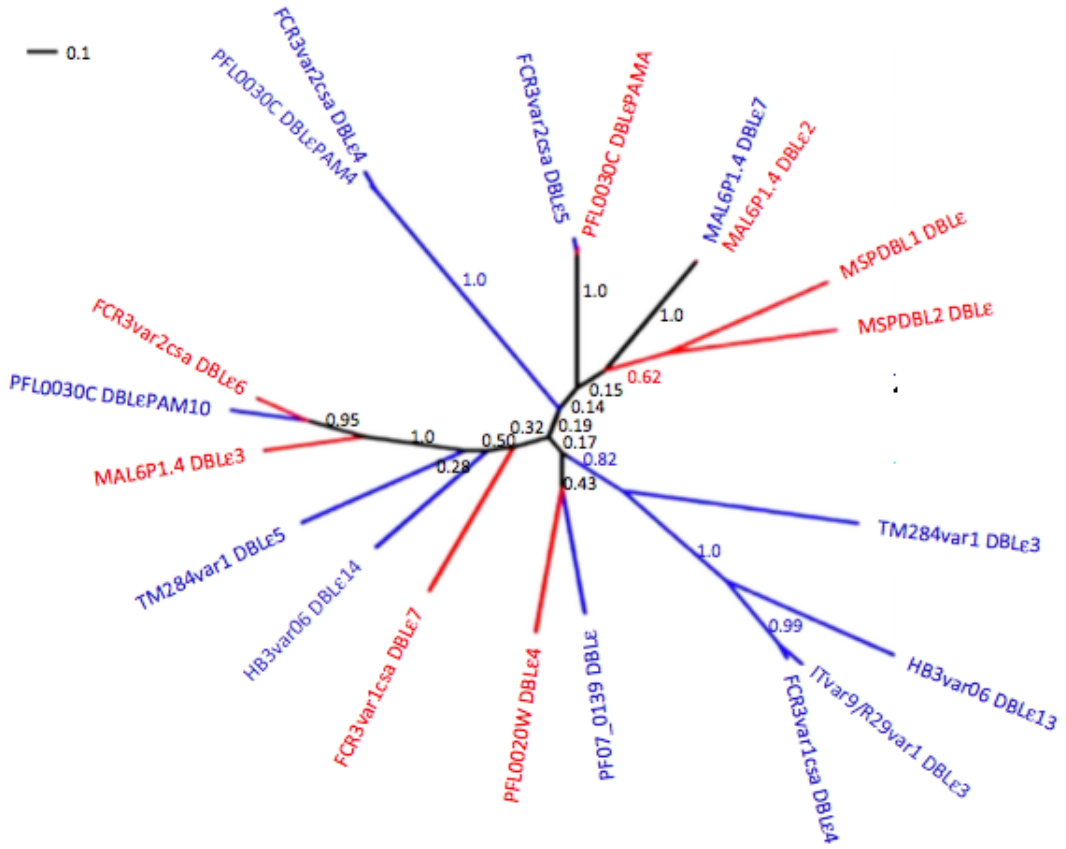
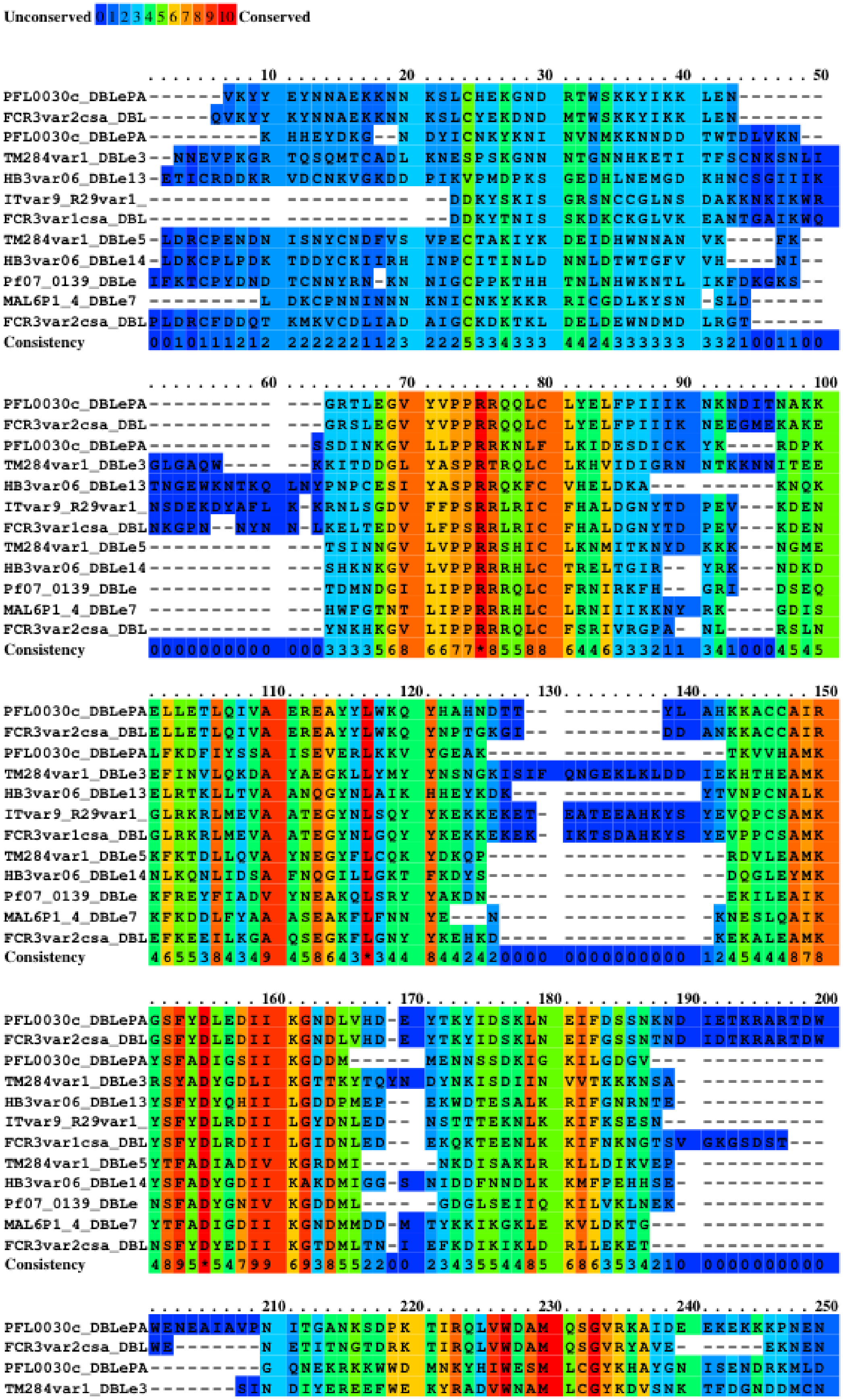


Figure 4.15: **Phylogenetic Relationships between IgM Binding and Non-Binding DBLε Domains.** Phylogenetic analysis of eight IgM binding (red) and twelve non-binding (blue) DBLε domains shows a lack of clustering. Bipartition bootstrap support (%) is indicated by edge numbers. Amino acid substitutions per site are indicated by scale bar (Pleass et al., 2015).



Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
10 20 30 40 50
PFL0030c_DBLePA ------VKYY EYNNAEKKNN KSLCHEKGND RTWSKKYIKK LEN-------
FCR3var2csa_DBL -----QVKYY KYNNAEKKNN KSLCYEKDND MTWSKKYIKK LEN-------
PFL0030c_DBLePA ---------K HHEYDKG--N DYICNKYKNI NVNMKKNNDD TWTDLVKN--
TM284var1_DBLe3 --NNEVPKGR TQSQMTCADL KNESPSKGNN NTGNNHKETI TFSCNKSNLI
HB3var06_DBLe13 -ETICRDDKR VDCNKVGKDD PIKVPMDPKS GEDHLNEMGD KHNCSGIIIK
ITvar9_R29var1_ ---------- ---------- --DDKYSKIS GRSNCCGLNS DAKKNKIKWR
FCR3var1csa_DBL ---------- ---------- --DDKYTNIS SKDKCKGLVK EANTGAIKWQ
TM284var1_DBLe5 -LDRCPENDN ISNYCNDFVS VPECTAKIYK DEIDHWNNAN VK----FK--
HB3var06_DBLe14 -LDKCPLPDK TDDYCKIIRH INPCITINLD NNLDTWTGFV VH----NI--
Pf07_0139_DBLe IFKTCPYDND TCNNYRN-KN NIGCPPKTHH TNLNHWKNTL IKFDKGKS--
MAL6P1_4_DBLe7 ---------L DKCPNNINNN KNICNKYKKR RICGDLKYSN -SLD------
FCR3var2csa_DBL PLDRCFDDQT KMKVCDLIAD AIGCKDKTKL DELDEWNDMD LRGT------
Consistency 0010111212 2222221123 2225334333 4424333333 3321001100
60 70 80 90 100
PFL0030c_DBLePA ---------- ---GRTLEGV YVPPRRQQLC LYELFPIIIK NKNDITNAKK
FCR3var2csa_DBL ---------- ---GRSLEGV YVPPRRQQLC LYELFPIIIK NEEGMEKAKE
PFL0030c_DBLePA ---------- --SSDINKGV LLPPRRKNLF LKIDESDICK YK----RDPK
TM284var1_DBLe3 GLGAQW---- --KKITDDGL YASPRTRQLC LKHVIDIGRN NTKKNNITEE
HB3var06_DBLe13 TNGEWKNTKQ LNYPNPCESI YASPRRQKFC VHELDKA--- ------KNQK
ITvar9_R29var1_ NSDEKDYAFL K-KRNLSGDV FFPSRRLRIC FHALDGNYTD PEV---KDEN
FCR3var1csa_DBL NKGPN--NYN N-LKELTEDV LFPSRRLRIC FHALDGNYTD PEV---KDEN
TM284var1_DBLe5 ---------- ---TSINNGV LVPPRRSHIC LKNMITKNYD KKK---NGME
HB3var06_DBLe14 ---------- ---SHKNKGV LVPPRRRHLC TRELTGIR-- YRK---NDKD
Pf07_0139_DBLe ---------- ---TDMNDGI LIPPRRRQLC FRNIRKFH-- GRI---DSEQ
MAL6P1_4_DBLe7 ---------- ---HWFGTNT LIPPRRRHLC LRNIIIKKNY RK----GDIS
FCR3var2csa_DBL ---------- ---YNKHKGV LIPPRRRQLC FSRIVRGPA- NL----RSLN
Consistency 0000000000 0003333568 6677*85588 6446333211 3410004545
110 120 130 140 150
PFL0030c_DBLePA ELLETLQIVA EREAYYLWKQ YHAHNDTT-- --------YL AHKKACCAIR
FCR3var2csa_DBL ELLETLQIVA EREAYYLWKQ YNPTGKGI-- --------DD ANKKACCAIR
PFL0030c_DBLePA LFKDFIYSSA ISEVERLKKV YGEAK----- ---------- --TKVVHAMK
TM284var1_DBLe3 EFINVLQKDA YAEGKLLYMY YNSNGKISIF QNGEKLKLDD IEKHTHEAMK
HB3var06_DBLe13 ELRTKLLTVA ANQGYNLAIK HHEYKDK--- ---------- YTVNPCNALK
ITvar9_R29var1_ GLRKRLMEVA ATEGYNLSQY YKEKKEKET- EATEEAHKYS YEVQPCSAMK
FCR3var1csa_DBL GLRKRLMEVA ATEGYNLGQY YKEKKEKEK- IKTSDAHKYS YEVPPCSAMK
TM284var1_DBLe5 KFKTDLLQVA YNEGYFLCQK YDKQP----- ---------- --RDVLEAMK
HB3var06_DBLe14 NLKQNLIDSA FNQGILLGKT FKDYS----- ---------- --DQGLEYMK
Pf07_0139_DBLe KFREYFIADV YNEAKQLSRY YAKDN----- ---------- --EKILEAIK
MAL6P1_4_DBLe7 KFKDDLFYAA ASEAKFLFNN YE---N---- ---------- -KNESLQAIK
FCR3var2csa_DBL EFKEEILKGA QSEGKFLGNY YKEHKD---- ---------- -KEKALEAMK
Consistency 4655384349 458643*344 8442420000 0000000000 1245444878
160 170 180 190 200
PFL0030c_DBLePA GSFYDLEDII KGNDLVHD-E YTKYIDSKLN EIFDSSNKND IETKRARTDW
FCR3var2csa_DBL GSFYDLEDII KGNDLVHD-E YTKYIDSKLN EIFGSSNTND IDTKRARTDW
PFL0030c_DBLePA YSFADIGSII KGDDM----- MENNSSDKIG KILGDGV--- ----------
TM284var1_DBLe3 RSYADYGDLI KGTTKYTQYN DYNKISDIIN VVTKKKNSA- ----------
HB3var06_DBLe13 YSFYDYQHII LGDDPMEP-- EKWDTESALK RIFGNRNTE- ----------
ITvar9_R29var1_ YSFYDLRDII LGYDNLED-- NSTTTEKNLK KIFKSESN-- ----------
FCR3var1csa_DBL YSFYDLRDII LGIDNLED-- EKQKTEENLK KIFNKNGTSV GKGSDST---
TM284var1_DBLe5 YTFADIADIV KGRDMI---- -NKDISAKLR KLLDIKVEP- ----------
HB3var06_DBLe14 YSFADYGDII KAKDMIGG-S NIDDFNNDLK KMFPEHHSE- ----------
Pf07_0139_DBLe NSFADYGNIV KGDDML---- -GDGLSEIIQ KILVKLNEK- ----------
MAL6P1_4_DBLe7 YTFADIGDII KGNDMMDD-M TYKKIKGKLE KVLDKTG--- ----------
FCR3var2csa_DBL NSFYDYEDII KGTDMLTN-I EFKDIKIKLD RLLEKET--- ----------
Consistency 4895*54799 6938552200 2343554485 6863534210 0000000000
210 220 230 240 250
PFL0030c_DBLePA WENEAIAVPN ITGANKSDPK TIRQLVWDAM QSGVRKAIDE EKEKKKPNEN
FCR3var2csa_DBL WE-------N ETITNGTDRK TIRQLVWDAM QSGVRYAVE- -----EKNEN
PFL0030c_DBLePA ---------G QNEKRKKWWD MNKYHIWESM LCGYKHAYGN ISENDRKMLD
TM284var1_DBLe3 -------SIN DIYEREEFWE KYRADVWNAM LCGYKDVSNK TFDGNDDMCN

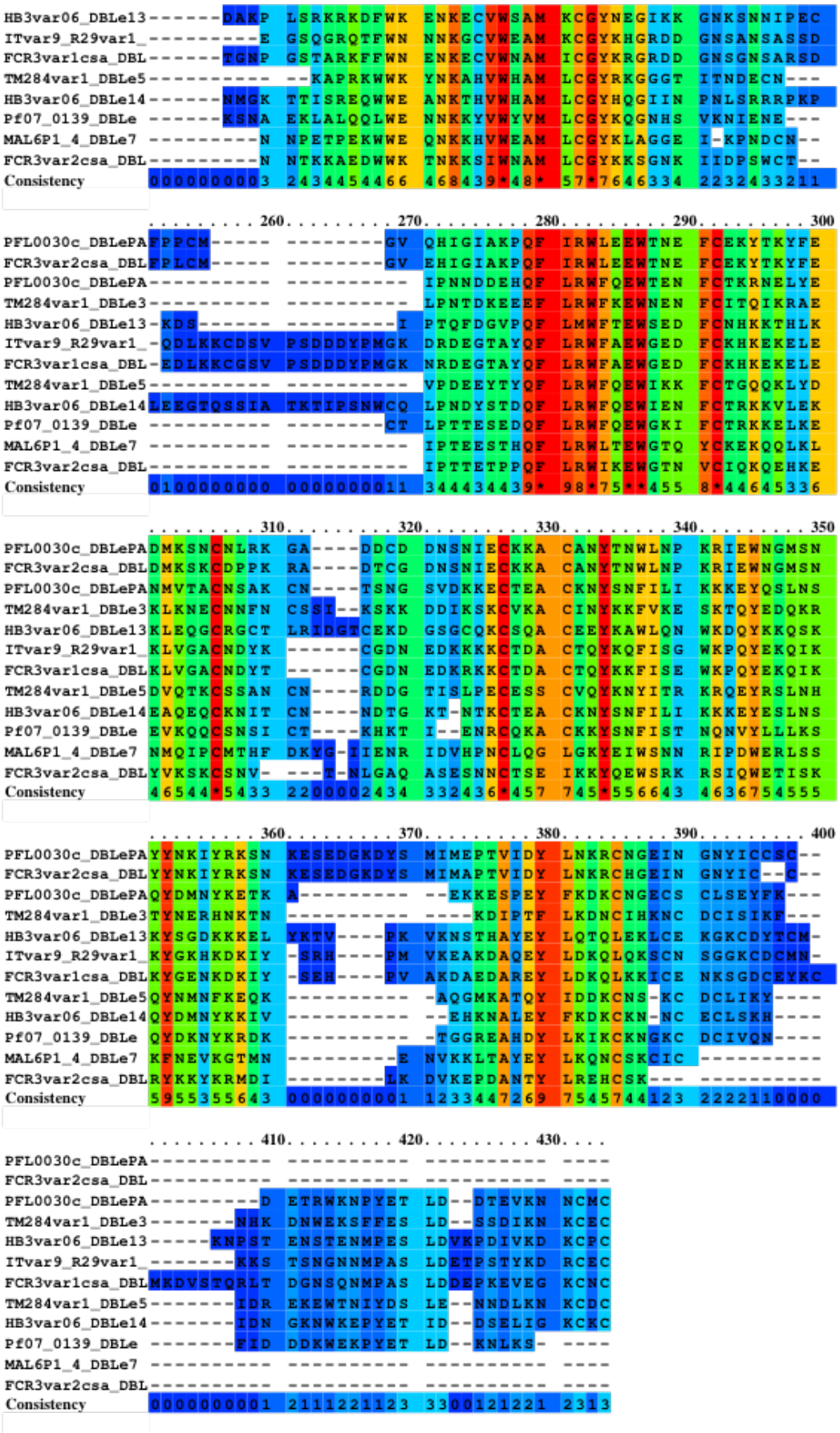
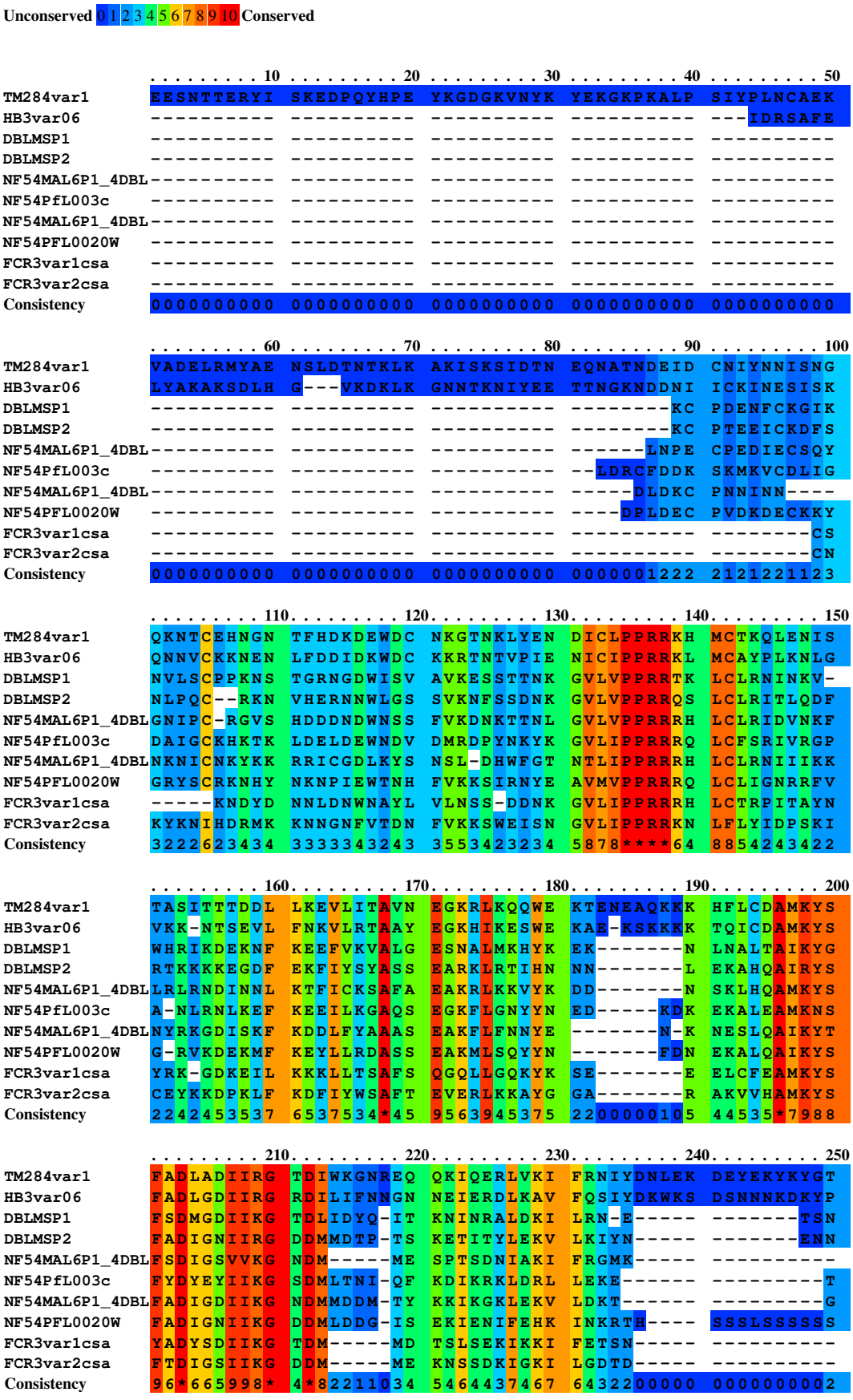


Figure 4.16: **Sequence alignment of known non IgM binding DBL domains highlighting sequence conservation.** Sequences are conserved between non IgM-binding DBL domains in locations of homology blocks common to all DBL domains.


Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
TM284var1
HB3var06
DBLMSP1
DBLMSP2
NF54MAL6P1_4DBL
NF54PfL003c
NF54MAL6P1_4DBL
NF54PFL0020W
FCR3var1csa
FCR3var2csa
Consistency

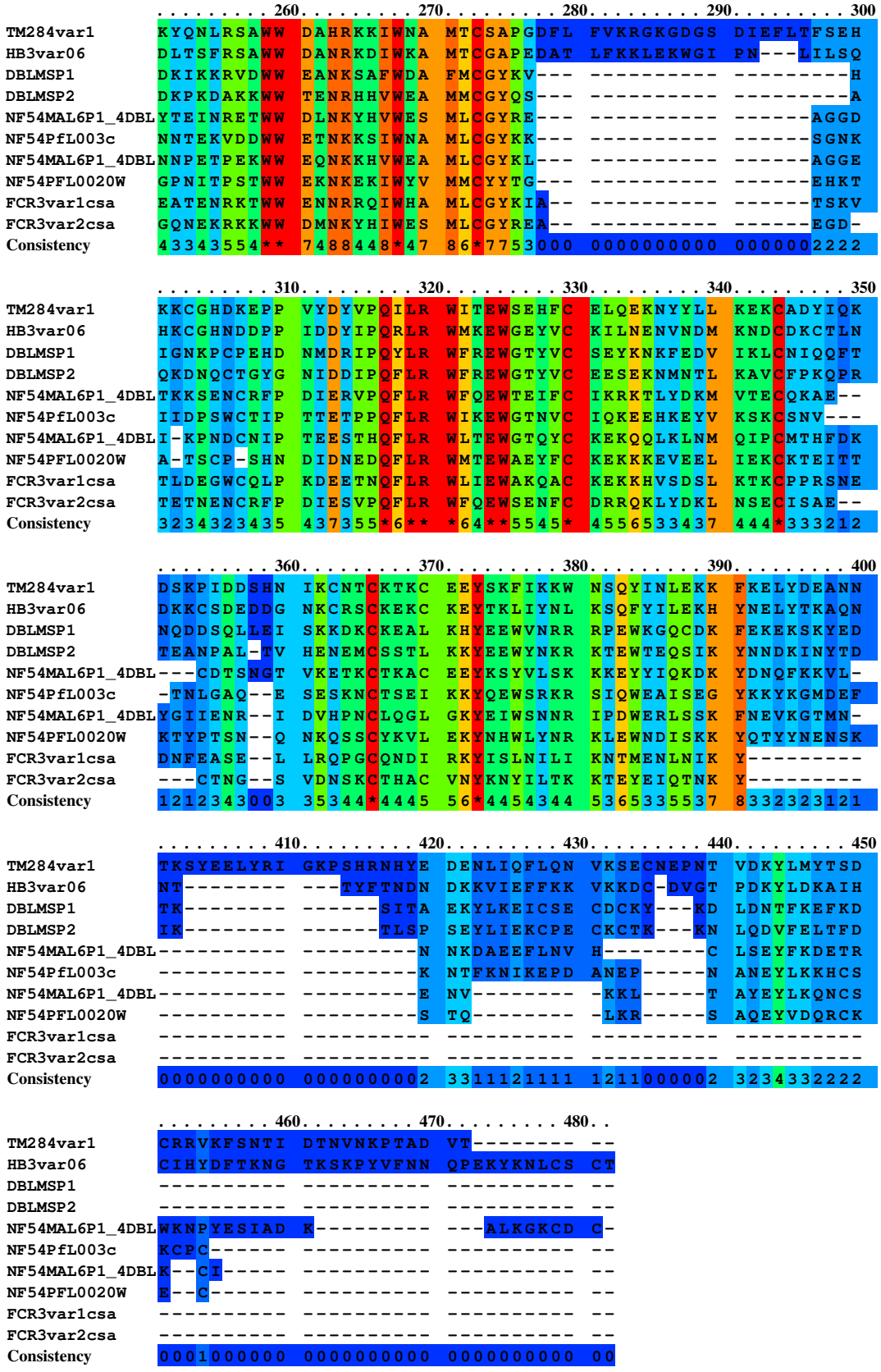
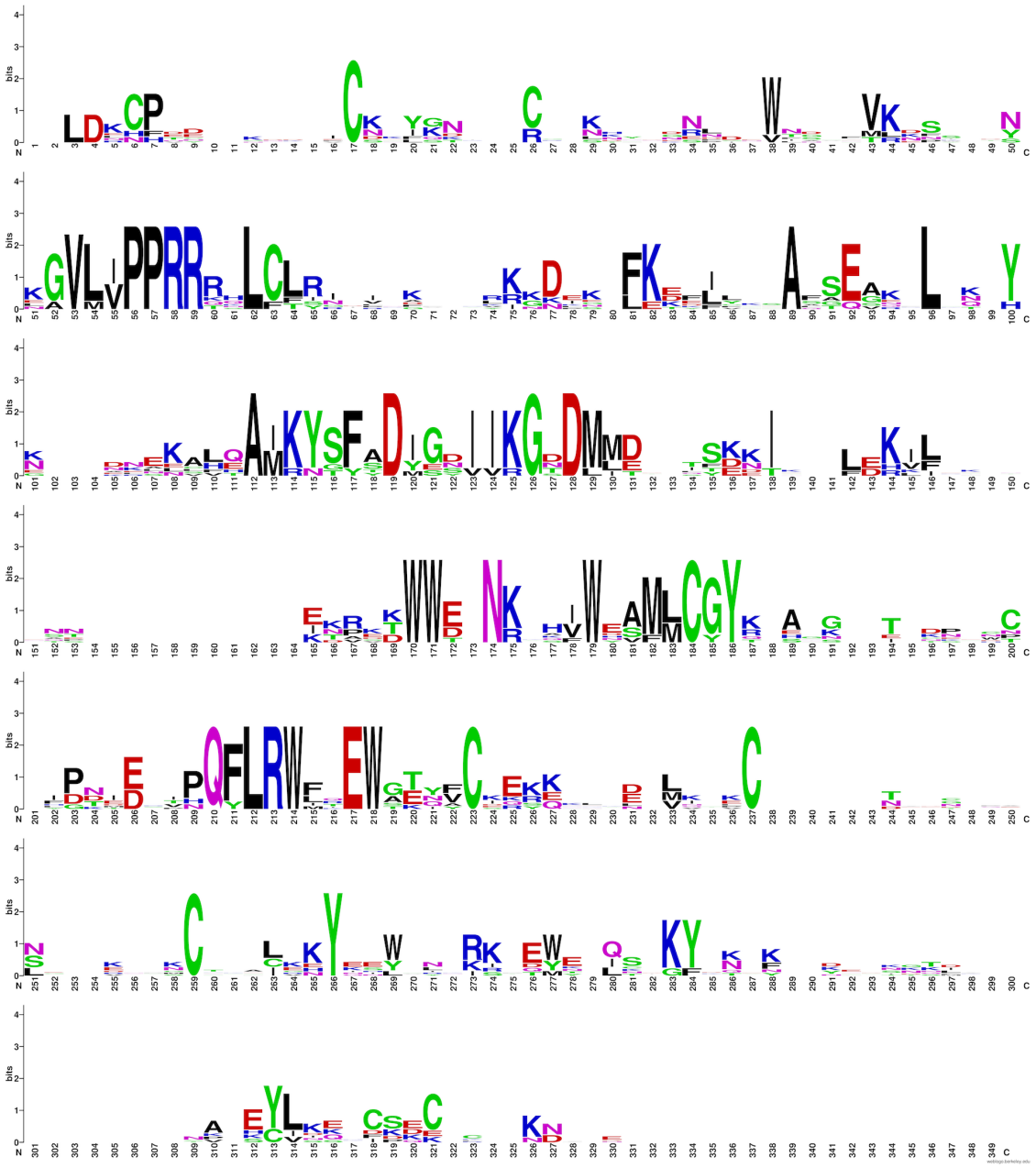
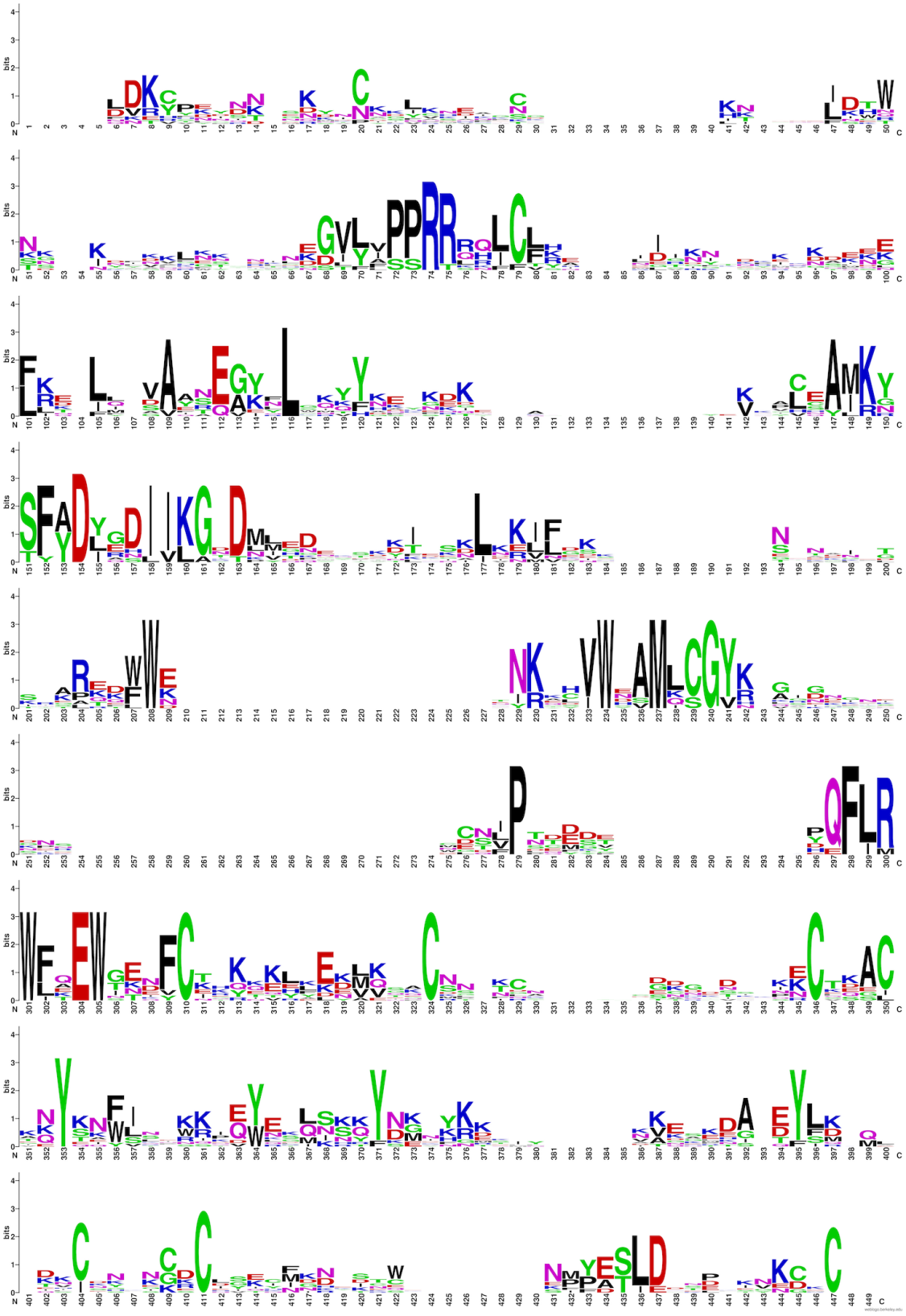


Figure 4.17: **Sequence alignment of known IgM binding DBL domains highlighting sequence conservation.** Sequences are conserved between IgM binding DBL domains in locations of homology blocks common to all DBL domains.



(a) IgM-binding DBL domains



(b) Non IgM-binding DBL domains

Figure 4.18: Sequence conservation logos for known IgM binding and non-binding DBL domains highlight only core DBL homology blocks.

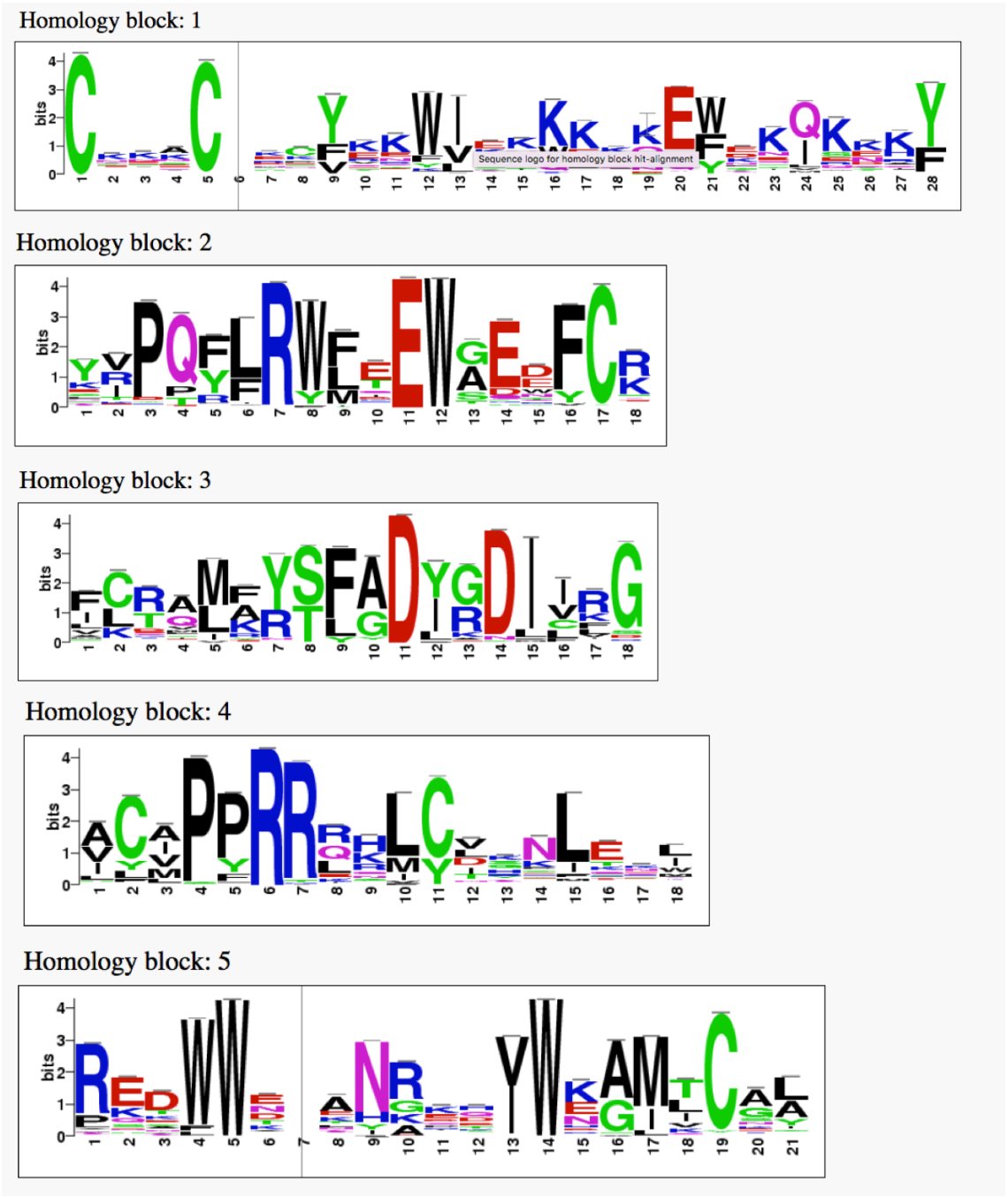


Figure 4.19: **Homology blocks common to all DBL domains.** Homology blocks 1 - 5 form the core structure of DBL domains and are the most conserved in the seven genomes analysed by Rask et al. (2010).

To further investigate whether conserved regions could be identified that are common to IgM-binding domains, homology block (HB) analysis was performed based on the homology blocks characterised by Rask et al. (2010). HBs contained in the known IgM-binding and non-binding DBL domains described by Pleass et al. (2015), were identified. HBs common to all DBL domains (HB1 - HB5) were removed from the analysis, as were HBs which only occurred once. A data mining search then was performed to identify binding motifs amongst HBs common only to IgM-binding DBL domains (Appendix H) but no HB motifs unique to these domains were identified. Some single HBs were unique to either IgM-binding or non-binding DBL domains, however these HBs were considered rare from the seven-genome analysis conducted by Rask et al. (2010). In such cases, the HB was contained in two DBL domains from the same domain class: such domains might be expected to have sequence homology and structural similarities.

4.4.5 Structural analysis of IgM-binding DBL domains

Since sequence analysis suggested that quaternary structure of the DBL domains was important for binding, we aimed to use structural analysis to identify binding regions. The predicted effect of DBL mutations using the structural model of IgM binding was inconclusive, therefore the structure of known IgM-binding DBL domains was compared to structures of known non-binders.

4.4.5.1 Structural analysis of IgM binding and non-binding MSPDBL domain isolates

Although sequence analysis of known IgM binding DBL isolates did not highlight significant residues (Crosnier et al., 2016). The structures of IgM binding DBLMSP and DBLMSP2 isolates were modelled based on the known crystal structure of the DBL domains from MSPDBL2. The non-binding DBL domains (isolate 028 and strain 7G9 in DBLMSP) were superimposed and highlighted in red (Figure 4.20) (Crosnier et al., 2016). This highlights differences in structure which are likely to be involved in IgM-binding and were observed in helices h1, h2a, h4 and h7 (Crosnier et al., 2016). There are, however, polymorphisms in h7 frequently observed around the same positions amongst DBL domains that have retained their IgM-binding ability, therefore the IgM-binding site is unlikely to be in h7 (Crosnier et al., 2016).

4.4.5.2 Structural analysis of all known IgM binding and non-binding DBL domains

The structures of all known IgM binding DBL domains and non-binding DBL domains (Table 4.1) were similarly superimposed, with binders (red) superimposed on non-binding domains (Appendix I). Where the structure of a DBL domain was not known, it was modelled based on the most closely related DBL domain with known structure. While this highlighted the structural diversity of DBL domains, differences in structure between IgM binding and non-binding domains were not clear. The analysis was repeated for DBL domains solely of DBLε class and similarly for DBLζ class. However, the structural diversity of the DBL domains was too high to identify differences between binding and non-binding domains.

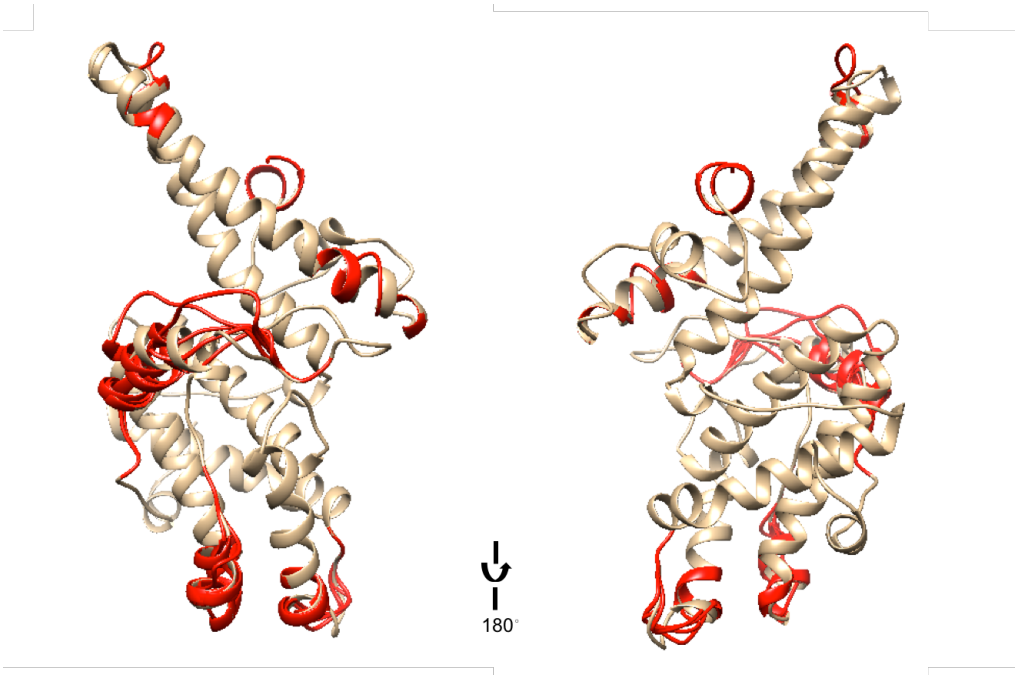


Figure 4.20: **Structural conservation of known IgM binding MSPDBL DBL domain isolates vs non-binding isolates.** DBL domains from known IgM binding MSPDBL isolates (cream) vs non-binding isolates (red). Homology models of isolates were produced using the Phyre2 software based on the crystal structure of MSPDBL2 (Kelley et al., 2015) and superimposed showing structural differences. Figure adapted from (Crosnier et al., 2016).

4.5 Summary

The aim of this chapter was to identify critical residues and binding motifs within the MSPDBL1 and MSPDBL2 DBL domains which are responsible for IgM binding. In order to do this, a library of DBL mutants was produced using the DBL-Fc fusion constructs from Chapter 3 which were mutated by random mutagenesis. Out of a total of 47 DBL mutants produced from selection of 137 positive colonies, 20 DBL mutants were selected (10 from each of MSPDBL1 and MSPDBL2) based on the molecular simulation of DBL-IgM binding kindly provided by Dan Czajkowsky (Shanghai Institute of Applied Physics, Shanghai, China).

Improved protein folding was observed and for three mutants possible homodimer formation was identified. A number of methods were optimised in this chapter in an aim to improve protein yield and folding. Firstly, for CHO-KI cell growth, HAMs growth and selection media was used which contains an additional number of supplements (such as zinc, hypoxanthine and thymidine). Cells were grown in parallel using both DMEM and HAMs and it was found that the cells grew better in HAMs.

Secondly, protein purification was optimised in order to harvest all recombinant protein produced by the cells. While HBSS is commonly used as a buffer solution for cells, sodium phosphate is a more commonly used binding buffer for affinity purification. Protein G-IgG binding has strongest affinity at pH 7.0 therefore 20 mM sodium phosphate pH 7.0 was used. The cell supernatant was diluted 1:1 with the new binding buffer prior to loading in order to adjust the supernatant to the composition of the buffer. This produced improved binding. Elution was also optimised to prevent harsh conditions for the eluted protein. A higher pH is optimal to prevent denaturing of the protein. pH was increased from 2.5 to 2.7 which still maintained effective elution while preventing denaturing.

IgM-binding of the recombinant mutant DBL domains was investigated and none of the mutants produced a knock-out of IgM binding. There was evidence that there was improved binding for the three mutants which had improved protein folding. One of these, an MSPDBL1 mutant, has an interesting mutation in which an Asparagine in helix 2a has been replaced with Tyrosine. Protein-protein interaction prediction using the online meta-PPISP server identified this region of MSPDBL1 as a protein-protein interface by analysis of surface residues. This analysis also suggested that the binding interface is different in each of MSPDBL1 and MSPDBL2.

Functional analysis of the mutant library was disappointing due to low protein yield and difficulties in protein folding. Sequence analysis was therefore used to identify common binding motifs. Phlogenetic analysis of known IgM binding and non-binding DBL domains from *P. falciparum* recently showed a lack of clustering (Pleass et al., 2015). Sequence and homology block analysis are consistent with this result.

The structures of IgM-binding MSPDBL1 and MSPDBL2 isolates were modelled using the Phyre 2 server and superimposed to highlight any differences in structure (Crosnier et al., 2016). This analysis highlighted regions in helices h1, h2a and h4, consistent with PPIP prediction results. A similar structural analysis of all known IgM binding and non-binding DBL domains in *P. falciparum* was unable to confirm this result and only highlighted the extreme structural variation between DBL domains.

Further work needs to be done to investigate the role of residues in α-helix 2a on IgM binding. This could be done experimentally using site-directed mutagenesis of the helix residues. However, in order to do this, low yield and poor folding of the DBL-Fc fusion constructs need to be addressed. In Chapter 5, an extended hinge version of the constructs was designed in order to improve protein folding and homodimer formation.

Chapter 5

Improving the DBL-Fc fusion constructs

5.1 Background

Expressing recombinant DBL protein with an Fc-fusion has exciting implications for future therapeutics. To date, no DBL domains have yet been expressed as Fc-fusion proteins, which is a newly emerging and highly promising technology, and there are no malarial Fc-fusions with FDA approval. However, a number of commercially successful Fc-fusion based therapies are currently available in the clinic with a wide range of applications from autoimmune disease to haemophilia (Summarised in Table 1.2). These exciting molecules harness the interaction of the salvage neonatal receptor (FcRn) which, by improving half-life, improves the therapeutic potential of the fused antigen. Further improvements arise from the formation of a homodimer which increases the avidity of ligand receptor interactions.

In Chapter 3, novel Fc-fusion proteins were produced from DBL domains belonging to *P. falciparum* merozoite surface proteins. DBL domains are particularly interesting in malaria as they are present in a number of protein families and in many cases are responsible for cytoadherence, a property highly associated with virulence (Singh et al., 2006; Chattopadhyay et al., 2004; Vogt et al., 2003; Flick et al., 2001; Miller et al., 1997; Chen et al., 1998, 2000; Smith et al., 1998, 2000; Barragan et al., 2000a; Buffet et al., 1999). The DBL-containing MSP proteins have been shown to localise to the merozoite surface and bind to the MSP1 complex, suggesting their involvement in erythrocyte invasion. Another interesting feature of DBL domains on the merozoite surface is their ability to bind to the Fc portion of natural IgM, a characteristic also seen in DBL domains located on the erythrocyte surface (Crosnier et al., 2016). Merozoite DBL domains were expressed as Fc-fusion proteins in order to further investigate their IgM binding properties as well as to provide proof of concept for Fc fusion proteins using malarial antigens.

In Chapter 3 the recombinant DBL-Fc fusions were shown to bind IgM. However it was identified that the recombinant proteins produced were monomeric, consisting of a single chain of Fc (Cγ2 and Cγ3 domains) fused to a single DBL domain (Figure 5.1). Although the DBL domains proved functional in relation to protein G and IgM binding, the addition of these tightly packed molecules prevented dimerisation of the Fc fusion construct. This suggests that the DBL domains at $\sim$ 35 kDa were too large and perhaps inhibited disulphide bond formation between the two chains, preventing the formation of homodimers.

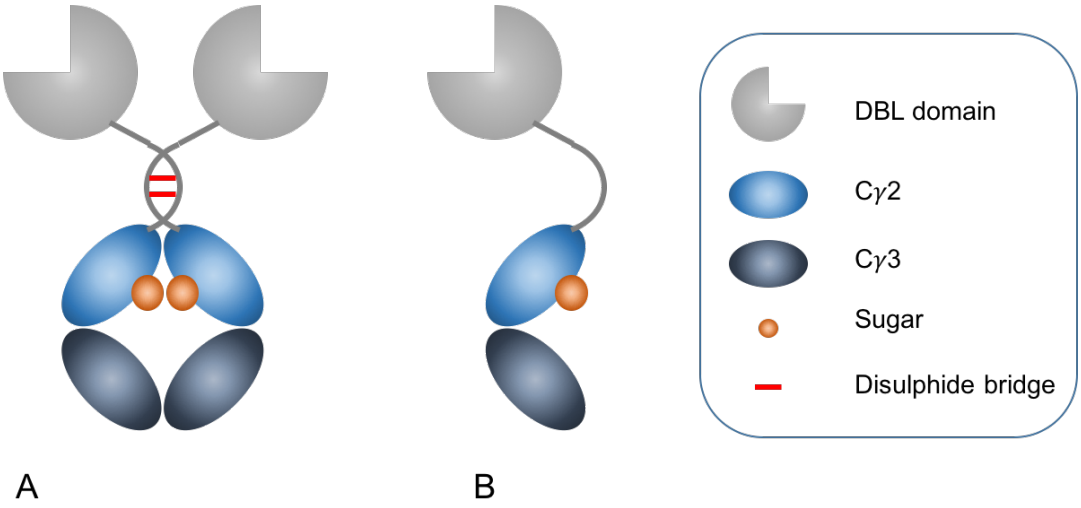


Figure 5.1: Schematic representation of DBL Fc-fusion protein. (A) The expected DBL Fc-fusion protein was a homodimer. (B) The DBL Fc-fusion protein produced in Chapter 3 is monomeric suggesting that fusion of the DBL to the IgG Fc prevents homodimer formation.

The aim of this chapter is to improve the Fc-fusion constructs created in Chapter 3 and produce dimeric DBL-Fc fusion proteins with improved IgM binding. The homodimeric DBL-Fc fusion should consist of each 25 kDa chain of the Fc (consisting Cγ2 and Cγ3 domains) linked to a single DBL domain (35 kDa). The two chains will be joined together in the hinge region by disulphide bonds to form a homodimer of 120 kDa. Most Fc-fusion proteins consist of small antigens or antigenic regions ($<$ 25 kDa) fused to the IgG Fc. The relative size of the fused antigen may therefore be critical for formation of the homodimeric Fc-fusion molecule. One of the largest Fc-fusion proteins available commercially at 200 kDa, Eloctate, is known to consist of 39% single chain molecules (Peters et al., 2013; Buyue et al., 2014). The antihaemophilic factor VIII (80 kDa) is the base protein for this Fc-fusion which has a tightly packed structure containing five globular domains (Ngo et al., 2008; Shen et al., 2008). This suggests that the size of the fusion molecule may influence dimer formation. However, the 251 kDa Fc-fusion drug rilonacept has the IL-1 receptor as the base protein and forms homodimers. The structure of this receptor is less tightly packed than the DBL domain, and its shape likened to a question mark (Schreuder et al., 1997). This suggests that the structure of the base protein also influences homodimer formation, with tightly packed structures preventing dimerisation.

The relative size of the merozoite DBL domains as well as their compact structure,

consisting of 3 subdomains made up of a tight α-helical structure, may impact on homodimer formation within the Fc. It has been shown that an extended linker region can improve the stability of fusion proteins by increasing the distance between the base protein and the tag. Fc-fusion proteins already contain a flexible hinge region from IgG1. In this chapter we have modified and increased the length of the hinge region in order to decrease the proximity of the two DBL domains from the Fc tag. It is hypothesised that this may result in improved folding of the DBL Fc-fusion.

5.1.1 Extended hinge Fc-fusion construct

Many studies have investigated the effect of linker length and composition in fusion proteins (Argos, 1990; George and Heringa, 2002; Bird et al., 1988; Alfthan et al., 1995; Robinson and Sauer, 1998; Crasto and Feng, 2000) and have been extensively reviewed by Chen et al. (2013); Zhang et al. (2009). In particular, Argos (1990) investigated natural protein-linking domains in a database of 51 protein structures in which common residues were identified. It was noted that small residues were favoured as well as lysine (Lys), glutamine (Gln) and asparagine (Asn), while hydrophobic residues tend to be avoided (Argos, 1990). Twelve years on, George and Heringa (2002) performed a similar study although this time on a database containing 1280 proteins. Proline (Pro), arginine (Arg), phenylalanine (Phe), threonine (Thr), glutamic acid (Glu) and glutamine (Gln) were identified as common to linker regions. In both studies, it was concluded that smaller residues were preferred while hydrophobic residues were uncommon. Pro, Thr and Gln were identified as most favourable (Chen et al., 2013). The use of smaller amino acids may increase flexibility of the linker while polar-charged residues are thought to improve stability. Proline is a unique amino acid with a cyclic side chain which makes it highly rigid. In recombinant fusion proteins, it is thought that proline prevents interaction of the linker region with the fused proteins and increases the stiffness of the linker (Chen et al., 2013). Naturally, proline-rich inter-domain linker regions are common, such as within the hinge region of IgG3 and IgA1 (Chen et al., 2013; Woof and Russell, 2011).

Studies of linker regions in natural multi-domain proteins has influenced the design of linkers in recombinant fusion proteins. A sophisticated linker design has the potential to improve the expression and stability of a recombinant fusion protein (Chen et al., 2013). A linker region may have a flexible design if a degree of movement between the protein

domains is required. Alternatively, in situations where a flexible hinger has resulted in poor expression or loss of functionality, a rigid, helical structure can be used to increase separation of the protein domains. Linker regions may also be cleavable and have the ability to release domains *in vivo* (Chen et al., 2013).

In addition to the careful consideration of linker composition, the linker length must be optimised. A linker may be extended to increase separation between the linked proteins in order that they each retain their functionality. Silacci et al. (2014) investigated the length of the linker region in the Fynomer-Fc fusion protein by producing four mutants whose hinges varied in length. It was concluded that the fusion protein with the longest linker region displayed the most potent inhibitory activity (Silacci et al., 2014). However, it should be noted that an extended hinge region may increase susceptibility to protease cleavage (Chen et al., 2013).

Currently all Fc-fusions in clinic are of the IgG1 subclass and in chapter 3, a DBL IgG1-Fc fusion was produced. The flexibility of the hinge regions within the four IgG subclasses is variable with IgG3 the most flexible, followed by IgG1 (Roux et al., 1997). The IgG3 hinge is distinct from the other three subclasses and is 62 amino acids in length compared to the other subclasses which have much shorter hinges ranging 12 - 15 amino acids (Table 5.1). Each IgG hinge consists of distinct upper, middle and lower regions. The upper region connects the Fab arms to the middle hinge and is responsible for a degree of flexibility (Roux et al., 1997). The middle hinge region is a relatively inflexible proline-rich region containing the interchain cysteine residues. The lower hinge connects the Fc to the middle hinge region and is responsible for further flexibility in the hinge (Roux et al., 1997). Although the IgG3 middle hinge region is significantly longer than the IgG1 middle hinge, the extra flexibility of the IgG3 hinge is more likely due to the extra residues in the upper hinge.

Table 5.1: Amino acid sequences of hinge regions from the four IgG subclasses.

IgG subclass	UH	MH	LH
IgG1	EPKSCDKTHT	CPPCP	APELLGGP
IgG2	ERK	CCVECPPCP	APPVAGP
IgG3	ELKTPLGDTTGT	CPRCP(EPKSCDTPPPCPRCP)$_3$	APELLGGP
IgG4	ESKYGPP	CPSCP	APEFLGGP

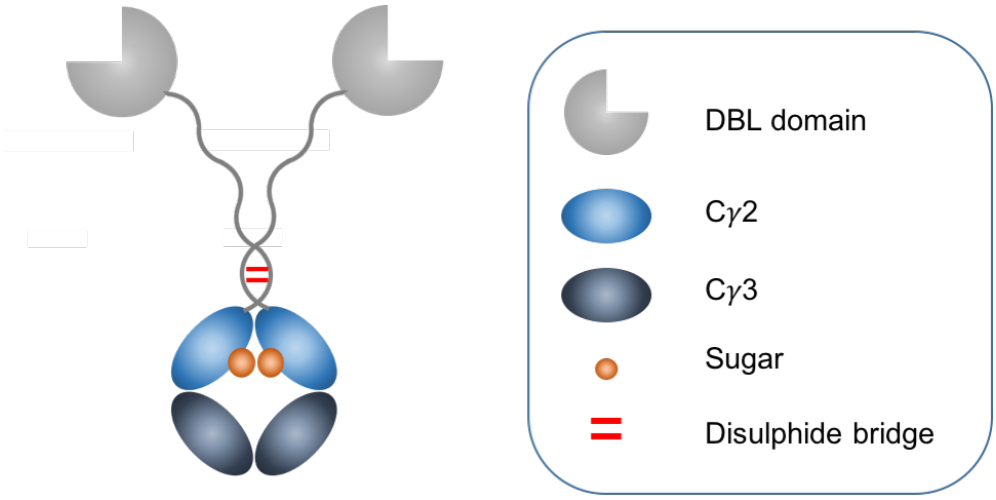


Figure 5.2: Schematic representation of a dimeric Fc-fusion protein with extended hinge region. The IgG1 hinge regions has been repeated four times with the three additional hinges closest to the DBL domain mutated (C=A) to prevent formation of further disulphide bonds.

In this chapter DBL Fc-fusion constructs have been produced with an extended IgG1 hinge consisting of four successive IgG Fc hinge regions designed to extend the distance between the DBL and Fc domains. The length of the extended IgG1 hinge was increased to resemble that of the more flexible IgG3 hinge. The original wild type IgG1 hinge is located next to the Cγ2 region of the Fc and three additional IgG1 hinges have been added before the DBL domain (Figure 5.2). The additional three hinges have been mutated by replacing the two cysteine residues with alanine residues. This is to prevent disulphide bond formation within the extended hinge region and to provide an extra degree of flexibility. The original hinge region retains cysteine residues responsible for disulphide bridge formation which aid in polymerisation of the construct.

5.1.2 Location of the DBL domain

When this work was started, the crystal structure of MSPDBL2 had not yet been resolved and the exact location of the DBL domain within the full-length MSP was unclear. The DBL domains expressed as Fc-fusion proteins in Chapter 3 are shorter than the DBL domain that has since been crystallised. There are six amino acids missing at the N terminus including a cysteine residue, and seven residues missing at the C terminus. In this chapter, the full-length DBL domains have been expressed in order to investigate the effect of the missing residues on the functionality of the DBL domains.

5.2 Objectives

The aim of this chapter is to produce improved DBL Fc-fusion proteins. The specific objectives include:

- Creating modified DBL Fc-fusion constructs where the hinge region has been extended, increasing the distance between the DBL domain and the Fc region.
- Expressing a modified, full length DBL domain which includes the extra amino acids present in the crystal structure of MSPDBL2.
- Characterising the new constructs and identifying any functional improvements.

5.3 Methods

5.3.1 pFMCS-hIgG1-Fc2 (modified pFUSE-hIgG1-Fc2 plasmid)

A modified pFUSE-hIgG1-Fc2 plasmid was used for cloning. This contains a modified multiple cloning site for ease of sub-cloning but the rest of the plasmid remains the same. The multiple cloning site of pFUSE-hIgG2-Fc2 contains extra base residues between restriction sites, therefore sub-cloning the same insert between different restriction sites requires additional bases to keep the insert in frame. pFMCS-hIgG2-Fc2 was developed such that an insert added between any restriction sites remain in frame. Additional restriction sites BamHI and XhoI were also added for convenience (Figure 5.3).

pFUSE-hIgG1-Fc2

EcoRI EcoRV NcoI BglII

ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCAC GAATTC GATATC GG CCATGG TT AGATCT GACAAAACTCACACATGCCCACCGTGCCCA

pFMCS-hIgG1-Fc2

EcoRI EcoRV NcoI BamHI XhoI BglII

ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACT GAATTC GATATC CCATGG GGATCC CTCGAG AGATCT GACAAAACTCACACATGCCCACCGTGCCCA

pFMCS (parking)

EcoRI EcoRV NcoI BamHI XhoI BglII NheI

ATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACT GAATTC GATATC CCATGG GGATCC CTCGAG AGATCT GCTAGC TGGCCAGACATGATAAGATACATTGATGA

Figure 5.3: Multiple cloning site of original pFUSE-hIgG1-Fc2 plasmid vs modified pFMCS-hIgG1-Fc2. In the modified plasmid, additional restriction sites BamHI and XhoI have been added. The bases between restriction sites EcoRV/NcoI and NcoI/BglII have been removed for convenience. The parking plasmid pFMCS contains the modified multiple cloning site from pFMCS-hIgG1-Fc2 but has the hIgG1-Fc2 removed. This is used as a parking plasmid for the DBL inserts.

A parking plasmid pFMCS was developed for sub-cloning for the DBL inserts. This contains the modified pFMCS-hIGg1-Fc2 multiple cloning site and has the IgG1-Fc2 portion removed.

5.3.2 pFMCS-4HF-hIgG1-Fc2 (extended flexible hinge construct)

Another modified pFUSE-hIgG1-Fc2 plasmid was used for cloning the flexible four-hinge constructs. This is the same as the pFMCS-hIgG1-Fc2 (with modified multiple cloning site) but contains an additional extended hinge region. The extended hinge region (three hinges) with cysteine residues mutated to alanine was synthesised separately with compatible restriction sites and sub-cloned into pFMCS-hIgG1-Fc2 to produce the pFMCS-4HF-hIgG1-Fc2 construct (Figure 5.4).

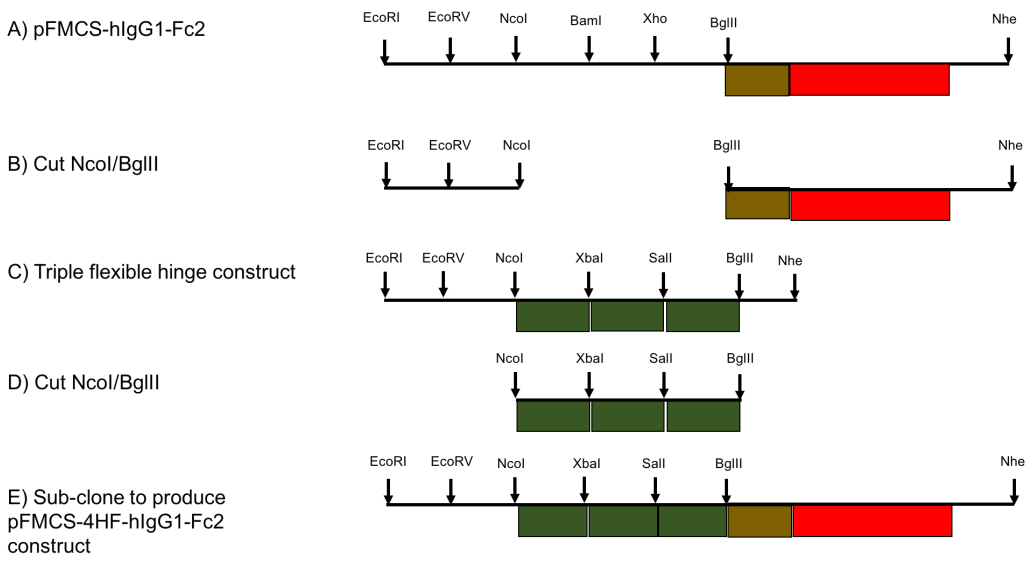


Figure 5.4: **Sub-cloning of the pFMCS-4HF-hIgG1-Fc2 construct.** The flexible four-hinge construct was produced by the addition of a triple hinge insert containing three modified flexible IgG1 hinge regions to the pFMCS-hIgG1-Fc2 plasmid. The pFMCS-hIgG1-Fc2 construct (A) was cut by restriction digest with NcoI and BglII (B). The triple hinge insert was synthesised with the required restriction sites (C) and cut with NcoI and BglII (D) before being sub-cloned into the pFMCS-hIgG1-Fc2 plasmid (E).

5.3.3 Plasmid Synthesis

A series of plasmids was synthesised by Eurofins (UK):

1. **pFMCS-3.4 FL:** pFMCS parking plasmid containing the full length 3.4 DBL (Appendix A.8).
2. **pFMCS-3.8 FL:** pFMCS parking plasmid containing the full length 3.8 DBL (Appendix A.9).
3. **pFMCS-3.4 short:** pFMCS parking plasmid containing the short 3.4 DBL domain (Appendix A.10).
4. **pFMCS-3.8 short:** pFMCS parking plasmid containing the short 3.8 DBL domain (Appendix A.11).
5. **pFMCS-hIgG1-Fc2-3.4 FL:** pFMCS-hIgG1-Fc2 containing the full length 3.4 DBL (Appendix A.13).
6. **pFMCS-hIgG1-Fc2-3.8 FL:** pFMCS-hIgG1-Fc2 containing the full length 3.8 DBL (Appendix A.14).

Short 3.4 and 3.8 DBL pFMCS-hIgG1-Fc2 constructs (5 and 6) were synthesised ready for transfection and direct comparison of recombinant protein produced with the short recombinant DBL-Fc fusion proteins from Chapter 3.

The four pFMCS parking plasmids (1 - 4) contained DBL inserts ready for sub-cloning into the pFMCS-4HF-hIgG1-Fc2 construct.

5.3.4 Sub-cloning the four DBL domains (short and full-length, 3.4 and 3.8) into the pFMCS-4HF-hIgG1-Fc2 construct

Plasmid preparation

The PfMCS-4HF-hIgG1-Fc2 plasmid was prepared for insertion of the DBL domains using the same methods as Chapter 3. Plasmid was digested using the appropriate restriction enzymes (EcoRv/NcoI for sub-cloning 3.4 DBL domains and EcoRI/NcoI for 3.8 DBL domains), run by gel electrophoresis, excised and purified. The plasmid was then de-phosphorylated to prevent self-ligation (Section 3.3.5). Recovered DNA was quantified by nanodrop spectro-photometer (Section 2.2.15).

Insert preparation

The pFMCS parking plasmids were similarly prepared for sub-cloning by cutting out the DBL domain using the appropriate restriction enzymes (3.4 DBLs: EcoRv/NcoI and 3.8 DBLs: EcoRI/NcoI). The product was run by gel electrophoresis and DBL inserts excised and purified (Section 3.3.5).

Ligation

The following reaction with a molar ratio of 1:3 vector:insert was set-up on ice for ligation reactions containing the short DBLs:

- 10x T4 DNA ligase buffer 2 µl
- Vector DNA (4.3 kb) 50 ng
- Insert DNA (860 bp) 30 ng
- T4 DNA ligase 1 µl
- dH_2O up to 20 µl.

The following reaction keeping a molar ration of 1:3 vector:insert was used for ligation reactions containing the full length DBLs:

- 10x T4 DNA ligase buffer 2 µl
- Vector DNA (4.3 kb) 50 ng
- Insert DNA (905 bp) 32 ng
- T4 DNA ligase 1 µl
- dH_2O up to 20 µl.

The ligation reactions were held at 16° C for 16 hours using a Progene thermocycler.

Transfection

Cells were transformed using the same methods as in Chapter 3. Single colonies were grown and DNA extracted by mini-prep which was then screened to check that it contained the insert by restriction digest (Section 3.3.7). Plasmids containing inserts were identified and DNA sent for sequencing for confirmation (Section 2.2.13).

5.3.5 Expression of recombinant protein

Recombinant protein was produced in CHO-KI cels using the same methods as in Chapters 3 and 4 (Section 3.3.8). CHO-KI cells were transfected using FuGENE 6 transfection reagent and selected using zeocin selection media. Monoclonal populations were produced by selecting for high-expressing colonies formed from single cells using immunoblot analysis. After 10 days in culture, cell supernatant was purified using a protein-G sepharose HiTrap column on the ÄKTA-FPLC. Immunoblot analysis was used to test for the presence of recombinant protein in FPLC fractions which were then pooled and concentrated by ultrafiltration.

5.3.6 IgM-binding ELISA

The series of constructs were screened for IgM binding by ELISA. The same methods were used as in Chapters 3 and 4 (Section 3.3.10).

5.3.7 Western blot analysis

The series of constructs were characterised by Western blot analysis. The same methods were used as in Chapters 3 and 4 (Section 4.3.8).

5.3.8 Gel filtration of recombinant proteins

The series of constructs were characterised by gel filtration using the same methods as in Chapters 3 and 4 (Section 3.3.12).

5.4 Results

5.4.1 Sub-cloning the four pFMCS-4HF-hIgG1-Fc2-DBL constructs

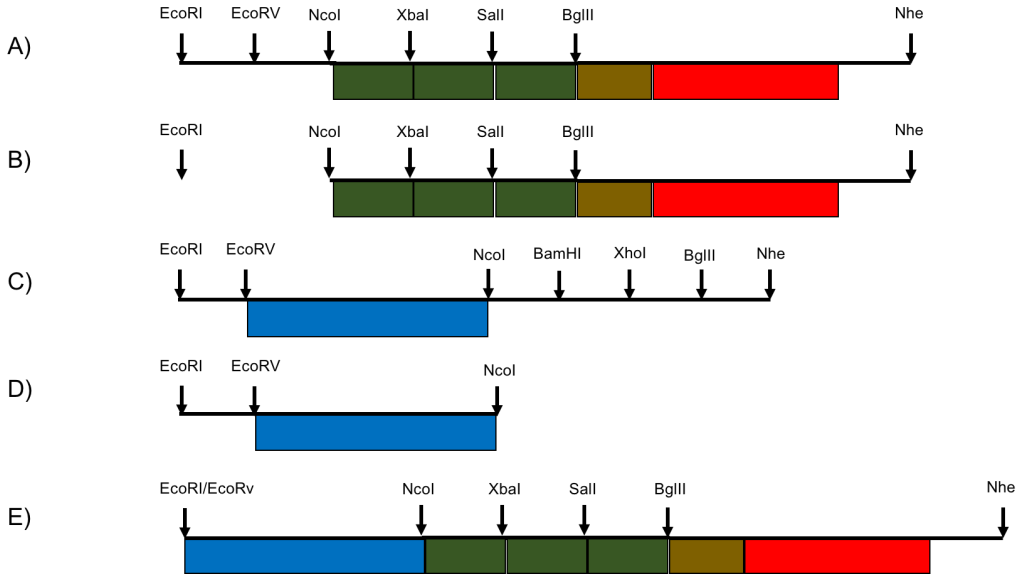
Four constructs were produced by sub-cloning DBL inserts into the pFMCS-4HF-hIgG1-Fc2 plasmid (Figure 5.5).

- pFMCS-4HF-hIgG1-Fc2-3.4DBL-short
- pFMCS-4HF-hIgG1-Fc2-3.4DBL-FL
- pFMCS-4HF-hIgG1-Fc2-3.8DBL-short
- pFMCS-4HF-hIgG1-Fc2-3.8DBL-FL

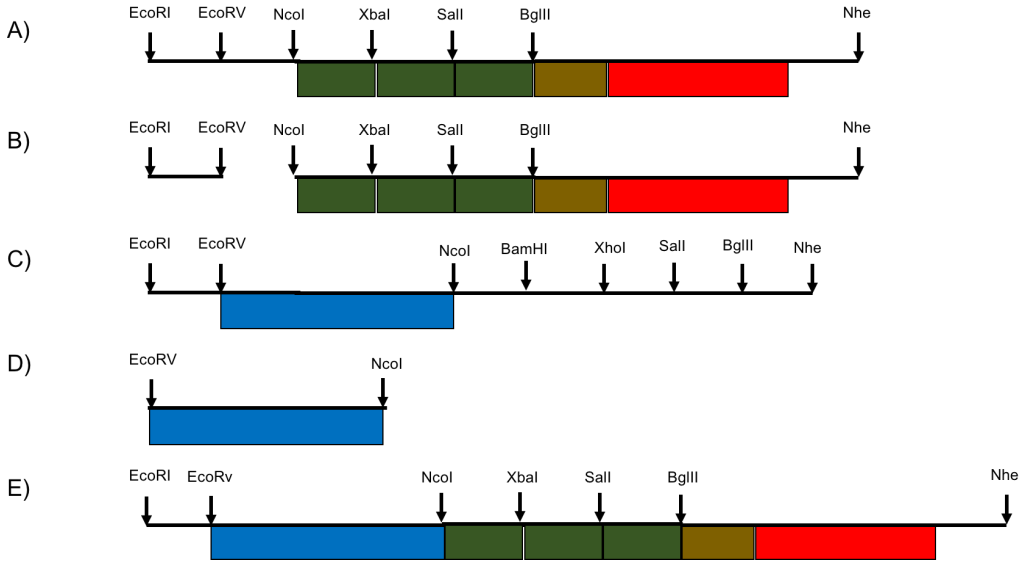
The pFMCS-4HF-hIgG1-Fc2 plasmid and inserts were prepared by restriction digest with the required restriction enzymes and then ligated to create the new constructs. These were transformed into *E. coli* cells and colonies screened for presence of inserts by restriction digest. DNA extracted by mini-prep was sent for sequencing to confirm each construct before DNA was used to transform CHO-K1 cells for protein expression.

5.4.2 Producing full-length DBL domains in pFMCS-IgG1-Fc2

The original pFUSE 3.4 DBL and pFUSE 3.8 DBL constructs were re-transfected and new monoclonal cell lines produced in parallel with the new full length constructs. The full length constructs were synthesised directly into the pFMCS-IgG1-Fc2 plasmid. Having been extracted by maxi-prep, the DNA was confirmed by sequencing and used to transform CHO-K1 cells for protein expression.



(a) Sub-cloning 3.4 DBL inserts into pFMCS-4HF-IgG1-Fc2



(b) Sub-cloning 3.8 DBL inserts into pFMCS-4HF-IgG1-Fc2

Figure 5.5: **Sub-cloning the (a) 3.4 and (b) 3.8 DBL inserts into the pFMCS-4HF-IgG1-Fc2 construct.** Sub-cloning the 3.4 DBL inserts (both short and full length) into the pFMCS-4HF-IgG1-Fc2 (A) involved digesting the plasmid using EcoRV and NcoI (B). The 3.4 DBL inserts (both short and full length) were both synthesised into the parking pFMCS plasmid (C) which was prepared by cutting with EcoRV and NcoI (D). The insert was then ligated into the cut plasmid (E). Similarly, the 3.8 DBL inserts (both short and full length) were sub-cloned into pFMCS-4HF-IgG1-Fc2 (A) by digesting the plasmid using EcoRI and NcoI (B). The inserts, synthesised into the parking pFMCS plasmid (C) were prepared by cutting with EcoRI and NcoI (D). The insert was then ligated into the cut plasmid (E).

5.4.3 Expression of the DBL-Fc fusion recombinant proteins

CHO-K1 cells were transfected with the eight constructs (summarised in Section 5.4.3.1), using the FuGENE transfection reagent and the same methods as Chapters 3 and 4 (Section 3.3.8), and monoclonal populations of high-expressing cells were produced (Section 3.3.8).

5.4.3.1 Summary of DBL-Fc fusion proteins produced

The eight recombinant DBL-Fc fusion proteins produced are summarised in Figure 5.6. The original two DBL-Fc fusion proteins produced in Chapter 3 will be referred to as DBL3.4 IgG1 Fc and DBL3.8 IgG1 Fc. The full-length versions of these, which were synthesised directly into pFMCS-hIgG1-Fc2 before expression are referred to as DBL 3.4 FL IgG1 Fc and DBL 3.8 FL IgG1 Fc.

The four extended-hinge constructs were produced by sub-cloning the inserts into the pFMCS-4HF-IgG1-Fc2 plasmid. This produced two short DBL Fc fusion recombinant proteins with the extended hinge which will be referred to as DBL 3.4 4HF IgG1 Fc and DBL 3.8 4HF IgG1 Fc. The full-length versions of these are referred to as DBL 3.4 FL 4HF IgG1 Fc and DBL 3.8 FL 4HF IgG1 Fc.

Figure 5.6: Schematic representation of the eight DBL-Fc fusion constructs. (1) The short 3.4 DBL and (2) short 3.8 DBL were cloned into the pFUSE-hIgG1-Fc vector containing the IgG1 hinge region and Fc (Chapter 3). (3) The full length 3.4 DBL and (4) full length 3.8 DBL have been cloned into the pFMCS-hIgG1-Fc vector as for (1 - 2). This plasmid is the same as the pFUSE-hIgG1-Fc but with a modified multiple cloning site (MCS). (5) The short 3.4 DBL and (6) short 3.8 DBL have been cloned into a modified pFUSE-hIgG1-Fc vector containing the modified IgG1 hinge region. In this construct, the IgG1 hinge has been repeated 4 times, in 3 of which cysteine residues have been replaced by alanine to increase flexibility. (7) The full length 3.4 DBL and (8) full length 3.8 DBL have similarly been cloned into the modified pFUSE-hIgG1-Fc vector containing the modified IgG1 hinge region.

5.4.3.2 FPLC Protein purification

The eight recombinant DBL-Fc fusion proteins were purified from supernatant using a protein G sepharose column by affinity chromatography (Section 3.3.8). A linear gradient of 0.1 M glycine-HCl, pH 2.7 was used to elute protein from the column. FPLC profiles are summarised in Table 5.2 and can be found in Appendix J.

Table 5.2: Summary of FPLC profiles from eight DBL-Fc fusion constructs.

Construct	Profile (mAU)
Original 3.4 DBL short	700
Original 3.8 DBL short	400
Full-length 3.4 DBL	600
Full-length 3.8 DBL	500
3.4 DBL short with extended hinge	700
3.8 DBL short with extended hinge	400
Full-length 3.4 DBL with extended hinge	1400
Full-length 3.8 DBL with extended hinge	800

The FPLC purification profiles from recombinant proteins with the original IgG1 hinge were compared (Appendix J, Figure .21). There were four Fc-fusion proteins produced with the original hinge, two 3.4 DBL domains (one short and one full-length) and similarly two 3.8 DBL domains. The full-length DBL fusion proteins had similar profiles to the short DBL fusion proteins, suggesting that the extra residues included at the beginning and end of the DBL domains did not improve protein expression.

FPLC purification profiles from the original DBL-Fc fusions were compared to recombinant protein containing the original DBL domains along with the extended hinge region (Appendix J, Figure .22). The profiles of the fusion proteins with extended hinges were similar to those containing the original hinge, suggesting that the extended region does not improve protein expression.

FPLC purification profiles from the full-length DBL-Fc fusion proteins containing the original IgG1 hinge were compared to the full-length DBL-Fc fusion proteins containing the extended hinge region (Appendix J, Figure .23). The purification profile for the full-length 3.4 DBL domain is much bigger with an extended hinge compared to the original IgG1 hinge for the same volume of cell supernatant. However, the 3.8 FL DBL hinge doesn't appear to improve protein purification. This might suggest that the cells express the

extended hinge, full-length 3.4 DBL-Fc fusion better than the same full-length 3.4 DBL domain with the original hinge. However, protein expression varies between cells therefore a monoclonal population may be derived from a cell which is a better general expresser of protein.

Fractions were collected and immunoblotted with αhIgG-AP (Section 3.3.8) to show that protein was present in the fractions corresponding to the FPLC (Figure 5.7).

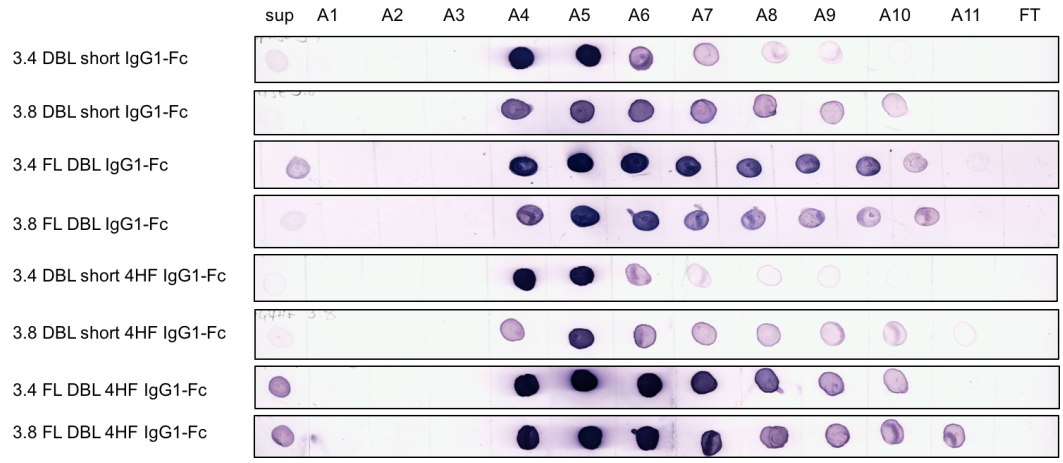


Figure 5.7: **Immunoblot of fraction from the FPLC elution of the eight DBL-Fc fusion recombinant proteins.** Fractions A1-A11 from the FPLC elution were tested by immunoblotting for the presence of protein. These were compared to the cell supernatant (sup) prior to purification as well as the flow-through (FT) to check that all protein had been removed.

5.4.4 Characterisation of recombinant protein

In Chapter 3, recombinant DBL-Fc fusion proteins had difficulty forming homodimers. In order to characterise the new series of DBL-Fc fusion proteins, non-reduced samples were separated by SDS-PAGE and transferred to a membrane. IgG-Fc was detected using αhIgG-Fc-AP. The addition of the extended hinge region appears to improve dimerisation of the DBL-Fc fusion proteins (Figure 5.8). However, monomeric DBL-Fc is still present and the addition of the hinge region also leads to aberrant folding intermediates. This suggests that the extension of the hinge may increase susceptibility of the recombinant protein to cleavage by proteases.

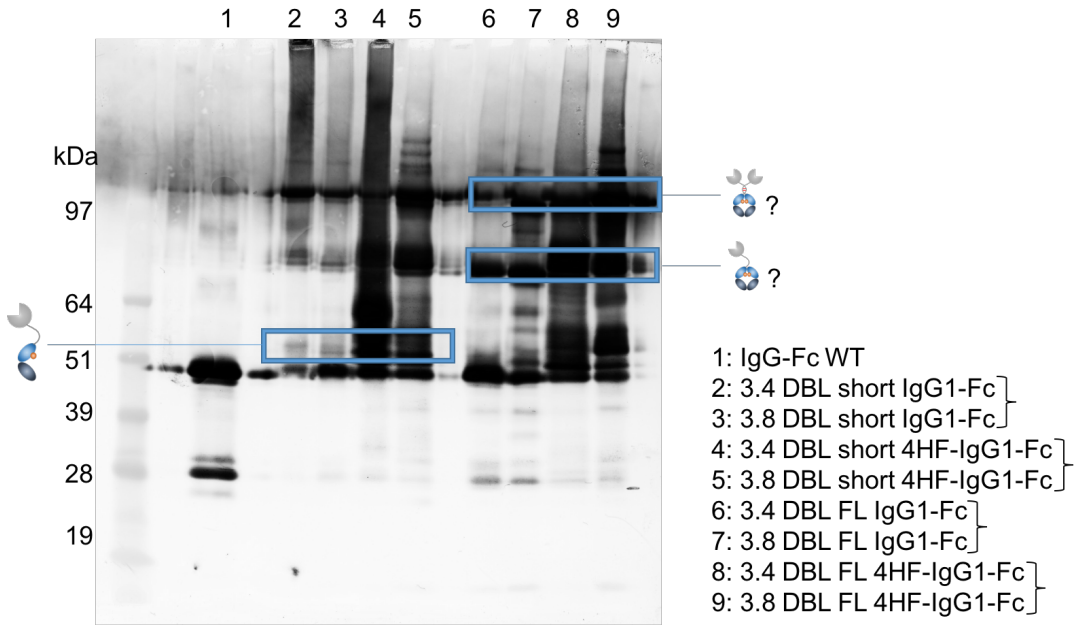
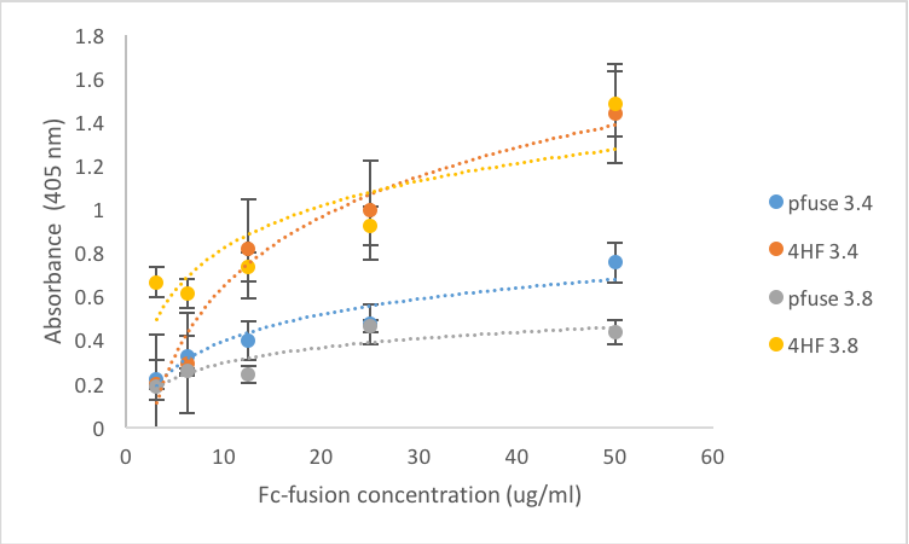


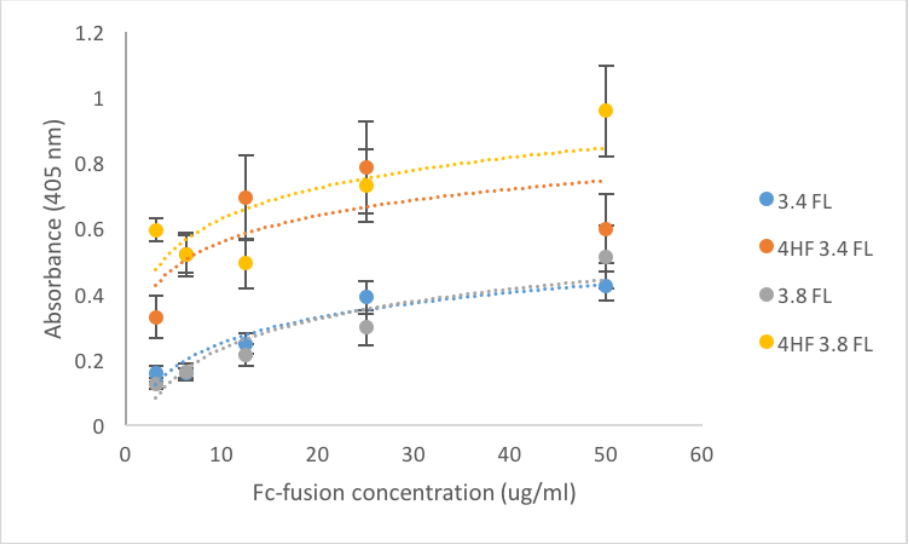
Figure 5.8: **Non-reduced αhIgG1-Fc Western blot shows improved folding of recombinant DBL-Fc fusions with the addition of the extended hinge.** SDS-PAGE of recombinant DBL-Fc fusion proteins was transferred to a membrane and Fc detected using αhIgG-Fc alkaline phosphatase-conjugated antibody. An IgG-Fc WT control was used in **Lane 1**. **Lanes 2 and 3** show the original 3.4 DBL-Fc and 3.8 DBL-Fc fusion proteins respectively. **Lanes 4 and 5** show these short DBL-Fc fusion with the extended hinge. **Lanes 6 and 7** show the full-length version of the 3.4 DBL-Fc and 3.8 DBL-Fc fusion proteins respectively. **Lanes 8 and 9** show these full-length DBL-Fc fusions with the extended hinge.

5.4.5 Recombinant DBL-Fc fusion proteins containing the extended hinge show improved hIgM binding

In Chapter 3 it was shown by sandwich ELISA that the recombinant DBL-Fc fusion proteins bind human IgM. DBL 3.4 was shown to bind hIgM better than DBL 3.8, in agreement with a study by Crosnier et al. (2016) where DBL 3.4 was shown to have ten-fold higher affinity than the 3.8 DBL domain. The sandwich ELISA was repeated for the eight DBL-Fc fusion recombinant proteins in order to identify differences in IgM binding amongst the constructs. Human IgM was captured by coating of an αhIgM on the ELISA plate. The DBL-Fc fusions were then allowed to bind hIgM and were detected using αhIgG-Fc-AP. It was shown that the recombinant DBL-Fc fusions containing the extended hinge region bind hIgM better than those containing the original IgG1 hinge (Figure 5.9). This was true for both sets of recombinant proteins, those containing the original short DBL domains (Figure 5.9a) as well as those containing the full-length DBL domains (Figure 5.9b). There was, however, no notable improvement in binding in the full-length DBL constructs over the original short DBL domains.



(a) IgM-binding of recombinant short DBL-Fc fusion proteins



(b) IgM-binding of recombinant full-length DBL-Fc fusion proteins

Figure 5.9: **IgM-binding sandwich ELISA showing improved binding by the new flexible 4H constructs.** (a) The short DBL-Fc fusion constructs show improved IgM binding with the addition of the flexible 4-hinge (4HF). (b) The full length DBL-Fc fusion constructs show improved IgM-binding with the addition of the flexible 4-hinge (4HF). Results are shown from four repeat experiments.

5.5 Summary

The aim of Chapter 3 was to produce novel recombinant DBL-Fc fusion proteins from the DBL domains unique to two interesting MSP3 proteins, MSPDBL1 and MSPDBL2. DBL inserts were sub-cloned into the pFUSE-hIgG1-Fc2 plasmid (which contains the IgG1 Fc and hinge regions) for protein expression in CHO-K1 cells. The subsequent recombinant protein produced, although functional with respect to protein G and hIgM binding, was mostly monomeric, suggesting that the Fc-fusion protein had difficulty polymerising. The aim of this chapter was to improve the DBL-Fc fusion constructs in order to produce homodimers and to investigate whether there was improved hIgM-binding to the new recombinant protein.

There were two possible explanations for the lack of dimer formation by the original recombinant DBL-Fc fusions. Firstly, the DBL inserts were identified as shorter in length compared to the DBL domains whose crystal structure was resolved by Hodder et al. (2012). New DBL-Fc fusions were therefore produced using the full-length DBL domains. This involved the addition of six residues prior to the DBL sequence and seven at the end. There were no functional improvements observed by adding the extra residues and protein expression, which had previously proved difficult to produce adequate yields, was comparable. Secondly, the relative size and compact structure of the DBL domains compared to the Fc domain was proposed to prevent polymerisation of the fusion proteins.

A modified Fc-fusion construct with an extended hinge region was designed to increase the distance between the Fc and the DBL domain. The hinge region was designed to give an extra degree of flexibility to the molecule and it was hypothesised that this would enable formation of a dimeric Fc-fusion protein. The hinge region was constructed from four repeated IgG1 hinge regions. The original hinge remains in proximity to the Fc and three successive modified IgG1 hinges, in which cysteine residues are replaced with alanine, add an extra element of flexibility. The two cysteine residues remain in the original Fc-proximal hinge region; these are essential for the formation of disulphide bonds needed to produce homodimers.

Both of the original short DBL domains and the full-length DBL domains were sub-cloned into the modified construct containing the extended flexible hinge. An improved FPLC profile was seen from purification of the full-length 3.4 DBL-Fc fusion containing the extended flexible hinge compared to the same full-length 3.4 DBL-Fc fusion

with the original hinge. This suggests either improved protein-G binding or improved protein expression. Any difference in expression could be due to differences in the protein expression capacity of different monoclonal cell populations. It is therefore unclear whether protein production was improved by the extended hinge region, or by better expressing cells. There was no difference in expression seen from purification of the two full-length recombinant 3.8 DBL domains with either the extended hinge or the original hinge. In this case, the extended hinge construct did not improve protein G binding or protein expression.

In all cases, protein yield was still low, improvements are therefore still needed in order to produce adequate quantities of protein. However, the recombinant protein was able to form homodimers, although there is evidence that proteolytic cleavage is occurring. Other modifications to the hinge region may be necessary to further improve protein folding.

There was an improvement in hIgM-binding seen in the recombinant proteins which contained the modified hinge. Both 3.4 DBL and 3.8 DBL domains showed improved hIgM-binding ability when expressed in an Fc-fusion with the modified hinge. For the full-length 3.4 DBL and full-length 3.8 DBL, the improvement in binding from the modified hinge was even more striking. The ability to improve functionality suggests that the modified, extended hinge construct might have useful applications in therapeutics.

Chapter 6

Discussion

6.1 Overview

Malaria is a vast global problem with approximately half of the world's population at risk, particularly in vulnerable communities where the burden is greatest. There were 198 million clinical cases of malaria in 2013, with 584,000 deaths (WHO, 2015). Malaria vaccine development is urgently needed and the blood stages are pivotal, because it is these stages of infection that causes clinical symptoms. Merozoite surface proteins MSPDBL1 and MSPDBL2 have recently received recognition as vaccine candidates due to their role in clinical immunity to malaria (Chiu et al., 2015).

The aim of this study was to identify residues and sequence motifs critical for IgM-binding of the DBL domains from MSPDBL1 and MSPDBL2 *P. falciparum* merozoite surface proteins. The vaccine potential of MSPDBL1 and MSPDBL2 is only just being recognised and these proteins, which have been shown to localise to the merozoite surface, are thought to be involved in erythrocyte invasion (Lin et al., 2014). However, their receptors on the erythrocyte surface have not yet been identified. In a recent study by Crosnier et al. (2016), the MSP DBL domains have also been shown to bind the Fc portion of human IgM. Although MSPDBL1 and MSPDBL2 are the only two merozoite surface proteins to contain DBL domains, such domains are common in other surface-protein families such as PfEMP1. In PfEMP1 they are responsible for the cytoadherent properties of infected erythrocytes, some of which also bind IgM, and are associated with virulence. To further investigate these interesting binding characteristics of MSPDBL1 and MSPDBL2, this study aimed to:

1. Express the DBL domains from MSPDBL1 and MSPDBL2 as recombinant Fc-fusion proteins.

2. Produce a library of recombinant DBL mutants to investigate IgM binding.

3. Analyse a wealth of sequence data and structural information from all known IgM-binding DBL domains to provide further insight into the IgM-binding region.

Fc-fusion proteins are an exciting and successful field of bio-engineering due to their interaction with the FcRn which facilitates recycling and improves half-life (Levin et al., 2015; Czajkowsky et al., 2012). Having experienced impressive growth rates compared to other biologics in the past few years, Fc fusion proteins are now well established

in the clinic (Aggarwal, 2014). The most successful Fc fusion to date is etanercept, a TNF receptor-fusion that is able to block the action of the tumour necrosis factor (TNF) (Czajkowsky et al., 2012; Ioannidis et al., 2013). It is used in rheumatoid arthritis and other inflammatory conditions to successfully reduce inflammation.

A crucial first stage of this study was to produce recombinant DBL domains from MSPDBL1 and MPSPDBL2. These were expressed as Fc-fusion proteins which, while enabling further study of these DBL domains, could have the potential to be developed for vaccine or therapeutic use. Malaria vaccine development is urgently required as current vaccines have limited efficacy or are limited by logistical issues. The DBL-Fc fusions could also be developed for the treatment of IgM-related diseases due to their affinity for IgM.

DBL Fc-fusion constructs were produced by fusing DBL domains to the pFuse-hIgG1-Fc2 mammalian expression vector. This plasmid is commercially available and has previously been used in our laboratory to produce successful Fc-fusion proteins with the malaria antigens $MSP1_{19}$ and $MSP1_{42}$ (Mekhaiel et al., 2011). These merozoite DNAs were codon optimised for mammalian expression, and N-linked glycan sites removed from the DBL domain in order to prevent unnecessary glycosylation of the recombinant proteins. Glycosylation, however, has been shown to be necessary for the Fc glycoprotein domain. Mammalian expression allowed for the formation of disulphide bonds within the DBL domains which are known to be critical for structural stability. Recombinant protein was produced from CHO-K1 cells, as has previously been used for the successful expression of Fc-fusion proteins (Mekhaiel et al., 2011). CHO-K1 cells are the mostly commonly used expression system for commercial Fc-fusion proteins (Czajkowsky et al., 2012).

There were several difficulties with the construction of the DBL Fc constructs and working with the DNA proved problematic. Bacterial growth was challenging, particularly for the 3.8 DBL domain, which suggests that the DNA may be toxic to *E. coli*. Bacteria were found to grow better in a rich TB growth media rather than LB medium. However, bacterial colonies that did manage to grow mostly contained the correct insert. The use of the TOPO parking plasmid for efficient amplification of a large amount of the TOPO vector plus the correct insert was found to be an essential first step.

Protein expression also proved difficult, with low yields of monomeric fusion protein produced rather than the predicted 120 kDa homodimer. Most Fc fusions currently with FDA approval contain small ($\sim$25 kDa) antigens fused to the Fc. However, two Fc fusion

proteins (eloctate and rilonacept) that have had success in the clinic consist of much larger fused proteins. Eloctate, an antihaemophilic factor fusion, has a molecular weight of 220 kDa but is 39% single chain molecule (Peters et al., 2013; Buyue et al., 2014). Rilonacept, a 251 kDa IL-1R fusion, forms homodimers but the shape of IL-1R molecule isn't as tightly packed as a DBL domain and has been likened to a question mark in structure (Schreuder et al., 1997). The malarial Fc-fusion proteins produced by Mekhaiel et al. (2011) containing the fused $MSP1_{19}$ and $MSP1_{42}$ were able to form homodimers. Thus it may be that the size or shape of the merozoite DBL domains inhibits homodimer formation.

There was evidence of glycoform production, proteolytic cleavage and poorly folded protein. Gel filtration also suggested that aggregates were formed. However, monomeric Fc fusions were shown to be functional with relation to IgM binding, as the recombinant 3.4 DBL fusion bound better than the 3.8 DBL. This is in agreement with a study by Crosnier et al. (2016) which saw ten-fold improvement in 3.4 DBL-IgM binding over 3.8. DBL Fc fusions were therefore successfully expressed as functional recombinant proteins, albeit monomeric, which allowed further investigation into the IgM-binding properties of these merozoite protein domains.

In order to investigate the IgM-binding site, a library of DBL mutants was produced using random mutagenesis. Unlike the site-directed mutagenesis approach, this study did not rely on any previous knowledge or assumptions about the binding site. Although a molecular simulation of DBL-IgM binding provided by Dan Czajkowsky (Shanghai Institute of Applied Physics, China) predicts the regions involved in binding based on homology, the aim of this study was to either confirm the accuracy of this model or to provide data for model improvement.

Production of the mutant Fc-fusion constructs again proved difficult, particularly in the production of sufficient quantities of DNA for ligation. The Genemorph II mutagenesis kit (Agilent) was used to limit mutational bias. The PCR had to be scaled up, as after gel purification of the product, the DNA concentrations often proved too low for the ligation step. After ligation, the false positive rate for 3.4 DBL mutant colonies was high. Although lower for 3.8 DBL mutants, we obtained very few colonies after 3.8 DBL mutant ligation. The desired mutation frequency was eight within the DBL domain insert (nine mutations per kb). This corresponds to approximately four amino acid mutations within the DBL

insert, after silent mutations are taken into account. The manufacturer's instructions were followed in order to get this mutational frequency. However, the average mutation rate achieved was two amino acid substitutions per DBL domain. This is lower than predicted and is most likely due to inaccuracies in quantifying the concentration of the target DNA by nanodrop spectrophotometer.

After creation of the mutant plasmids, protein production proved difficult, with low yields and poor protein folding observed. Protein produced was predominantly monomeric, although there was evidence of slight improvements from previous attempts due to extensive method optimisation. The selection of high-expressing mammalian cell clones allowed for maximum protein production and optimised purification allowed effective harvesting of protein. However the yields still varied between mutants, suggesting that yield was highly dependent on the clonal cell line that was selected. Overall, even taking into account variation across the mutants and improved methods, yield was still low and inadequate for further functional investigation. This problem was addressed in Chapter 5.

Purification by protein G sepharose was effective after thorough optimisation of the methods giving recombinant protein that was functional with relation to protein G binding. However, fragments were seen by SDS-page and Western blotting, possibly due to poor protein folding or proteolytic cleavage. Due to limitations in resolution it was not possible to separate these fragments from the monomeric Fc-fusion by gel filtration, therefore the calculations of protein concentration may not accurately reflect the concentrations of intact monomers. This may have an effect on the ELISA data presented here as, although equal concentrations of each mutant protein preparation was used, this might not reflect equal concentrations of monomeric Fc-fusion. However, it was assumed that each mutant would experience the same proteolytic cleavage and difficulties in folding and thus comparable amounts of monomeric protein would be produced. The ELISA results were subsequently disappointing, with high variations in repeat experiments. IgM binding between mutants has been compared from ELISA results but any differences observed needs further investigation.

The ELISAs showed comparable IgM binding for most DBL mutants, and three mutants showed slight improvements in binding - 3.4 DBL mutant 87, and 3.8 DBL mutants 44 and 65. Due to the high variation observed in repeat experiments these results were not significant, however one of the mutants identified has an interesting

mutation which is worth highlighting. 3.4 Mutant 87 contains an asparagine mutation in helix 2a where it has been replaced with tyrosine. This region was also highlighted by structural analysis of isolates, where four potential binding regions were identified (Figure 4.20), one of which was also identified by protein-protein interaction prediction tools and corresponded to helix 2a. Improved binding might be surprising since the aromatic residue tyrosine is located less commonly on protein-protein interfaces (Ma et al., 2003). However, the conformation of tyrosine on the DBL surface is important. If improved binding is observed, this suggests that the aromatic ring is buried under the surface, leaving the hydroxyl group exposed. Thus tyrosine might act as a polar residue on the surface, replacing the polar asparagine residue which may be involved in binding. A resulting conformational change in the binding region may explain any improved binding. It has been suggested that binding is not only dependent on polar hotspots but also on flexibility of the binding site (Ma et al., 2003). Introducing tyrosine might improve binding by introducing a favourable change in flexibility.

A recent study by Semblat et al. (2015) using an equivalent DBL-IgM homology model from PfEMP1 tested the function of amino acids predicted to be involved in IgM binding. The results suggested that the model needed revision as none of the residues had an effect on binding. Data presented here are in agreement with this study. The predicted involvement of helix 2a disagrees with the molecular simulation of binding, although the region is highlighted by further structural analysis. Since the validity of the model is in question, further investigation into the involvement of helix 2a in IgM-binding is worth exploring.

Site-directed mutagenesis could be used to further investigate the role of helix 2a and to identify the critical residues involved in IgM-binding, although protein yield and folding difficulties need to be addressed. In Chapter 5, an extended IgG1 hinge region was developed with this aim. The extended hinge allowed an increase in distance between the Fc and the fused DBL domain and was designed with a degree of flexibility to allow proper folding of both regions. The extended hinge was constructed from repeating the IgG1 hinge region four times. The first region, which is proximal to the Fc, retains its cysteine residues but these were replaced with alanine in the additional three hinge regions. It was hypothesised that this would increase flexibility in the extended hinge as well as increasing the distance from the Fc, allowing improved protein folding and homodimer

formation.

Homodimer formation of the extended hinge DBL-Fc fusion proteins was observed. However, protein fragments were still present suggesting that the constructs still had some difficulty folding. Such fragments also suggest that there is proteolytic cleavage, which is to be expected as the increased length of a hinge is known to increase susceptibility to cleaving (Chen et al., 2013).

An extended DBL domain was also expressed with both the original and the extended flexible hinge, but showed no improvement in IgM binding. When this work was started, the DBL domain boundaries were unclear and therefore a short version of these DBL domains was used, compared to the recently solved crystal structure (Hodder et al., 2012). A minimal IgM-binding region has recently been identified in the DBL4ζ domain of TM284var1 (Semblat et al., 2015). The short domains expressed here fall within the equivalent minimal binding region, which might explain why there was no observed improvement in binding using the full length DBL domains. The data presented therefore support the minimal binding region of DBL4ζ and suggest that the minimal binding region of the merozoite DBL domains is equivalent to that of PfEMP1 DBLs.

6.2 Expression of merozoite DBL domains as Fc fusion proteins

Novel DBL Fc-fusion proteins have been produced, providing a framework for vaccine or drug development using the DBL domains from merozoite surface proteins which are known to be involved in erythrocyte invasion. These interesting and unique DBL domains have been shown to bind the Fc of hIgM via the Cμ4 domain, a characteristic of eight other DBL domains from PfEMP1 that are located on the erythrocyte surface (Crosnier et al., 2016; Pleass et al., 2015). Fc-fusion proteins have shown clinical success and the addition of the Fc from IgG1 allows improved half-life due to the interaction with the neonatal Fc receptor (FcRn). Expression of the merozoite DBL domains as recombinant Fc-fusion proteins allowed investigation into the region involved in IgM binding. The IgM-binding property is thought to allow the parasite to evade immunity by masking itself in hIgM, preventing recognition by specific and highly effective IgG Abs, but no critical residues or binding motifs have previously been identified (Czajkowsky et al., 2010; Pleass et al., 2015).

An Fc-fusion molecule typically consists of two IgG1-Fc chains attached to two molecules of the fused protein, bound together by disulphide bonds. Recombinant protein was initially expressed as monomeric DBL-Fc. The structure of the (∼35 kDa) DBL domains and their size is thought to prevent homodimer formation, leading to the presence of multiple fragments visible by Western blot. Protein yield was also low, typically producing yields of 0.2 - 5 μg/ml of culture supernatant, in comparison to a yield of 5 - 10 μg/ml for unfused pFUSE-hIgG1-Fc2. Protein production needs to be improved in order for the proteins to be further developed for clinical use. However, the monomeric Fc-fusions were functional with respect to protein G and IgM binding. The Fc-fusion drug Aloctate consists of a large proportion of monomeric protein, shown to be functionally comparable to the homodimer form of the molecule. Monomeric DBL-Fc fusions therefore allowed further analysis of the IgM binding region. Extensive method development gave some improvement to the protein produced. Expression and purification methods were optimised and limited amounts of homodimer were produced for some mutants, however yield was still low and protein folding not optimal.

Due to the fragmented nature of these recombinant proteins, there were difficulties obtaining accurate protein concentrations for functional investigations. FPLC peaks gave an estimate of protein yield which was calculated using the Beer-Lambert

equation. However, estimates were thought to be high by comparison to known protein concentrations on a protein gel (data not shown). A BioRad colorimetric assay was also used which is based on the Bradford method and calculates total protein quantity. Both of these estimates were thought to be high as all protein fragments and improperly folded proteins were quantified in the total amount. Functional analysis by ELISA was comparable but not quantitative for this reason. SPR is frequently used for quantitative binding analysis, but could not be used here due to the presence of protein fragments which might interfere with affinity. For the same reason, circular dichroism analysis was also not possible. This would have been useful to investigate the secondary structure of the DBL mutants and identify mutations which significantly change the structure of the mutants, as well as to confirm proper folding of the wild types.

The protein-folding problems were addressed by using an extended, flexible hinge based on the IgG1 hinge. The aim was to improve protein folding by increasing the distance between the Fc and the DBL and to allow a degree of flexibility. The current IgG1 hinge was repeated three times and in these repeated regions, cysteine residues were mutated to alanines to eliminate further disulphide bond formation and to give a higher degree of flexibility. Improvements to homodimer formation were observed but improper folding was still notable. Intermediate bands observed by Western blot may also be due to cleavage which has previously been reported in the IgG1 hinge region of a recombinant human IgG1 antibody expressed in CHO-KI cells (Yan et al., 2009). Extending the linker region may also make the hinge more susceptible to cleavage as has been reported in other fusion proteins Liu and May (2012); Zhang et al. (2009). Analysis of the four-hinge constructs was difficult as resolution was not good enough above 97 kDa on the Western blot. It is expected that the addition of the extended region adds a molecular weight of 10 kDa, but this was not noticeable at ranges of 120 kDa. However, definite improvements to folding were observed between protein containing the original hinge and the extended hinge although protein yields still remained an issue.

This study is an important first step as improvements in protein production have been observed, and the results highlight the potential that further development of the hinge region may have on protein folding. Ongoing work in the laboratory on extended hinge regions suggests that the flexible nature of the hinge may not be necessary and that improved folding is observed in a similar construct where the IgG1 hinge has been

repeated four times without mutating the cysteines. Therefore it may be the decreased proximity which improves protein folding in the fusion proteins. In this study, the first logical step was to develop the IgG1 hinge regions. An IgG3 hinge, at 62 amino acids in length, is longer than the hinge regions in the other IgG1 subclasses and has a high degree of flexibility. An IgG3-based hinge could be designed, or one based on the Cμ2 region of IgM may also be worth investigating. IgM is a natural oligomer with no hinge region, but it has an extra Cμ2 domain which could be used instead of the IgG1 hinge to decrease the proximity of the Fc and the DBL domain. Further work is necessary but there is hope that a sophisticated extended hinge region will provide optimal protein folding for the successful development of DBL-Fc fusions as vaccine candidates.

CHO-K1 cell expression

CHO-K1 cells are well established in the production of therapeutics, and are the most commonly used mammalian expression system for Fc-fusion proteins. Mammalian expression was important in this study for proper disulphide bond formation, which is critical for the stability of the DBL domain. Human HeK cells are used for the expression of the Fc fusion protein Alprolix and were also tested in this study. Although HeK cells are known to produce protein quickly, they were not successful for transfection in this case. CHO-K1 cells were therefore used for production of the DBL-Fc fusion. Yield problems were addressed by developing methods for selection of high-expressing clonal cell lines. This facilitated protein production, and it was also found that the CHO-KI cells grow better in Hams F-10 growth media instead of DMEM. However, a bacterial expression system could be developed to further improve yield. *E. coli* expression systems are cost effective and known to produce high protein yields. This is a commonly used system for the expression of *P. falciparum* recombinant protein where glycosylation is not necessary. However, care would be needed to ensure proper disulphide bond formation.

pFUSE-hIgG1-Fc2 plasmid

The pFUSE-hIgG1-Fc2 plasmid is a commercially available plasmid for the construction of Fc-fusion proteins and has shown to be successful (Mekhaiel et al., 2011). This plasmid is known to produce yields on the scale of micrograms per millilitre although typical yields here were 0.1 - 0.5 μg/ml. pFUSE-hIgG1-Fc2 contains the hEF1-HTLV promoter, a

strong promoter based on the Elongation Factor-1α core promoter 1. Protein yield is often dependent on promoter strength, however it has been suggested that weaker promoters are preferred in cases where proteins may be toxic.

Difficulties in bacterial growth may be related to the origin of replication on the pFUSE plasmid. However, it is unclear which origin of replication is used in the plasmid, therefore it is uncertain whether the plasmid is high or low copy number. Slow growth rates are usually linked to low copy number plasmids. However, if the DNA is toxic to the cells a low copy number plasmid is preferred. Therefore, it may be that the pFUSE plasmid is high copy number and may not ultimately be the best vector for this work.

Commercially available plasmids for the expression of Fc-fusion proteins are limited. Another plasmid used in the laboratory which produces a high yield of Fc fusion protein has a low copy number. This plasmid is patent protected and therefore the promoter is unknown. Further plasmid development is needed to optimise protein yield.

Protein purification

Protein was purified using the AKTA explorer FPLC system. This is a basic system that is no longer in production. The newer, highly sophisticated AKTA pure machine has a separate sample loop which prevents cell supernatant from being passed through the main pumps and the mixing chamber. This mixing chamber contains a magnetic flea which could be detrimental to protein and could explain some of the protein fragments observed. A new machine is currently being installed in the laboratory and it will be interesting to see whether improvements are seen in future protein purification.

Protein G sepharose columns were used to isolate the Fc-fusion proteins by affinity purification. Protein A is commonly used for the purification of Fc-fusion proteins, however the binding affinity of IgG is stronger to protein A than to protein G, and therefore a lower elution pH is required. It was decided that low pH should be avoided in order to prevent denaturing of the protein and to maintain the biological activity of the molecule, therefore protein G sepharose was used. However, purification by protein A could be further investigated.

6.3 Accuracy of the molecular simulation of DBL-IgM binding

A model for IgM binding was constructed by developing homology models of IgM-binding DBL domains based on known DBL structures and docking these to a model of IgM (Czajkowsky et al., 2010). There were multiple possible binding conformations but these were narrowed down by placing constraints on the termini of the IgM-binding domains due to the central disposition of the IgM-binding DBL domains within the PfEMP1 (Czajkowsky et al., 2010). The accessibility of the domains within the PfEMP1 and the localisation of ligand-binding residues in other DBL domains were also used to narrow down the binding conformations (Czajkowsky et al., 2010). The model shows a striking number of charged residues at immediate distances from oppositely charged residues on the IgM surface (Czajkowsky et al., 2010). The model corresponds particularly well to the DARC recognition site on *P. knowlesi* DBL (Singh et al., 2006).

A merozoite model of IgM binding was based on this PfEMP1 DBL-IgM docking model. The DBL structures are homologous and therefore the binding sites are assumed to be similar. However, due to the way that the potential conformations were narrowed down in the PfEMP1 model, based on constraints due to the PfEMP1 protein structure, it might be expected that the model is not accurate for the merozoite DBL domains. Protein interface prediction suggests that the binding sites may even be different for MSPDBL1 compared to MSPDBL2.

A recent study of the minimal binding region for the DBL4ζ from TM284var1 using the PfEMP1-IgM homology model provides further evidence that the model might not be accurate (Semblat et al., 2015). Residues predicted to be critical for IgM binding were mutated and shown not to affect binding and the authors have called for a revision of the model. Data presented here highlight the need for further model development.

Although the homology model was used to select for mutants, a set of random mutations was used rather than assuming the model to be correct and using it to produce mutants by site-directed mutagenesis. The model was used to select mutants for expression, but few mutations were actually in the regions predicted by the model. Charged residues were therefore selected as well as those that had cysteine mutations. Using improved recombinant proteins, future work might involve developing a larger library of mutants in order to further investigate the IgM binding region. However, constraints on available facilities may limit a large scale-study.

6.4 The role of helix 2a in IgM binding

Despite the disappointing results from ELISA investigations of the mutant DBL Fc-fusion proteins, regions of interest were identified following structural analysis of the IgM-binding DBL domains. MSPDBL isolates were modelled, based on the crystal structure of 3D7, and known IgM binders were superimposed on non-binders. This highlighted structural differences in helices 1, h2a, h4 and h7 (Crosnier et al., 2016). Polymorphisms in h7 are frequently observed in DBL domains, therefore any differences highlighted in this loop were assumed to be unlikely due to changes in binding ability. Therefore the binding site is more likely in one of the other three loops identified.

Analysis of surface residues predicted regions in MSPDBL1 and -2 likely to be involved in protein-protein interactions. This analysis highlighted helix 2a as a binding interface common to both MPSDBL1 and MPSDBL2 and is therefore likely to be involved in IgM binding. Helix 2a is located on the surface of the DBL within subdomain 2 and is therefore open to receptor interactions (Figure 6.1). However, MSPDBL1 and MPSDBL2 have also been shown to bind the MSP1 complex on the erythrocyte surface (Lin et al., 2014). This may occur through a common binding site. Competition studies would therefore be interesting to investigate whether the merozoite DBLs are still able to bind IgM when in complex.

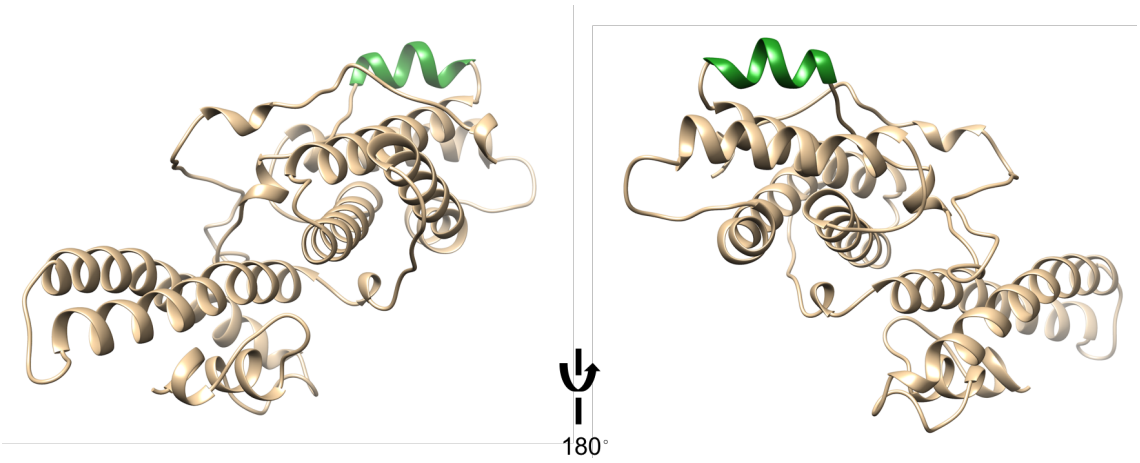


Figure 6.1: Location of MPSDBL1 helix 2a which is highlighted in green.

Two studies of critical binding residues for the Duffy antigen showed that the binding region lay in a central region spanning the fifth to the eight cysteine residues (C5 - C8) of the PvRII and PkαRI DBL domains (Ranjan and Chitnis, 1999; Singh et al., 2003). A

further study by Mayor et al. (2005) aimed to identify receptor-binding regions of several DBL domains from EBP and PfEMP1 protein families using deletion constructs. The receptor-binding residues for DBL domains that bind sialic acid on glycophorin A for erythrocyte invasion as well as complement receptor-1 and chondroitin sulphate A for cytoadherence were mapped to central regions (Mayor et al., 2005). The region identified spanned C5 to C8 except for the ICAM-1 binding region of the JDP8-DBLβ which also required the C2 domain for ICAM-1 binding. However, it was shown that the region spanning C1 - C4, as with the other DBLs investigated, was not required for binding. Helix 2a of the merozoite DBL domains is located in this region which corresponds to subdomain 1. This suggests that the IgM-binding region of the merozoite DBL may be a unique binding site.

6.5 Minimal IgM-binding region

Short versions of the merozoite DBL domains were originally expressed due to debate on their exact location when this work first started. Since the crystal structure of MSPDBL2 has been resolved, the exact location of the DBL domain is clearer (Hodder et al., 2012). Full-length DBL domains were expressed (Chapter 5) along with the shorter versions used previously, in order to investigate whether there were any differences in protein folding or binding between the DBLs of different lengths.

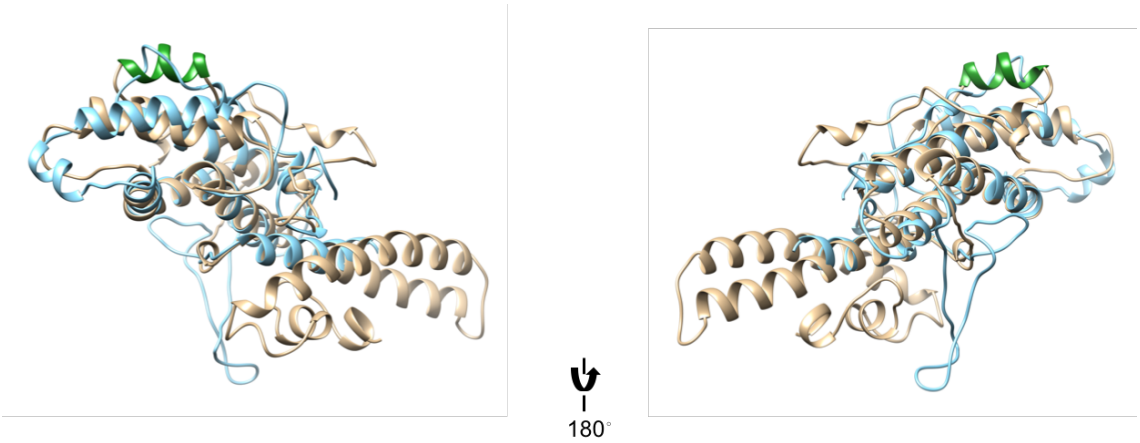


Figure 6.2: **Minimal IgM binding region of DBLζ from TM284var1 compared to the short merozoite DBL domain.** The TM284var1 DBLζ minimal IgM binding region (blue) lies within the short merozoite DBL (cream) and contains helix 2a (green).

No differences were observed, although the full length DBL domains produced higher protein yields with the addition of the extended hinge. The minimal IgM-binding region of DBLζ from TM284var1 has been identified as a central region consisting of subdomain 2 and flanking parts of subdomains 1 and 3 (Semblat et al., 2015). This region of the DBL showed no loss of IgM binding affinity by SPR analysis (Semblat et al., 2015). The short DBL domains produced here contain subdomain 2 with longer flanking regions of subdomains 1 and 3 than the minimal binding region (Figure 6.2). Data presented here are in agreement with the minimal IgM-binding region which lies within the short DBL domains.

6.6 Merozoite DBL-Fc fusion proteins as reagents for IgM purification

Currently, IgM purification is difficult due to it's size, and current purification reagents have limited efficacy. This is a problem for the manufacture of monoclonal IgM antibodies. C1q and mannose-binding lectin (MBL) are susceptible to proteolysis and such affinity ligands have IgM recovery rates of less than 95%. Merozoite Fc-fusion proteins may be developed as novel IgM binding reagents for purification due to their high affinity for IgM and their ability to be purified using protein G. However, IgM is sensitive to low pH therefore the elution methods developed here would need optimisation.

6.7 Merozoite DBL-Fc fusion proteins as malaria vaccines

The clinical success of Fc-fusion proteins has focussed on their use for drug delivery. A multimeric Fc fusion would be expected to enhance the interaction with the FcRn and further increase the plasma half-life and vaccine activity of the fused protein (Czajkowsky et al., 2015). In our laboratory, a single Fc has been engineered so that it can form hexameric oligomers which bind to high-affinity Fc receptors (Mekhaiel et al., 2011; Czajkowsky et al., 2015). The next step for malaria vaccine development, which is ongoing in our laboratory, is to further develop this multimeric scaffold. This may facilitate the production of a multimeric DBL-Fc fusion protein. A vaccine based on such a molecule may show improved immunogenicity for inducing adhesion-blocking antibodies and would have superior half-life. Further work is needed to identify binding motifs and critical residues. The methods developed here may also be used to produce other malarial Fc-fusion proteins which may have the ability to prime an immune response to malaria.

6.8 Merozoite DBL-Fc fusion proteins as therapeutics for the treatment of IgM-related diseases

DBL-Fc fusion proteins may be used as therapeutics for IgM-mediated diseases due to their high affinity for IgM. Cryoglobulinemia is a condition defined by the presence of cryoglobulins in serum. Cryoglobulins are immunoglobulins or immune complexes which precipitate from serum below temperatures of 37°C (Oliver et al., 2005). Cryoglobulinemia can be IgM-mediated by IgM cryoglobulins in type I cryoglobulinemia or by IgM-IgG complexes in type II (Terrier et al., 2012). Symptoms include fatigue, skin ulcers, neuropathy and kidney manifestations. Cryoglobulinemia may be present on its own or

may be due to an underlying infection or diseases such as myeloma, macroglobulinemia or B-cell non-Hodkgin's lymphoma (Terrier et al., 2012). The reduction of IgM by DBL-Fc fusion proteins could be used as therapeutics for these conditions. Complement activation can lead to an intense inflammatory response in type II cryoglobulinemia where immune complexes are deposited in blood vessel walls. These manifest clinically as purpuric skin eruptions and ulcers (Schamberg and Lake-Bakaar, 2007). The reduction of IgM in serum may also reduce complement activation and improve symptoms of vasculitis, which can range from mild to severe.

IgM is the main underlying component of Schnitzler's syndrome, a rare auto-inflammatory disease which manifests as a rash and joint pain, as well as fever and enlarged lymph nodes (Lipsker, 2010). Current treatments are ineffective and while symptoms can be debilitating, patients are also prone to lymphoproliferative disorders such as Waldenstrőm's macroglubulinemia (WM) and lymphoma (Lipsker, 2010). Treatment of the underlying IgM monoclonal gammopathy by the DBL-Fc fusion proteins could be investigated.

IgM reacts against myelin-associated glycoprotein (MAG) in anti-MAG demyelinating neuropathy, resulting in nerve damage and peripheral neuropathy (Lunn and Nobile-Orazio, 2012). Intravenous immunoglobulin treatment (IVIG) produces only short-term benefits, and a small clinical trial of rituximab was of poor quality and is awaiting a larger study (Lunn and Nobile-Orazio, 2012). Rituximab is currently the main treatment for IgM gammopathy in Waldenstrőm's macroglobulinemia. However, a common site effect is IgM flare which leads to the worsening of IgM-related symptoms (Treon et al., 2004). Such an IgM flare has an occurrence rate of 40 - 50 % when rituximab is used as a monotherapy for WM (Treon, 2009). It is possible that DBL-Fc fusion proteins could be used in combination with rituximab to control IgM levels in serum.

6.9 Concluding remarks

Novel merozoite DBL-Fc fusion proteins have been produced and used to investigate IgM binding. The experimental results described here were limited by protein yield and lack of homodimer formation. This was addressed by designing a flexible, extended hinge region which led to improved homodimer formation of the Fc-fusion proteins. The hinge design, however, is not optimal and could be further improved to make multimeric Fc-fusion proteins for vaccine development.

While sequence analysis was unsuccessful, structural analysis identified a region which is potentially involved in IgM binding. Helix 2a was predicted by structural modelling of MSPDBL isolates known to bind IgM and identified from analysis of surface residues as a protein-protein interaction interface. However, further experimental refinements are required to investigate this result and to identify any critical residues in this region.

This study has provided a foundation to further develop the scaffold and identify the IgM-binding site on the merozoite DBL domains. Binding to natural IgM is thought to prevent detection of the parasite by the more effective IgG Ab. Thus, since DBL-Fc fusion proteins may have immunogenicity for inducing anti-adhesion antibodies, a blood-stage vaccine based on these proteins may have the potential to reduce parasitemia in the blood (Fried and Duffy, 2015; Avril et al., 2011). The DBL-Fc fusion proteins have high affinity for IgM and may also be developed for therapeutic use against IgM-mediated diseases or as affinity ligands for IgM purification.

7 References

Adachi, T., Harumiya, S., Takematsu, H., Kozutsumi, Y., Wabl, M., Fujimoto, M., and Tedder, T. F. (2012). CD22 serves as a receptor for soluble IgM. *European journal of immunology*, 42(1):241–247.

Aggarwal, S. R. (2014). What's fueling the biotech engine—2012 to 2013. *Nature Biotechnology*, 32(1):32–39.

Akhouri, R., Sharma, A., Malhotra, P., and Sharma, A. (2008). Role of Plasmodium falciparum thrombospondin-related anonymous protein in host-cell interactions. - PubMed - NCBI. *Malaria journal*, 7(1):63.

Alberts, B. (2008). *Molecular Biology of the Cell*. Reference edition. Garland Science.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). The Adaptive Immune System.

Alfthan, K., Takkinen, K., Sizmann, D., Soderlund, H., and Teeri, T. T. (1995). Properties of a single-chain antibody containing different linker peptides. *Protein Engineering Design and Selection*, 8(7):725–731.

Andersen, J. T., Dalhus, B., Viuff, D., Ravn, B. T., Gunnarsen, K. S., Plumridge, A., Bunting, K., Antunes, F., Williamson, R., Athwal, S., Allan, E., Evans, L., Bjørås, M., Kjærulff, S., Sleep, D., Sandlie, I., and Cameron, J. (2014). Extending serum half-life of albumin by engineering neonatal Fc receptor (FcRn) binding. *The Journal of biological chemistry*, 289(19):13492–13502.

Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *Journal of molecular biology*, 211(4):943–958.

Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M., and Dwek, R. A. (2007). The impact of glycosylation on the biological function and structure of human immunoglobulins. . *Annual review of immunology*, 25(1):21–50.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, 22(2):195–201.

Artavanis-Tsakonas, K. and Riley, E. M. (2002). Innate immune response to malaria: rapid induction of IFN-gamma from human NK cells by live Plasmodium falciparum-infected erythrocytes. *Journal of immunology (Baltimore, Md : 1950)*, 169(6):2956–2963.

Atkin, J. D., Pleass, R. J., Owens, R. J., and Woof, J. M. (1996). Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly. *Journal of immunology (Baltimore, Md : 1950)*, 157(1):156–159.

Atkinson, S. C., Armistead, J. S., Mathias, D. K., Sandeu, M. M., Tao, D., Borhani-Dizaji, N., Tarimo, B. B., Morlais, I., Dinglasan, R. R., and Borg, N. A. (2015). The Anopheles-midgut APN1 structure reveals a new malaria transmission-blocking vaccine epitope. *Nature structural & molecular biology*, 22(7):532–539.

Audard, V., Georges, B., Vanhille, P., Toly, C., Deroure, B., Fakhouri, F., Cuvelier, R., Belenfant, X., Surin, B., Aucouturier, P., Mougenot, B., and Ronco, P. (2008). Renal lesions associated with IgM-secreting monoclonal proliferations: revisiting the disease spectrum. *Clinical journal of the American Society of Nephrology : CJASN*, 3(5):1339–1349.

Avril, M., Cartwright, M. M., Hathaway, M. J., and Smith, J. D. (2011). Induction of strain-transcendent antibodies to placental-type isolates with VAR2CSA DBL3 or DBL5 recombinant proteins. *Malaria journal*, 10(1):36.

Bakema, J. E. and van Egmond, M. (2011). The human immunoglobulin A Fc receptor FcαRI: a multifaceted regulator of mucosal immunity. *Mucosal Immunology*, 4(6):612–624.

Baldini, L., Nobile Orazio, E., Guffanti, A., Barbieri, S., Carpo, M., Cro, L., Cesana, B., Damilano, I., and Maiolo, A. T. (1994). Peripheral neuropathy in IgM monoclonal gammopathy and wäldenstrom's macroglobulinemia: A frequent complication in elderly males with low MAG-reactive serum monoclonal component. *American journal of hematology*, 45(1):25–31.

Barfod, L., Dalgaard, M. B., Pleman, S. T., Ofori, M. F., Pleass, R. J., and Hviid, L. (2011). Evasion of immunity to Plasmodium falciparum malaria by IgM masking of protective IgG epitopes in infected erythrocyte surface-exposed PfEMP1. *Proceedings of the National Academy of Sciences of the United States of America*, 108(30):12485–12490.

Baron, S., Crutcher, J. M., and Hoffman, S. L. (1996). Malaria. In *Medical Microbiology*. University of Texas Medical Branch at Galveston, Galveston (TX).

Baron, S. and Klimpel, G. R. (1996). Immune Defenses. In *Medical Microbiology*. University of Texas Medical Branch at Galveston, Galveston (TX).

Barragan, A., Fernandez, V., Chen, Q., von Euler, A., Wahlgren, M., and Spillmann, D. (2000a). The Duffy-binding-like domain 1 of Plasmodium falciparumerythrocyte membrane protein 1 (PfEMP1) is a heparan sulfate ligand that requires 12 mers for binding. *Blood*, 95(11):3594–3599.

Barragan, A., Kremsner, P. G., Wahlgren, M., and Carlson, J. (2000b). Blood group A antigen is a coreceptor in Plasmodium falciparum rosetting. *Infection and immunity*, 68(5):2971–2975.

Berger, S. S., Turner, L., Wang, C. W., Petersen, J. E. V., Kraft, M., Lusingu, J. P. A., Mmbando, B., Marquard, A. M., Bengtsson, D. B. A. C., Hviid, L., Nielsen, M. A., Theander, T. G., and Lavstsen, T. (2013). Plasmodium falciparum Expressing Domain Cassette 5 Type PfEMP1 (DC5-PfEMP1) Bind PECAM1. *PLoS One*, 8(7):e69117.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(Web Server issue):W252–8.

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S., and Whitlow, M. (1988). Single-chain antigen-binding proteins. *Science*, 242(4877):423–426.

Birkholtz, L.-M., Blatch, G., Coetzer, T. L., Hoppe, H. C., Human, E., Morris, E. J., Ngcete, Z., Oldfield, L., Roth, R., Shonhai, A., Stephens, L., and Louw, A. I. (2008). Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria journal*, 7(1):1.

Boes, M., Esau, C., Fischer, M. B., Schmidt, T., Carroll, M., and Chen, J. (1998). Enhanced B-1 cell development, but impaired IgG antibody responses in mice deficient in secreted IgM. *Journal of immunology (Baltimore, Md : 1950)*, 160(10):4776–4787.

Boes, M., Schmidt, T., Linkemann, K., Beaudette, B. C., Marshak-Rothstein, A., and Chen, J. (2000). Accelerated development of IgG autoantibodies and autoimmune disease in the absence of secreted IgM. *Proceedings of the National Academy of Sciences of the United States of America*, 97(3):1184–1189.

Bolad, A., Farouk, S. E., Israelsson, E., Dolo, A., Doumbo, O. K., Nebié, I., Maiga, B., Kouriba, B., Luoni, G., Sirima, B. S., Modiano, D., Berzins, K., and Troye-Blomberg, M. (2005). Distinct interethnic differences in immunoglobulin G class/subclass and immunoglobulin M antibody responses to malaria antigens but not in immunoglobulin G responses to nonmalarial antigens in sympatric tribes living in West Africa. *Scandinavian Journal of Immunology*, 61(4):380–386.

Boudin, C., Chumpitazi, B., Dziegiel, M., Peyron, F., Picot, S., Hogh, B., and Ambroise-Thomas, P. (1993). Possible role of specific immunoglobulin M antibodies to Plasmodium falciparum antigens in immunoprotection of humans living in a hyperendemic area, Burkina Faso. *Journal of Clinical Microbiology*, 31(3):636–641.

Brambell, F. W. R. (1966). The Transmission of Immunity from Mother to Young and the Catabolism of Immunoglobulins. *The Lancet*, 288(7473):1087–1093.

Brändlein, S., Pohle, T., Ruoff, N., Wozniak, E., Müller-Hermelink, H.-K., and Vollmers, H. P. (2003). Natural IgM antibodies and immunosurveillance mechanisms against epithelial cancer cells in humans. *Cancer research*, 63(22):7995–8005.

Brandtzaeg, P. and Prydz, H. (1984). Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature*, 311(5981):71–73.

Brown, J., Greenwood, B. M., and Terry, R. J. (1986). Cellular mechanisms involved in recovery from acute malaria in Gambian children. *Parasite immunology*, 8(6):551–564.

Buffet, P. A., Gamain, B., Scheidig, C., Baruch, D., Smith, J. D., Hernandez-Rivas, R., Pouvelle, B., Oishi, S., Fujii, N., Fusai, T., Parzy, D., Miller, L. H., Gysin, J., and Scherf, A. (1999). Plasmodium falciparum domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. *Proceedings of the National Academy of Sciences*, 96(22):12743–12748.

Burgess, B. R., Schuck, P., and Garboczi, D. N. (2005). Dissection of merozoite surface protein 3, a representative of a family of Plasmodium falciparum surface proteins, reveals an oligomeric and highly elongated molecule. *Journal of Biological Chemistry*, 280(44):37236–37245.

Buyue, Y., Liu, T., Kulman, J. D., Toby, G. G., Kamphaus, G. D., Patarroyo-White, S., Lu, Q., Reidy, T. J., Mei, B., Jiang, H., Pierce, G. F., Sommer, J. M., and Peters, R. T. (2014). A single chain variant of factor VIII Fc fusion protein retains normal in vivo efficacy but exhibits altered in vitro activity. *PLoS One*, 9(11):e113600.

Camp, B. J. and Magro, C. M. (2012). Cutaneous macroglobulinosis: a case series. *Journal of Cutaneous Pathology*, 39(10):962–970.

Carlson, J., Helmby, H., Hill, A. V., Brewster, D., Greenwood, B. M., and Wahlgren, M. (1990). Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet*, 336(8729):1457–1460.

Charles A Janeway, J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). Pathogens have evolved various means of evading or subverting normal host defenses. *Nature reviews Immunology*, 8(3):205–217.

Chattopadhyay, R., Taneja, T., Chakrabarti, K., Pillai, C. R., and Chitnis, C. E. (2004). Molecular analysis of the cytoadherence phenotype of a Plasmodium falciparum field isolate that binds intercellular adhesion molecule—1. *Molecular and biochemical parasitology*, 133(2):255–265.

Chen, K. and Cerutti, A. (2011). The function and regulation of immunoglobulin D. *Current opinion in immunology*, 23(3):345–352.

Chen, Q., Barragan, A., Fernandez, V., Sundström, A., Schlichtherle, M., Sahlén, A., Carlson, J., Datta, S., and Wahlgren, M. (1998). Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum. *The Journal of experimental medicine*, 187(1):15–23.

Chen, Q., Heddini, A., Barragan, A., Fernandez, V., Pearce, S. F., and Wahlgren, M. (2000). The semiconserved head structure of Plasmodium falciparum erythrocyte

membrane protein 1 mediates binding to multiple independent host receptors. *The Journal of experimental medicine*, 192(1):1–10.

Chen, X., Zaro, J. L., and Shen, W.-C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, 65(10):1357–1369.

Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A. H. (1999). Directed evolution of a fungal peroxidase. *Nature Biotechnology*, 17(4):379–384.

Chiu, C. Y. H., Hodder, A. N., Lin, C. S., Hill, D. L., Li Wai Suen, C. S. N., Schofield, L., Siba, P. M., Mueller, I., Cowman, A. F., and Hansen, D. S. (2015). Antibodies to the Plasmodium falciparum Proteins MSPDBL1 and MSPDBL2 Opsonize Merozoites, Inhibit Parasite Growth, and Predict Protection From Clinical Malaria. *The Journal of infectious diseases*, 212(3):406–415.

Cooper, M. D. and Herrin, B. R. (2010). How did our complex immune system evolve? *Nature reviews Immunology*, 10(1):2–3.

Cooper, N. R. (1985). The classical complement pathway: activation and regulation of the first complement component. *Advances in immunology*, 37:151–216.

Couper, K. N., Phillips, R. S., Brombacher, F., and Alexander, J. (2005). Parasite-specific IgM plays a significant role in the protective immune response to asexual erythrocytic stage Plasmodium chabaudi AS infection. *Parasite immunology*, 27(5):171–180.

Cowman, A. F., Berry, D., and Baum, J. (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *The Journal of Cell Biology*, 198(6):961–971.

Crasto, C. J. and Feng, J.-a. (2000). LINKER: a program to generate linker sequences for fusion proteins. *Protein Engineering*, 13(5):309–312.

Crosnier, C., Iqbal, Z., Knuepfer, E., Maciuca, S., Perrin, A. J., Kamuyu, G., Goulding, D., Bustamante, L. Y., Miles, A., Moore, S. C., Dougan, G., Holder, A. A., Kwiatkowski, D. P., Rayner, J. C., Pleass, R. J., and Wright, G. J. (2016). Binding of Plasmodium falciparum merozoite surface proteins DBLMSP and DBLMSP2 to human

immunoglobulin M is conserved amongst broadly diverged sequence variants. *Journal of Biological Chemistry*, page jbc.M116.722074.

Crosnier, C., Wanaguru, M., McDade, B., Osier, F. H., Marsh, K., Rayner, J. C., and Wright, G. J. (2013). A Library of Functional Recombinant Cell-surface and Secreted P. falciparumMerozoite Proteins. *Molecular & cellular proteomics : MCP*, 12(12):3976–3986.

Czajkowsky, D. M., Andersen, J. T., Fuchs, A., Wilson, T. J., Mekhaiel, D., Colonna, M., He, J., Shao, Z., Mitchell, D. A., Wu, G., Dell, A., Haslam, S., Lloyd, K. A., Moore, S. C., Sandlie, I., Blundell, P. A., and Pleass, R. J. (2015). Developing the IVIG biomimetic, Hexa-Fc, for drug and vaccine applications. *Scientific reports*, 5:9526.

Czajkowsky, D. M., Hu, J., Shao, Z., and Pleass, R. J. (2012). Fc-fusion proteins: new developments and future perspectives. *EMBO molecular medicine*, 4(10):1015–1028.

Czajkowsky, D. M., Salanti, A., Ditlev, S. B., Shao, Z., Ghumra, A., Rowe, J. A., and Pleass, R. J. (2010). IgM, Fc mu Rs, and malarial immune evasion. *Journal of immunology (Baltimore, Md : 1950)*, 184(9):4597–4603.

Czajkowsky, D. M. and Shao, Z. (2009). The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35):14960–14965.

Demanga, C. G., Daher, L.-J., Prieur, E., Blanc, C., Pérignon, J.-L., Bouharoun-Tayoun, H., and Druilhe, P. (2010). Toward the rational design of a malaria vaccine construct using the MSP3 family as an example: contribution of antigenicity studies in humans. *Infection and immunity*, 78(1):486–494.

Dinglasan, R. R., Armistead, J. S., Nyland, J. F., Jiang, X., and Mao, H. Q. (2013). Single-dose microparticle delivery of a malaria transmission-blocking vaccine elicits a long-lasting functional antibody response. *Current Molecular Medicine*, 13(4):479–487.

Douglas, A. D., Williams, A. R., Illingworth, J. J., Kamuyu, G., Biswas, S., Goodman, A. L., Wyllie, D. H., Crosnier, C., Miura, K., Wright, G. J., Long, C. A., Osier, F. H., Marsh, K., Turner, A. V., Hill, A. V. S., and Draper, S. J. (2011). The blood-stage

malaria antigen PfRH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody. *Nature communications*, 2:601.

Druilhe, P., Spertini, F., Soesoe, D., Corradin, G., Mejia, P., Singh, S., Audran, R., Bouzidi, A., Oeuvray, C., and Roussilhon, C. (2005). A malaria vaccine that elicits in humans antibodies able to kill Plasmodium falciparum. *PLoS medicine*, 2(11):e344.

Edholm, E.-S., Bengten, E., and Wilson, M. (2011). Insights into the function of IgD. *Developmental and comparative immunology*, 35(12):1309–1316.

Ehrenstein, M. R. and Notley, C. A. (2010). The importance of natural IgM: scavenger, protector and regulator. *Nature reviews Immunology*, 10(11):778–786.

Elias, P. M. (2007). The skin barrier as an innate immune element. *Seminars in immunopathology*, 29(1):3–14.

Elloso, M. M., van der Heyde, H. C., Waa, J. A. v., Manning, D. D., and Weidanz, W. P. (1994). Inhibition of Plasmodium falciparum in vitro by human gamma delta T cells. *Journal of immunology (Baltimore, Md : 1950)*, 153(3):1187–1194.

Feldmann, M. (2002). Development of anti-TNF therapy for rheumatoid arthritis. *Nature reviews Immunology*, 2(5):364–371.

Fernández-Robledo, J. A. and Vasta, G. R. (2010). Production of recombinant proteins from protozoan parasites. *Trends in parasitology*, 26(5):244–254.

Flick, K., Ahuja, S., Chene, A., Bejarano, M. T., and Chen, Q. (2004). Optimized expression of Plasmodium falciparum erythrocyte membrane protein 1 domains in Escherichia coli. *Malaria journal*, 3(1):50.

Flick, K., Scholander, C., Chen, Q., Fernandez, V., Pouvelle, B., Gysin, J., and Wahlgren, M. (2001). Role of nonimmune IgG bound to PfEMP1 in placental malaria. *Science*, 293(5537):2098–2100.

Fournier-Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C.-W., and Tseng, V. S. (2014). SPMF: a Java open-source pattern mining library. *The Journal of Machine Learning Research*, 15(1):3389–3393.

Fried, M. and Duffy, P. E. (2015). Designing a VAR2CSA-based vaccine to prevent placental malaria. *Vaccine*, 33(52):7483–7488.

Gautam, S. and Loh, K.-C. (2011). Immunoglobulin-M purification — Challenges and perspectives. *Biotechnology Advances*, 29(6):840–849.

Geijtenbeek, T. B., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Middel, J., Cornelissen, I. L., Nottet, H. S., KewalRamani, V. N., Littman, D. R., Figdor, C. G., and van Kooyk, Y. (2000). DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*, 100(5):587–597.

George, R. A. and Heringa, J. (2002). An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering*, 15(11):871–879.

Ghetie, V., Hubbard, J. G., Kim, J.-K., Tsen, M.-F., Lee, Y., and Ward, E. S. (1996). Abnormally short serum half-lives of IgG in β2-microglobulin-deficient mice. *European journal of immunology*, 26(3):690–696.

Ghosh, S., Malhotra, P., Lalitha, P. V., Guha-Mukherjee, S., and Chauhan, V. S. (2002). Expression of Plasmodium falciparum C-terminal region of merozoite surface protein (PfMSP119), a potential malaria vaccine candidate, in tobacco. *Plant Science*, 162(3):335–343.

Ghumra, A., Semblat, J.-P., McIntosh, R. S., Raza, A., Rasmussen, I. B., Braathen, R., Johansen, F.-E., Sandlie, I., Mongini, P. K., Rowe, J. A., and Pleass, R. J. (2008). Identification of residues in the Cmu4 domain of polymeric IgM essential for interaction with Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1). *Journal of immunology (Baltimore, Md : 1950)*, 181(3):1988–2000.

Gould, H. J. and Sutton, B. J. (2008). IgE in allergy and asthma today. *Nature reviews Immunology*, 8(3):205–217.

Greenwood, B. M. and Targett, G. A. T. (2011). Malaria vaccines and the new malaria agenda. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(11):1600–1607.

Gregory, J. A., Li, F., Tomosada, L. M., Cox, C. J., Topol, A. B., Vinetz, J. M., and

Mayfield, S. (2012). Algae-Produced Pfs25 Elicits Antibodies That Inhibit Malaria Transmission. *PLoS One*, 7(5):e37179.

Hayes, J. M., Cosgrave, E. F. J., Struwe, W. B., Wormald, M., Davey, G. P., Jefferis, R., and Rudd, P. M. (2014). Glycosylation and Fc Receptors. In *Fc Receptors*, pages 165–199. Springer International Publishing, Cham.

Hermsen, C. C., Konijnenberg, Y., Mulder, L., Loé, C., van Deuren, M., van der Meer, J. W. M., van Mierlo, G. J., Eling, W. M. C., Hack, C. E., and Sauerwein, R. W. (2003). Circulating concentrations of soluble granzyme A and B increase during natural and experimental Plasmodium falciparum infections. *Clinical and experimental immunology*, 132(3):467–472.

Higgins, M. K. and Carrington, M. (2014). Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein science : a publication of the Protein Society*, 23(4):354–365.

Hodder, A. N., Czabotar, P. E., Uboldi, A. D., Clarke, O. B., Lin, C. S., Healer, J., Smith, B. J., and Cowman, A. F. (2012). Insights into Duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from Plasmodium falciparum. *The Journal of biological chemistry*, 287(39):32922–32939.

Hoffman, S. L., Goh, L. M. L., Luke, T. C., Schneider, I., Le, T. P., Doolan, D. L., Sacci, J., de la Vega, P., Dowler, M., Paul, C., Gordon, D. M., Stoute, J. A., Church, L. W. P., Sedegah, M., Heppner, D. G., Ballou, W. R., and Richie, T. L. (2002). Protection of humans against malaria by immunization with radiation-attenuated Plasmodium falciparum sporozoites. *The Journal of infectious diseases*, 185(8):1155–1164.

Honjo, K., Kubagawa, Y., Jones, D. M., Dizon, B., Zhu, Z., Ohno, H., Izui, S., Kearney, J. F., and Kubagawa, H. (2012). Altered Ig levels and antibody responses in mice deficient for the Fc receptor for IgM (FcμR). *Proceedings of the National Academy of Sciences of the United States of America*, 109(39):15882–15887.

Hviid, L. (2005). Naturally acquired immunity to Plasmodium falciparum malaria in Africa. *Acta tropica*, 95(3):270–275.

Hviid, L. (2010). The role of Plasmodium falciparum variant surface antigens in protective immunity and vaccine development. *Human vaccines*, 6(1):84–89.

Hviid, L., Kurtzhals, J. A., Adabayeri, V., Loizon, S., Kemp, K., Goka, B. Q., Lim, A., Mercereau-Puijalon, O., Akanmori, B. D., and Behr, C. (2001). Perturbation and proinflammatory type activation of V delta 1(+) gamma delta T cells in African children with Plasmodium falciparum malaria. *Infection and immunity*, 69(5):3190–3196.

Ioannidis, J. P. A., Karassa, F. B., Druyts, E., Thorlund, K., and Mills, E. J. (2013). Biologic agents in rheumatology: unmet issues after 200 trials and [dollar]200 billion sales. *Nature Reviews Rheumatology*, 9(11):665–673.

Jagannathan, P., Kim, C. C., Greenhouse, B., Nankya, F., Bowen, K., Eccles-James, I., Muhindo, M. K., Arinaitwe, E., Tappero, J. W., Kamya, M. R., Dorsey, G., and Feeney, M. E. (2014). Loss and dysfunction of Vδ2+ $\gamma\delta$ T cells are associated with clinical tolerance to malaria. *Science translational medicine*, 6(251):251ra117–251ra117.

Janeway, C. A. and Medzhitov, R. (2002). Innate immune recognition. *Annual review of immunology*, 20:197–216.

Jeppesen, A., Ditlev, S. B., Soroka, V., Stevenson, L., Turner, L., Dzikowski, R., Hviid, L., and Barfod, L. (2015). Multiple Plasmodium falciparum Erythrocyte Membrane Protein 1 Variants per Genome Can Bind IgM via Its Fc Fragment Fcμ. *Infection and immunity*, 83(10):3972–3981.

Johansen, Braathen, and Brandtzaeg (2000). Role of J Chain in Secretory Immunoglobulin Formation. *Scandinavian Journal of Immunology*, 52(3):240–248.

Juillerat, A., Lewit-Bentley, A., Guillotte, M., Gangnard, S., Hessel, A., Baron, B., Vigan-Womas, I., England, P., Mercereau-Puijalon, O., and Bentley, G. A. (2011). Structure of a Plasmodium falciparum PfEMP1 rosetting domain reveals a role for the N-terminal segment in heparin-mediated rosette inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 108(13):5243–5248.

Junghans, R. P. and Anderson, C. L. (1996). The protection receptor for IgG catabolism is the beta2-microglobulin-containing neonatal intestinal transport receptor. *Proceedings of the National Academy of Sciences*, 93(11):5512–5516.

Kaslow, D. C. and Biernaux, S. (2015). RTS,S: Toward a first landmark on the Malaria Vaccine Technology Roadmap. *Vaccine*, 33(52):7425–7432.

Kaveri, S. V., Silverman, G. J., and Bayry, J. (2012). Natural IgM in immune equilibrium and harnessing their therapeutic potential. *Journal of immunology (Baltimore, Md : 1950)*, 188(3):939–945.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6):845–858.

Kerr, M. A. (1990). The structure and function of human IgA. *The Biochemical journal*, 271(2):285–296.

Kinet, J.-P. (1999). The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. *Annual review of immunology*, 17:931–972.

King, B. F. and Wilkinson, B. J. (1981). Binding of human immunoglobulin G to protein A in encapsulated Staphylococcus aureus. *Infection and immunity*, 33(3):666–672.

Klimovich, V. B. (2011). IgM and its receptors: structural and functional aspects. *Biochemistry. Biokhimiia*, 76(5):534–549.

Koch, C., Boesman, M., and Gitlin, D. (1967). Maternofoetal transfer of gamma G immunoglobulins. *Nature*, 216(5120):1116–1117.

Krapp, S., Mimura, Y., Jefferis, R., Huber, R., and Sondermann, P. (2003). Structural Analysis of Human IgG-Fc Glycoforms Reveals a Correlation Between Glycosylation and Structural Integrity. *Journal of molecular biology*, 325(5):979–989.

Kubagawa, H., Oka, S., Kubagawa, Y., Torii, I., Takayama, E., Kang, D.-W., Gartland, G. L., Bertoli, L. F., Mori, H., Takatsu, H., Kitamura, T., Ohno, H., and Wang, J.-Y. (2009). Identity of the elusive IgM Fc receptor (FcmuR) in humans. *The Journal of experimental medicine*, 206(12):2779–2793.

Kumar, K. A., Sano, G.-i., Boscardin, S., Nussenzweig, R. S., Nussenzweig, M. C., Zavala, F., and Nussenzweig, V. (2006). The circumsporozoite protein is an immunodominant protective antigen in irradiated sporozoites. *Nature*, 444(7121):937–940.

Lau, C. K. Y., Turner, L., Jespersen, J. S., Lowe, E. D., Petersen, B., Wang, C. W., Petersen, J. E. V., Lusingu, J., Theander, T. G., Lavstsen, T., and Higgins, M. K. (2015). Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. *Cell host & microbe*, 17(1):118–129.

Levin, D., Golding, B., Strome, S. E., and Sauna, Z. E. (2015). Fc fusion as a platform technology: potential for modulating immunogenicity. *Trends in biotechnology*, 33(1):27–34.

Lin, C. S., Uboldi, A. D., Marapana, D., Czabotar, P. E., Epp, C., Bujard, H., Taylor, N. L., Perugini, M. A., Hodder, A. N., and Cowman, A. F. (2014). The Merozoite Surface Protein 1 Complex Is a Platform for Binding to Human Erythrocytes by Plasmodium falciparum. *The Journal of biological chemistry*, 289(37):25655–25669.

Lipsker, D. (2010). The Schnitzler syndrome. *Orphanet Journal of Rare Diseases*, 5(1):38.

Liu, H. and May, K. (2012). Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *mAbs*, 4(1):17–23.

Lubinski, J. M., Lazear, H. M., Awasthi, S., Wang, F., and Friedman, H. M. (2011). The Herpes Simplex Virus 1 IgG Fc Receptor Blocks Antibody-Mediated Complement Activation and Antibody-Dependent Cellular Cytotoxicity In Vivo. *Journal of virology*, 85(7):3239–3249.

Lunn, M. P. T. and Nobile-Orazio, E. (2012). Immunotherapy for IgM anti-myelin-associated glycoprotein paraprotein-associated peripheral neuropathies. *The Cochrane database of systematic reviews*, 5:CD002827.

Luster, A. D. (2002). The role of chemokines in linking innate and adaptive immunity. *Current opinion in immunology*, 14(1):129–135.

Ma, B., Elkayam, T., Wolfson, H., and Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences*, 100(10):5772–5777.

Maier, A. G., Cooke, B. M., Cowman, A. F., and Tilley, L. (2009). Malaria parasite proteins that remodel the host erythrocyte. *Nature reviews Microbiology*, 7(5):341–354.

Mallery, D. L., McEwan, W. A., Bidgood, S. R., Towers, G. J., Johnson, C. M., and James, L. C. (2010). Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proceedings of the National Academy of Sciences*, 107(46):19985–19990.

Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M., Wormald, M. R., Lellouch, A. C., Rudd, P. M., Woof, J. M., and Dwek, R. A. (1998). The Glycosylation and Structure of Human Serum IgA1, Fab, and Fc Regions and the Role of N-Glycosylation on Fcα Receptor Interactions. *The Journal of biological chemistry*, 273(4):2260–2272.

Mayor, A., Bir, N., Sawhney, R., Singh, S., Pattnaik, P., Singh, S. K., Sharma, A., and Chitnis, C. E. (2005). Receptor-binding residues lie in central regions of Duffy-binding-like domains involved in red cell invasion and cytoadherence by malaria parasites. *Blood*, 105(6):2557–2563.

McCall, M. B. B. and Sauerwein, R. W. (2010). Interferon-gamma–central mediator of protective immune responses against the pre-erythrocytic and blood stage of malaria. *Journal of Leukocyte Biology*, 88(6):1131–1143.

McColl, D. J. and Anders, R. F. (1997). Conservation of structural motifs and antigenic diversity in the Plasmodium falciparum merozoite surface protein-3 (MSP-3)1Note: Nucleotide sequence data reported in this manuscript are available in the EMBL, GenBank^{TM} and DDJB data bases under the accession numbers L07944, L28825, U08851 and U08852.1. *Molecular and biochemical parasitology*, 90(1):21–31.

McMullen, M. E., Hart, M. L., Walsh, M. C., Buras, J., Takahashi, K., and Stahl, G. L. (2006). Mannose-binding lectin binds IgM to activate the lectin complement pathway in vitro and in vivo. *Immunobiology*, 211(10):759–766.

Medzhitov, R. (2001). Toll-like receptors and innate immunity. *Nature reviews Immunology*, 1(2):135–145.

Medzhitov, R. and Janeway, C. A. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell*, 91(3):295–298.

Mekhaiel, D. N. A., Czajkowsky, D. M., Andersen, J. T., Shi, J., El-Faham, M., Doenhoff, M., McIntosh, R. S., Sandlie, I., He, J., Hu, J., Shao, Z., and Pleass, R. J. (2011). Polymeric human Fc-fusion proteins with modified effector functions. *Scientific reports*, 1:124.

Mercereau-Puijalon, O., Guillotte, M., and Vigan-Womas, I. (2008). Rosetting in Plasmodium falciparum: a cytoadherence phenotype with multiple actors. *Transfusion clinique et biologique : journal de la Société française de transfusion sanguine*, 15(1-2):62–71.

Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature*, 415(6872):673–679.

Miller, L. H., Rowe, J. A., Moulds, J. M., and Newbold, C. I. (1997). P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1 : Abstract : Nature. *Nature*, 388(6639):292–295.

Mogensen, T. H. (2009). Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses. *Clinical microbiology reviews*, 22(2):240–273.

Moon, J. J., Chu, H. H., Pepper, M., McSorley, S. J., Jameson, S. C., Kedl, R. M., and Jenkins, M. K. (2007). Naive CD4+ T Cell Frequency Varies for Different Epitopes and Predicts Repertoire Diversity and Response Magnitude. *Immunity*, 27(2):203–213.

Mueller, A.-K., Kohlhepp, F., Hammerschmidt, C., and Michel, K. (2010). Invasion of mosquito salivary glands by malaria parasites: Prerequisites and defense strategies. *International Journal For Parasitology*, 40(11):1229–1235.

Müller, R., Gräwert, M. A., Kern, T., Madl, T., Peschek, J., Sattler, M., Groll, M., and Buchner, J. (2013). High-resolution structures of the IgM Fc domains reveal principles of its hexamer formation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(25):10183–10188.

Nezlin, R. and Ghetie, V. (2004). Interactions of immunoglobulins outside the antigen-combining site. *Advances in immunology*, 82:155–215.

Ngo, J. C. K., Huang, M., Roth, D. A., Furie, B. C., and Furie, B. (2008). Crystal

structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. *Structure (London, England : 1993)*, 16(4):597–606.

Ocana-Morgner, C., Mota, M. M., and Rodriguez, A. (2003). Malaria Blood Stage Suppression of Liver Stage Immunity by Dendritic Cells. *The Journal of experimental medicine*, 197(2):143–151.

Oliver, M., Coton, T., Ragot, C., Chianéa, D., Moalic, J.-L., and Debonne, J.-M. (2005). Cryoglobulins: detection, type determination and quantitation. A study of healthy subjects and patients with chronic hepatitis C. *Annales de biologie clinique*, 63(1):59–65.

Ord, R. L., Rodriguez, M., and Lobo, C. A. (2015). Malaria invasion ligand RH5 and its prime candidacy in blood-stage malaria vaccine design. - PubMed - NCBI. *Human Vaccines & Immunotherapeutics*, 11(6):1465–1473.

Ord, R. L., Rodriguez, M., Yamasaki, T., Takeo, S., Tsuboi, T., and Lobo, C. A. (2012). Targeting sialic acid dependent and independent pathways of invasion in Plasmodium falciparum. *PLoS One*, 7(1):e30251.

Osier, F. H. A., Polley, S. D., Mwangi, T., Lowe, B., Conway, D. J., and Marsh, K. (2007). Naturally acquired antibodies to polymorphic and conserved epitopes of Plasmodium falciparum merozoite surface protein 3. *Parasite immunology*, 29(8):387–394.

Ouattara, A. and Laurens, M. B. (2015). Vaccines against malaria. *Clinical Infectious Diseases*, 60(6):930–936.

Ouchida, R., Mori, H., Hase, K., Takatsu, H., Kurosaki, T., Tokuhisa, T., Ohno, H., and Wang, J.-Y. (2012). Critical role of the IgM Fc receptor in IgM homeostasis, B-cell survival, and humoral immune responses. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40):E2699–706.

Panda, S. and Ding, J. L. (2015). Natural antibodies bridge innate and adaptive immunity. *Journal of immunology (Baltimore, Md : 1950)*, 194(1):13–20.

Peters, R. T., Toby, G., Lu, Q., Liu, T., Kulman, J. D., Low, S. C., Bitonti, A. J., and Pierce, G. F. (2013). Biochemical and functional characterization of a recombinant monomeric factor VIII-Fc fusion protein. *Journal of thrombosis and haemostasis : JTH*, 11(1):132–141.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

PG, H., A, C., and JH, P. (1982). Specific monoclonal IgM is a potent adjuvant in murine malaria vaccination. *Nature*, 302(5905):256–258.

Pleass, R. J., Areschoug, T., Lindahl, G., and Woof, J. M. (2001). Streptococcal IgA-binding proteins bind in the Calpha 2-Calpha 3 interdomain region and inhibit binding of IgA to human CD89. *The Journal of biological chemistry*, 276(11):8197–8204.

Pleass, R. J. and Holder, A. A. (2005). Antibody-based therapies for malaria. *Nature reviews Microbiology*, 3(11):893–899.

Pleass, R. J., Moore, S. C., Stevenson, L., and Hviid, L. (2015). Immunoglobulin M: Restrainer of Inflammation and Mediator of Immune Evasion by Plasmodium falciparum Malaria. *Trends in parasitology*, 0(0).

Pleass, R. J. and Woof, J. M. (2001). Fc receptors and immunity to parasites. *Trends in parasitology*, 17(11):545–551.

Polley, S. D., Tetteh, K. K. A., Lloyd, J. M., Akpogheneta, O. J., Greenwood, B. M., Bojang, K. A., and Conway, D. J. (2007). Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *The Journal of infectious diseases*, 195(2):279–287.

Prince, F. H. M., de Bekker-Grob, E. W., Twilt, M., van Rossum, M. A. J., Hoppenreijs, E. P. A. H., ten Cate, R., Koopman-Keemink, Y., Gorter, S. L., Raat, H., and van Suijlekom-Smit, L. W. A. (2011). An analysis of the costs and treatment success of etanercept in juvenile idiopathic arthritis: results from the Dutch Arthritis and Biologicals in Children register. *Rheumatology (Oxford, England)*, 50(6):1131–1136.

Qin, S. and Zhou, H.-X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics (Oxford, England)*, 23(24):3386–3387.

Quinn, P. M., Dunne, D. W., Moore, S. C., and Pleass, R. J. (2016). IgE-tailpiece associates with α-1-antitrypsin (A1AT) to protect IgE from proteolysis without compromising its ability to interact with FcεRI. *Scientific reports*, 6:20509.

Randall, T. D., King, L. B., and Corley, R. B. (1990). The biological effects of IgM hexamer formation. . *European journal of immunology*, 20:1971–1979.

Ranjan, A. and Chitnis, C. E. (1999). Mapping regions containing binding residues within functional domains of Plasmodium vivax and Plasmodium knowlesi erythrocyte-binding proteins. *Proceedings of the National Academy of Sciences*, 96(24):14067–14072.

Rapaka, R. R., Ricks, D. M., Alcorn, J. F., Chen, K., Khader, S. A., Zheng, M., Plevy, S., Bengten, E., and Kolls, J. K. (2010). Conserved natural IgM antibodies mediate innate and adaptive immunity against the opportunistic fungus Pneumocystis murina. *The Journal of experimental medicine*, 207(13):2907–2919.

Rask, T. S., Hansen, D. A., Theander, T. G., Gorm Pedersen, A., and Lavstsen, T. (2010). Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes–divide and conquer. *PLoS computational biology*, 6(9).

Ravetch, J. V. and Bolland, S. (2001). IgG Fc receptors. *Annual review of immunology*, 19:275–290.

Recker, M., Buckee, C. O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., Springer, A. L., Gupta, S., and Newbold, C. I. (2011). Antigenic Variation in Plasmodium falciparum Malaria Involves a Highly Structured Switching Pattern. *PLoS pathogens*, 7(3):e1001306.

Reddy, K. S., Pandey, A. K., Singh, H., Sahar, T., Emmanuel, A., Chitnis, C. E., Chauhan, V. S., and Gaur, D. (2014). Bacterially expressed full-length recombinant Plasmodium falciparum RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infection and immunity*, 82(1):152–164.

Riley, E. M. (1999). Is T-cell priming required for initiation of pathology in malaria infections? *Immunology Today*, 20(5):228–233.

Robinson, C. R. and Sauer, R. T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences*, 95(11):5929–5934.

Roopenian, D. C., Christianson, G. J., Sproule, T. J., Brown, A. C., Akilesh, S., Jung, N., Petkova, S., Avanessian, L., Choi, E. Y., Shaffer, D. J., Eden, P. A., and Anderson,

C. L. (2003). The MHC class I-like IgG receptor controls perinatal IgG transport, IgG homeostasis, and fate of IgG-Fc-coupled drugs. *Journal of immunology (Baltimore, Md : 1950)*, 170(7):3528–3533.

Roussilhon, C., Oeuvray, C., Müller-Graf, C., Tall, A., Rogier, C., Trape, J.-F., Theisen, M., Balde, A., Pérignon, J.-L., and Druilhe, P. (2007). Long-Term Clinical Protection from Falciparum Malaria Is Strongly Associated with IgG3 Antibodies to Merozoite Surface Protein 3. *PLoS medicine*, 4(11):e320.

Roux, K. H., Strelets, L., and Michaelsen, T. E. (1997). Flexibility of human IgG subclasses. *Journal of immunology (Baltimore, Md : 1950)*, 159(7):3372–3382.

Rudolph, M. G., Stanfield, R. L., and Wilson, I. A. (2006). How TCRs Bind MHCs, Peptides, and Coreceptors. *Annual review of immunology*, 24(1):419–466.

Sakamoto, H., Takeo, S., Maier, A. G., Sattabongkot, J., Cowman, A. F., and Tsuboi, T. (2012). Antibodies against a Plasmodium falciparum antigen PfMSPDBL1 inhibit merozoite invasion into human erythrocytes. *Vaccine*, 30(11):1972–1980.

Sakamoto, N., Shibuya, K., Shimizu, Y., Yotsumoto, K., Miyabayashi, T., Sakano, S., Tsuji, T., Nakayama, E., Nakauchi, H., and Shibuya, A. (2001). A novel Fc receptor for IgA and IgM is expressed on both hemotopoietic and non-hematopoietic tissues. *European journal of immunology*, 31:1310–1316.

Schamberg, N. J. and Lake-Bakaar, G. V. (2007). Hepatitis C Virus-related Mixed Cryoglobulinemia: Pathogenesis, Clinica Manifestations, and New Therapies. *Gastroenterology & Hepatology*, 3(9):695–703.

Schreuder, H., Tardif, C., Trump-Kallmeyer, S., Soffientini, A., Sarubbi, E., Akeson, A., Bowlin, T., Yanofsky, S., and Barrett, R. W. (1997). A new cytokine-receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist. *Nature*, 386(6621):194–200.

Scragg, I. G., Hensmann, M., Bate, C. A. W., and Kwiatkowski, D. (1999). Early cytokine induction by Plasmodium falciparum is not a classical endotoxin-like process. *European journal of immunology*, 29(8):2636–2644.

Seder, R. A., Chang, L.-J., Enama, M. E., Zephir, K. L., Sarwar, U. N., Gordon, I. J., Holman, L. A., James, E. R., Billingsley, P. F., Gunasekera, A., Richman, A., Chakravarty, S., Manoj, A., Velmurugan, S., Li, M., Ruben, A. J., Li, T., Eappen, A. G., Stafford, R. E., Plummer, S. H., Hendel, C. S., Novik, L., Costner, P. J. M., Mendoza, F. H., Saunders, J. G., Nason, M. C., Richardson, J. H., Murphy, J., Davidson, S. A., Richie, T. L., Sedegah, M., Sutamihardja, A., Fahle, G. A., Lyke, K. E., Laurens, M. B., Roederer, M., Tewari, K., Epstein, J. E., Sim, B. K. L., Ledgerwood, J. E., Graham, B. S., Hoffman, S. L., and Team, t. V. . S. (2013). Protection Against Malaria by Intravenous Immunization with a Nonreplicating Sporozoite Vaccine. *Science*, 341(6152):1359–1365.

Seixas, E., Cross, C., Quin, S., and Langhorne, J. (2001). Direct activation of dendritic cells by the malaria parasite, Plasmodium chabaudi chabaudi. *European journal of immunology*, 31(10):2970–2978.

Semblat, J.-P., Ghumra, A., Czajkowsky, D. M., Wallis, R., Mitchell, D. A., Raza, A., and Rowe, J. A. (2015). Identification of the minimal binding region of a Plasmodium falciparum IgM binding PfEMP1 domain. *Molecular and biochemical parasitology*, 201(1):76–82.

Semblat, J.-P., Raza, A., Kyes, S. A., and Rowe, J. A. (2006). Identification of Plasmodium falciparum var1CSA and var2CSA domains that bind IgM natural antibodies. *Molecular and biochemical parasitology*, 146(2):192–197.

Serghides, L., Smith, T. G., Patel, S. N., and Kain, K. C. (2003). CD36 and malaria: friends or foes? *Trends in parasitology*, 19(10):461–469.

Shen, B. W., Spiegel, P. C., Chang, C.-H., Huh, J.-W., Lee, J.-S., Kim, J., Kim, Y.-H., and Stoddard, B. L. (2008). The tertiary structure and domain organization of coagulation factor VIII. *Blood*, 111(3):1240–1247.

Shima, H., Takatsu, H., Fukuda, S., Ohmae, M., Hase, K., Kubagawa, H., Wang, J.-Y., and Ohno, H. (2010). Identification of TOSO/FAIM3 as an Fc receptor for IgM. - PubMed - NCBI. *International Immunology*, 22(3):149–156.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011).

Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1):539–539.

Silacci, M., Baenziger-Tobler, N., Lembke, W., Zha, W., Batey, S., Bertschinger, J., and Grabulovski, D. (2014). Linker Length Matters, Fynomer-Fc Fusion with an Optimized Linker Displaying Picomolar IL-17A Inhibition Potency. *The Journal of biological chemistry*, 289(20):14392–14398.

Simon, A., Asli, B., Braun-Falco, M., De Koning, H., Fermand, J. P., Grattan, C., Krause, K., Lachmann, H., Lenormand, C., Martinez-Taboada, V., Maurer, M., Peters, M., Rizzi, R., Rongioletti, F., Ruzicka, T., Schnitzler, L., Schubert, B., Sibilia, J., and Lipsker, D. (2013). Schnitzler's syndrome: diagnosis, treatment, and follow-up. *Allergy*, 68(5):562–568.

Simossis, V. A. and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Research*, 33(Web Server):W289–W294.

Singh, S., Soe, S., Weisman, S., Barnwell, J. W., Pérignon, J.-L., and Druilhe, P. (2009). A conserved multi-gene family induces cross-reactive antibodies effective in defense against Plasmodium falciparum. *PLoS One*, 4(4):e5410.

Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E., and Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature*, 439(7077):741–744.

Singh, S. K., Singh, A. P., Pandey, S., Yazdani, S. S., Chitnis, C. E., and Sharma, A. (2003). Definition of structural elements in Plasmodium vivax and P. knowlesi Duffy-binding domains necessary for erythrocyte invasion. *Biochemical Journal*, 374(Pt 1):193–198.

Sirima, S. B., Cousens, S., and Druilhe, P. (2011). Protection against malaria by MSP3 candidate vaccine. *The New England journal of medicine*, 365(11):1062–1064.

Smith, J. D., Craig, A. G., Kriek, N., Hudson-Taylor, D., Kyes, S., Fagan, T., Fagen, T., Pinches, R., Baruch, D. I., Newbold, C. I., and Miller, L. H. (2000). Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite

adhesion trait implicated in cerebral malaria. *Proceedings of the National Academy of Sciences*, 97(4):1766–1771.

Smith, J. D., Kyes, S., Craig, A. G., Fagan, T., Hudson-Taylor, D., Miller, L. H., Baruch, D., and Newbold, C. (1998). Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. *Molecular and biochemical parasitology*, 97(1-2):133–148.

Smith, J. D., Rowe, J. A., Higgins, M. K., and Lavstsen, T. (2013). Malaria's Deadly Grip: Cytoadhesion of Plasmodium falciparum Infected Erythrocytes. *Cellular microbiology*, 15(12):1976–1983.

Sola, R. J. and Griebenow, K. (2009). Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*, 98(4):1223–1245.

Sørensen, V., Sundvold, V., Michaelsen, T. E., and Sandlie, I. (1999). Polymerization of IgA and IgM: roles of Cys309/Cys414 and the secretory tailpiece. *Journal of immunology (Baltimore, Md : 1950)*, 162(6):3448–3455.

Springer, A. L., Smith, L. M., Mackay, D. Q., Nelson, S. O., and Smith, J. D. (2004). Functional interdependence of the DBLbeta domain and c2 region for binding of the Plasmodium falciparum variant antigen to ICAM-1. *Molecular and biochemical parasitology*, 137(1):55–64.

Stevenson, L., Huda, P., Jeppesen, A., Laursen, E., Rowe, J. A., Craig, A., Streicher, W., Barfod, L., and Hviid, L. (2015a). Investigating the function of Fc-specific binding of IgM to Plasmodium falciparum erythrocyte membrane protein 1 mediating erythrocyte rosetting. *Cellular microbiology*, 17(6):819–831.

Stevenson, L., Laursen, E., Cowan, G. J., Bandoh, B., Barfod, L., Cavanagh, D. R., Andersen, G. R., and Hviid, L. (2015b). α2-Macroglobulin Can Crosslink Multiple Plasmodium falciparum Erythrocyte Membrane Protein 1 (PfEMP1) Molecules and May Facilitate Adhesion of Parasitized Erythrocytes. *PLoS pathogens*, 11(7):e1005022.

Stevenson, M. M. and Riley, E. M. (2004). Innate immunity to malaria. *Nature reviews Immunology*, 4(3):169–180.

Stone, K. D., Prussin, C., and Metcalfe, D. D. (2010). IgE, mast cells, basophils, and eosinophils. *Journal of Allergy and Clinical Immunology*, 125(2):S73–S80.

Story, C. M., Mikulska, J. E., and Simister, N. E. (1994). A major histocompatibility complex class I-like Fc receptor cloned from human placenta: possible role in transfer of immunoglobulin G from mother to fetus. *The Journal of experimental medicine*, 180(6):2377–2381.

Stowers, A. W., Zhang, Y., Shimp, R. L., and Kaslow, D. C. (2001). Structural conformers produced during malaria vaccine production in yeast. *onlinelibrary.wiley.com*, 18:137–150.

Strohl, W. R. (2015). Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. *Biodrugs*, 29(4):215–239.

Strunk, R. C. and Bloomberg, G. R. (2006). Omalizumab for Asthma. *The New England journal of medicine*, 354(25):2689–2695.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12):2725–2729.

Terrier, B., Krastinova, E., Marie, I., Launay, D., Lacraz, A., Belenotti, P., de Saint-Martin, L., Quemeneur, T., Huart, A., Bonnet, F., Le Guenno, G., Kahn, J. E., Hinschberger, O., Rullier, P., Diot, E., Lazaro, E., Bridoux, F., Zenone, T., Carrat, F., Hermine, O., Leger, J. M., Mariette, X., Senet, P., Plaisier, E., and Cacoub, P. (2012). Management of noninfectious mixed cryoglobulinemia vasculitis: data from 242 cases included in the CryoVas survey. *Blood*, 119(25):5996–6004.

Tetteh, K. K. A., Osier, F. H. A., Salanti, A., Kamuyu, G., Drought, L., Failly, M., Martin, C., Marsh, K., and Conway, D. J. (2013). Analysis of Antibodies to Newly Described Plasmodium falciparum Merozoite Antigens Supports MSPDBL2 as a Predicted Target of Naturally Acquired Immunity. *Infection and immunity*, 81(10):3835–3842.

Tissot, J.-D., Sanchez, J.-C., Vuadens, F., Scherl, A., Schifferli, J. A., Hochstrasser, D. F., Schneider, P., and Duchosal, M. A. (2002). IgM are associated to Sp alpha (CD5 antigen-like). *Electrophoresis*, 23(7-8):1203–1206.

Tolia, N. H., enemark, e. J., Sim, B. K. L., and Joshua-Tor, L. (2005). Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. *Cell*, 3(10):746–746.

Treon, S. P. (2009). How I treat Waldenström macroglobulinemia. *Blood*, 114(12):2375–2385.

Treon, S. P., Branagan, A. R., Hunter, Z., Santos, D., Tournhilac, O., and Anderson, K. C. (2004). Paradoxical increases in serum IgM and viscosity levels following rituximab in Waldenstrom's macroglobulinemia. *Annals of Oncology*, 15(10):1481–1483.

Tsai, C. W., Duggan, P. F., Shimp, R. L., Miller, L. H., and Narum, D. L. (2006). Overproduction of Pichia pastoris or Plasmodium falciparum protein disulfide isomerase affects expression, folding and O-linked glycosylation of a malaria vaccine candidate expressed in P. pastoris. *Journal of biotechnology*, 121(4):458–470.

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.-T., Otsuki, H., Kaneko, O., Sattabongkot, J., Udomsangpetch, R., Sawasaki, T., Torii, M., and Endo, Y. (2008). Wheat Germ Cell-Free System-Based Production of Malaria Proteins for Discovery of Novel Vaccine Candidates. *Infection and immunity*, 76(4):1702–1708.

Urban, B. C., Ferguson, D. J. P., Pain, A., Willcox, N., Plebanski, M., Austyn, J. M., and Roberts, D. J. (1999). Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature*, 400(6739):73–77.

van Bemmelen, M. X., Beghdadi-Rais, C., Desponds, C., Vargas, E., Herrera, S., Reymond, C. D., and Fasel, N. (2000). Expression and one-step purification of Plasmodium proteins in Dictyostelium. *Molecular and biochemical parasitology*, 111(2):377–390.

VanBuskirk, K. M., Sevova, E., and Adams, J. H. (2004). Conserved residues in the Plasmodium vivax Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. *Proceedings of the National Academy of Sciences*, 101(44):15754–15759.

Vercammen, M., Scorza, T., El Bouhdidi, A., Van Beeck, K., Carlier, Y., Dubremetz, J. F., and Verschueren, H. (1999). Opsonization of Toxoplasma gondii tachyzoites

with nonspecific immunoglobulins promotes their phagocytosis by macrophages and inhibits their proliferation in nonphagocytic cells in tissue culture. *Parasite immunology*, 21(11):555–563.

Vidarsson, G., Dekkers, G., and Rispens, T. (2014). IgG subclasses and allotypes: from structure to effector functions. *Frontiers in immunology*, 5:520.

Vidarsson, G., Stemerding, A. M., Stapleton, N. M., Spliethoff, S. E., Janssen, H., Rebers, F. E., de Haas, M., and van de Winkel, J. G. (2006). FcRn: an IgG receptor on phagocytes with a novel role in phagocytosis. *Blood*, 108(10):3573–3579.

Vigan-Womas, I., Guillotte, M., Juillerat, A., Hessel, A., Raynal, B., England, P., Cohen, J. H., Bertrand, O., Peyrard, T., Bentley, G. A., Lewit-Bentley, A., and Mercereau-Puijalon, O. (2012). Structural Basis for the ABO Blood-Group Dependence of Plasmodium falciparum Rosetting. *PLoS pathogens*, 8(7):e1002781.

Vincendeau, P. and Daëron, M. (1989). Trypanosoma musculi co-express several receptors binding rodent IgM, IgE, and IgG subclasses. *Journal of immunology (Baltimore, Md : 1950)*, 142(5):1702–1709.

Vladutiu, A. O. (2000). Immunoglobulin D: properties, measurement, and clinical relevance. *Clinical and Diagnostic Laboratory Immunology*, 7(2):131–140.

Vogt, A. M., Barragan, A., Chen, Q., Kironde, F., Spillmann, D., and Wahlgren, M. (2003). Heparan sulfate on endothelial cells mediates the binding of Plasmodium falciparum-infected erythrocytes via the DBL1alpha domain of PfEMP1. *Blood*, 101(6):2405–2411.

Wan, T., Beavil, R. L., Fabiane, S. M., Beavil, A. J., Sohi, M. K., Keown, M., Young, R. J., Henry, A. J., Owens, R. J., Gould, H. J., and Sutton, B. J. (2002). The crystal structure of IgE Fc reveals an asymmetrically bent conformation. *Nature immunology*, 3(7):681–686.

Watkins, J. F. (1964). Adsorption of Sensitized Sheep Erythrocytes to HeLa Cells Infected with Herpes Simplex Virus. *Nature*, 202:1364–1365.

WHO (2015). WHO — World Malaria Report 2015. Technical report.

Wickramarachchi, T., Cabrera, A. L., Sinha, D., Dhawan, S., Chandran, T., Devi, Y. S., Kono, M., Spielmann, T., Gilberger, T. W., Chauhan, V. S., and Mohmmed, A. (2009). A novel Plasmodium falciparum erythrocyte binding protein associated with the merozoite surface, PfDBLMSP. *International Journal For Parasitology*, 39(7):763–773.

Wiersma, E. J., Collins, C., Fazel, S., and Shulman, M. J. (1998). Structural and functional analysis of J chain-deficient IgM. *Journal of immunology (Baltimore, Md : 1950)*, 160(12):5979–5989.

Wilson, T. J., Fuchs, A., and Colonna, M. (2012). Cutting Edge: Human FcRL4 and FcRL5 Are Receptors for IgA and IgG. *Journal of immunology (Baltimore, Md : 1950)*, 188(10):1102651–4745.

Wines, B. D., Powell, M. S., Parren, P. W., Barnes, N., and Hogarth, P. M. (2000). The IgG Fc contains distinct Fc receptor (FcR) binding sites: the leukocyte receptors Fc gamma RI and Fc gamma RIIa bind to a region in the Fc distinct from that recognized by neonatal FcR and protein A. *Journal of immunology (Baltimore, Md : 1950)*, 164(10):5313–5318.

Woof, J. M. (2002). The human IgA-Fc α receptor interaction and its blockade by streptococcal IgA-binding proteins. *Biochemical Society transactions*, 30(4):491–494.

Woof, J. M. and Russell, M. W. (2011). Structure and function relationships in IgA. *Mucosal Immunology*, 4(6):590–597.

Worth, C. L., Preissner, R., and Blundell, T. L. (2011). SDM–a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Research*, 39(suppl):W215–W222.

Yan, B., Yates, Z., Balland, A., and Kleemann, G. R. (2009). Human IgG1 hinge fragmentation as the result of H2O2-mediated radical cleavage. *The Journal of biological chemistry*, 284(51):35390–35402.

Yewhalaw, D., Wassie, F., Steurbaut, W., and Spanoghe, P. (2011). Multiple insecticide resistance: an impediment to insecticide-based malaria vector control program. *PLoS One*.

You, L. and Arnold, F. H. (1996). Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering*, 9(1):77–83.

Zhang, J., Yun, J., Shang, Z., Zhang, X., and Pan, B. (2009). Design and optimization of a linker for fusion protein construction. *Progress in Natural Science*, 19(10):1197–1200.

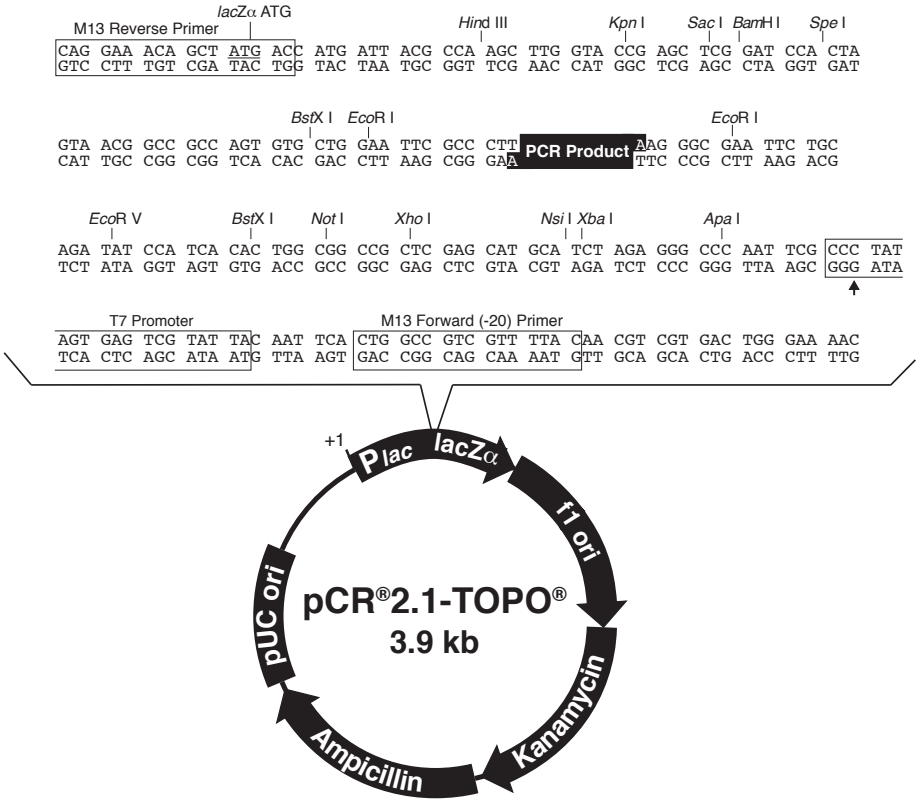
Zheng, K., Bantog, C., and Bayer, R. (2014). The impact of glycosylation on monoclonal antibody conformation and stability. *mAbs*, 3(6):568–576.

Appendices

A Plasmid Maps

A.1 pCR2.1-TOPO

The commercially available PCR 2.1 TOPO vector from Life Technologies was used as a parking vector to insert the DBL 3.4 and DBL 3.8 domains, forming the vectors PCR 2.1-TOPO 3.4 and PCR 2.1-TOPO 3.8. The cloned TOPO vectors were then used for sub-cloning the DBL domains into the pFUSE vector. The TOPO vector was used as a parking vector due to its cloning efficiency.



Comments for pCR®2.1-TOPO®
3931 nucleotides

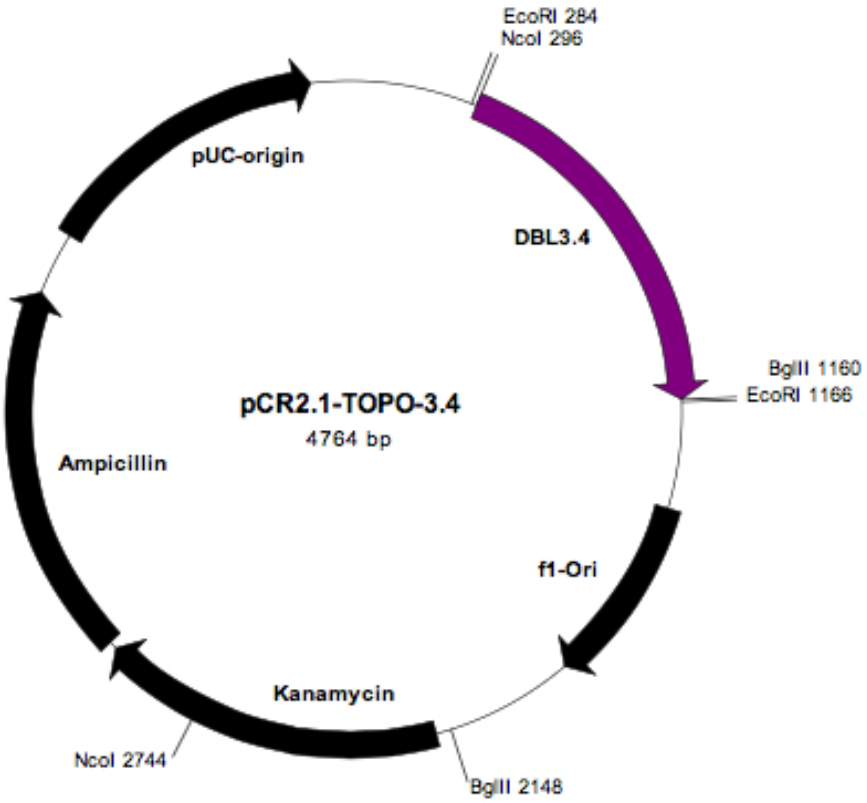
*LacZ*α fragment: bases 1-547
M13 reverse priming site: bases 205-221
Multiple cloning site: bases 234-357
T7 promoter/priming site: bases 364-383
M13 Forward (-20) priming site: bases 391-406
f1 origin: bases 548-985
Kanamycin resistance ORF: bases 1319-2113
Ampicillin resistance ORF: bases 2131-2991
pUC origin: bases 3136-3809

pCR2.1-TOPO

AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCG
GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTT
GTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCA
CTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGC
ATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTG
GCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCC
CAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTG
ACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCA
AGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTT
CACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGA
GCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAG
TCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGG
TAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCC
TCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGA
TCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTAT
TCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTT
GTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTG
CGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCCC
ACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGAC
CACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCA
GGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCT
GCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGAC
ATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCC
CGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG
TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAG
ATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTT
CCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCAT
ACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCA
GTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAC
GATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACT
GGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGT
GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCA
GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT
CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA
ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA
AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTC
AAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG

A.2 pCR2.1-TOPO-DBL3.4-short

Plasmid map for the pCR2.1-TOPO-3.4 vector containing the DBL 3.4 insert from MSPDBL1 (shown in purple). The 3.4 DBL domain has been inserted between NcoI restriction site at 296 bases and the BglII restriction site at 1160 bases. The sequence shows the DBL insert in purple with the flanking restriction sites in red. Neither NcoI (CCATGG) not BglII (AGATCT) are present in the insert sequence, but are both present in the TOPO plasmid, their positions being shown in yellow.

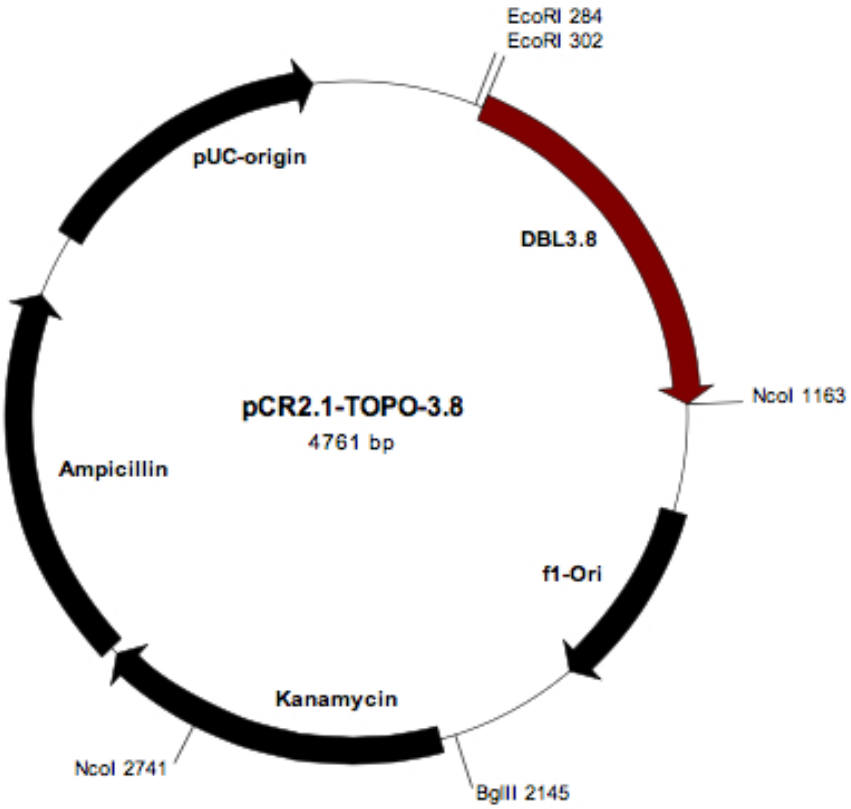


pCR2.1-TOPO-DBL3.4

AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCG
GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTT
GTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCA
CTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCCATGGTTttctgcaagggcatcaagaacgtgctgagctgccctcccaag
aacagcgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccag
acggaccaagctgtgcctgcggaacatcaacaaagtgtggcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaagg
tcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaacctgaacgccctgaccgccattaagtacggcttcagc
gacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaagatcctgag
aaacgaggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcg
gctacaaggtgcacatcggcaacaagccctgccccgagcacgacaacatggaccggatccccagtacctgcggtggtttcgcgag
tggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatcaagctgtgcaacatccagcagttcaccaaccagga
cgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggcggaggc
ccgagtggaagggccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctg
aaagaaatctgcagcgagtgcgactgcaagtacaaggacctggacaacAGATCTAAGGGCGAATTCTGCAGATATCCATCACACTG
GCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTC
GTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCC
CGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTG
GTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTT
CGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC
TTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAAT
AGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGC
CTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATTCAGGGCGCAAGGGCTGCTAAAGGAAGCG
GAACACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAA
GCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAA
TTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCG
CAGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCG
CTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAG
GGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGC
CACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGC
AGGATCTCCTGTCATCCCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCG
GCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGA
TCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCG
TGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTG
GCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCT
TTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGA
AAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTT
CGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGA
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA
CAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC
GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCT
GATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAC
TGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGC
TACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCAC
CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTG
TCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCT
TGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGT
CGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATT
ACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG

A.3 pCR2.1-TOPO-DBL3.8-short

Plasmid map for the pCR2.1-TOPO-3.8 vector containing the DBL 3.8 insert from MSPDBL2 (shown in maroon). The 3.8 DBL domain has been inserted between EcoRI restriction site at 302 bases and the NcoI restriction site at 1163 bases. The sequence shows the DBL insert in maroon with the flanking restriction sites in red. Neither EcoRI (GAATTC) nor NcoI (CCATGG) are present in the insert sequence but both are present within the TOPO plasmid, their positions being shown in yellow.

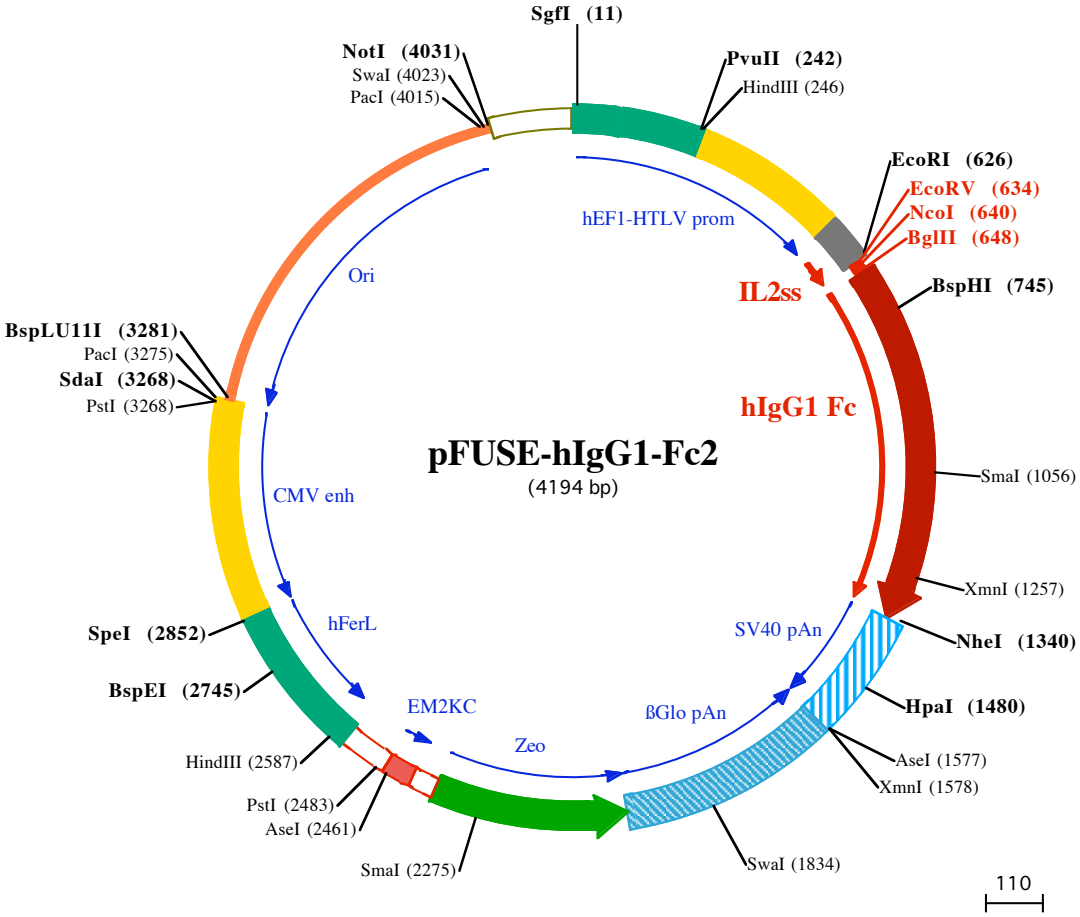


pCR2.1-TOPO-DBL3.8

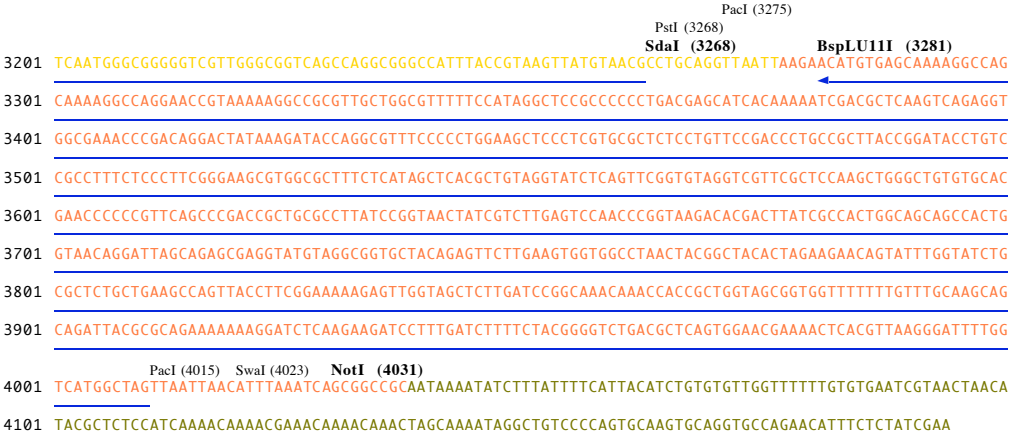
AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCG
GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTT
GTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCA
CTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGAATTCGATCTGCAAGGACTTCAGCAACCTGCCCCAGTGCCGGAAGAACG
TGCACGAGCGGAACAACTGGCTGGGCAGCAGCGTGAAGAACTTCGCCAGCGACAACAAGGGCGTGCTGGTGCCCCCCAGACGGCAG
AGCCTGTGCCTGAGAATCACCCTGCAGGACTTCCGGACCAAGAAGAAGAAAGAGGGCGACTTCGAGAAGTTCATCTACAGCTACGC
CAGCAGCGAGGCCCGGAAGCTGCGGACCATCCACAACAACAACCTGGAAAAGGCCCACCAGGCCATCCGGTACAGCTTCGCCGACA
TCGGCAACATCATCCGGGGCGACGACATGATGGACACCCCCACCAGCAAAGAGACAATCACCTATCTGGAAAAGGTGCTGAAGATC
TACAATGAGAACAACGACAAGCCCAAGGACGCCAAGAAGTGGTGGACCGAGAACCGGCACCACGTGTGGGAGGCCATGATGTGCGG
CTACCAGAGCGCCCAGAAAGACAACCAGTGCACCGGCTACGGCAACATCGACGATATCCCCCAGTTCCTGCGGTGGTTCCGCGAGT
GGGGCACCTACGTGTGCGAAGAGTCCGAGAAGAACATGAATACCCTGAAGGCCGTGTGCTTCCCCAAGCAGCCCAGAACCGAGGCC
AACCCTGCCCTGACCGTGCACGAGAACGAGATGTGCAGCAGCACCCTGAAGAAATACGAGGAATGGTACAACAAGCGCAAGACCGA
GTGGACCGAGCAGAGCATCAAGTATAACAACGATAAGATCAACTACGCCGACATCAAGACCCTGAGCCCCAGCGAGTACCTGATCG
AGAAGTGCCCCGAGTGCAAGTGCACCAAGAAAAACCTGCAGGACGCCATGGAAGGGCGAATTCTGCAGATATCCATCACACTGGCG
GCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTG
ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGC
ACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGC
CGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTA
TTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAA
CACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCG
CAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTG
CCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAG
GGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTT
GGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGG
CGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCAC
GACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGG
ATCTCCTGTCATCCCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCT
ACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCT
GGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGA
CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCG
GACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTA
CGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTGAAAAAGGAAGAGTATGAG
TATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAG
TAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCA
ACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAG
TAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAG
CTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGA
CGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGC
AACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGAT
AAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTA
CACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGT
CAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGAT
AATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCCTTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG
TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCT
TTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACC
GCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG

A.4 pFUSE-hIgG1-Fc2

The commercially available pFUSE-hIgG1-Fc2 vector from Invivogen was used as an expression vector.

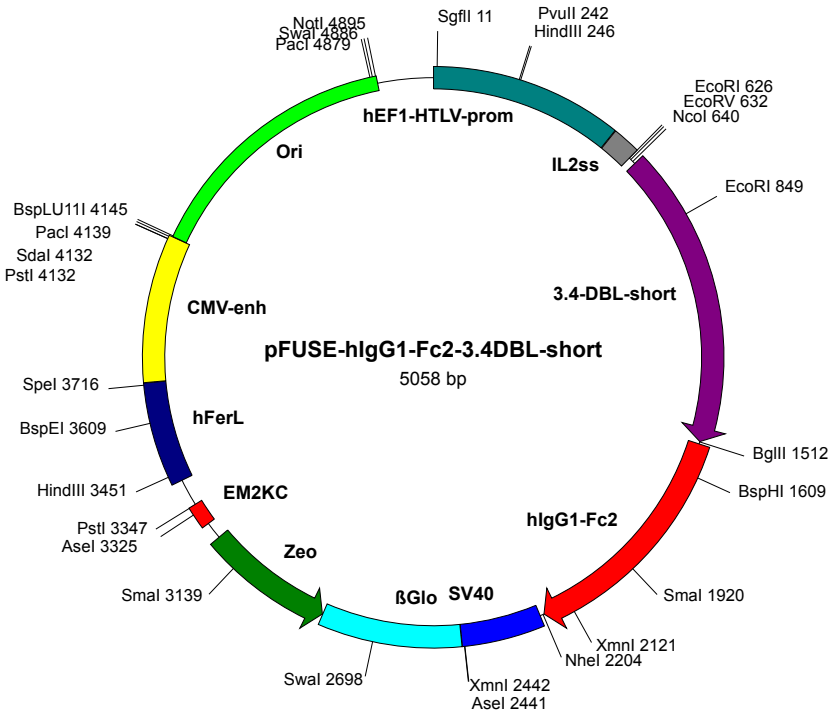



SgfI (11)
HindIII (246)
PvuII (242)
EcoRV (634)
BglII (648)
EcoRI (626)
NcoI (640)
BspHI (745)
SmaI (1056)
XmnI (1257)
NheI (1340)
HpaI (1480)
AseI (1577)
XmnI (1578)
SwaI (1834)
SmaI (2275)
AseI (2461)
PstI (2483)
HindIII (2587)
BspEI (2745)
SpeI (2852)


PacI (3275)
PstI (3268)
SdaI (3268)
BspLU11I (3281)
3201 TCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAG
3301 CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
3401 GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
3501 CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCAC
3601 GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
3701 GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
3801 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
3901 CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
PacI (4015)
SwaI (4023)
NotI (4031)
4001 TCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACA
4101 TACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

A.5 pFUSE-hIgG1-Fc2-DBL3.4-short

Plasmid map for the pFUSE-hIgG1-Fc2-DBL3.4-short vector containing the DBL 3.4 insert from MSPDBL1 (shown in purple). The 3.4 DBL domain has been inserted within the multiple cloning site between the NcoI restriction site at 640 bases and the BglII restriction site at 1512 bases. The sequence shows the DBL insert in purple in the pFUSE plasmid. The DBL is located before the start of the Fc sequence (which begins GACAAAA in dark red).



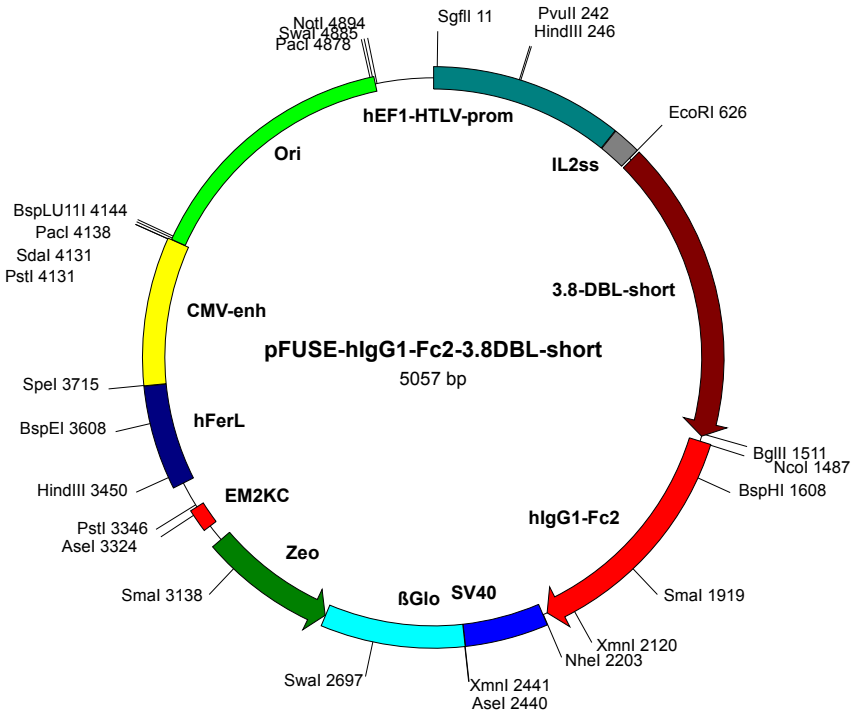
pFUSE-hIgG1-Fc2-3.4DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAATTCGATATCGGCCATGGTTttctgcaagggcatcaagaacgtgctgagctgccctcccaaga
acagcgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccagacggaccaagctgtgcctgcggaacatcaa
caaagtgtggcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaacctgaacgcc
ctgaccgccattaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaagatcctgagaa
acgaggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaacaagccctg
ccccgagcacgacaacatggacccggatccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatcaagctgtgc
aacatccagcagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggcggaggcccg
agtggaagggccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcgactgcaagta
caaggacctggacaacAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTC
ATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAA
AGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCT
CCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACC
TGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCT
TCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCT
GTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATG
CTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTA
AAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCA
TCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTT
AAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCC
CTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCA
CAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGAC
CTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGG
ACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCA
ACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAA
ACTAGGGCTGCAgggttcatagtgccactttcctgcactgccccatctcctgcccacccttccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAG
AAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAA
TCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTG
TTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCC
ATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGC
CATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCC
CTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATT
AAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTT
TCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC
CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
CAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTT
GTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTNSISAMVFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNI
NKVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDKILR
NEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKNKFEDVIKLC
NIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKY
KDLDNRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTK
PREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTC
LVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLS
PGK*VLA

A.6 pFUSE-hIgG1-Fc2-DBL3.8-short

Plasmid map for the pFUSE-hIgG1-Fc2-DBL3.8-short vector containing the DBL 3.8 insert from MSPDBL2 (shown in maroon). The 3.8 DBL domain has been inserted within the multiple cloning site between the EcoRI restriction site at 626 bases and the NcoI restriction site at 1487 bases. The sequence shows the DBL insert in maroon in the pFUSE plasmid. The DBL is located before the start of the Fc sequence (which begins GACAAA in dark red).

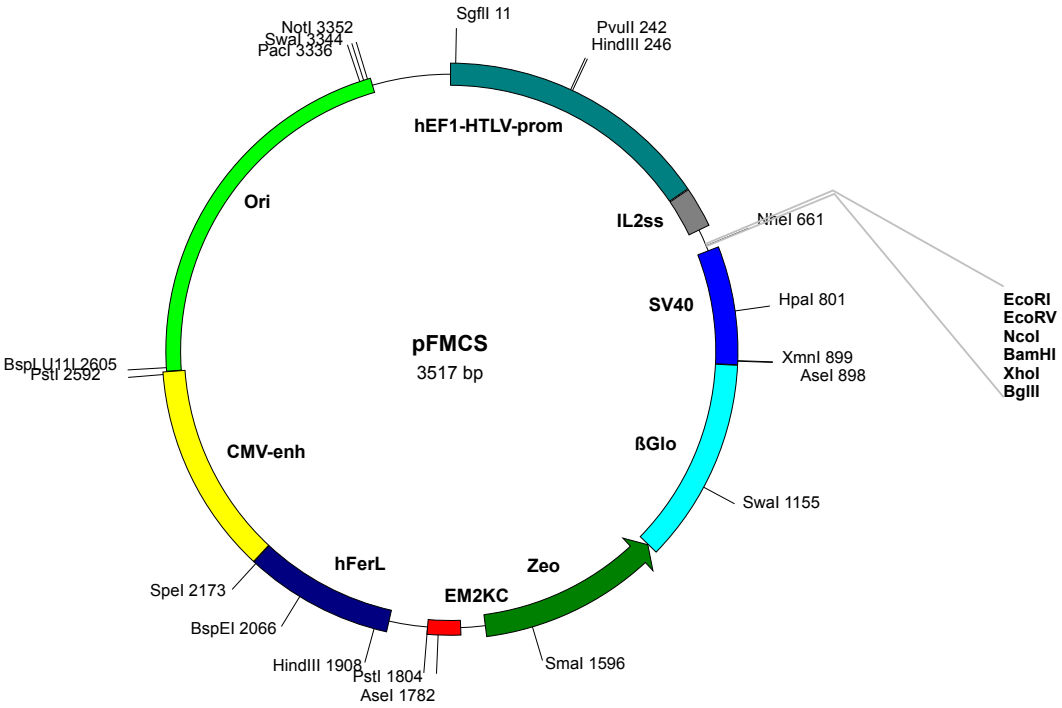


pFUSE-hIgG1-Fc2-38DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAATTCGatctgcaaggacttcagcaacctgccccagtgccggaagaacgtgcacgagcggaaca
actggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgcccccagacggcagagcctgtgcctgagaatcaccctgcaggacttccggaccaagaa
gaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcccaccaggccatccggtac
agcttcgccgacatcggcaacatcatccggggcgacgacatgatggacacccccaccagcaaagagacaatcacctatctggaaaaggtgctgaagatctacaatgagaacaacg
acaagcccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaaccagtgcaccggctacggcaa
catcgacgatatcccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaatacccgaaggccgtgtgcttcccccaagcagccc
agaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccgagtggaccgagcagagca
tcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacctgatcgagaagtgccccgagtgcaagtgcaccaagaaaaacctgcaggacgC
CATGGTTAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCAAAACCCAAGGACACCCTCATGATCTCC
CGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGG
AGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCC
CATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTC
AAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACA
GCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGG
TAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTT
TATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTA
CAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCT
GTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACT
GACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAAT
ATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCA
CGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCA
CTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCG
GCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCA
TGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCT
GCAgggttcatagtgccacttttcctgcactgccccatctcctgcccacccctttcccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAG
ACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTG
GCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAAT
GGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTA
CTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCG
TCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCG
TTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATG
TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTC
GGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACA
AACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC
GAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATC
GTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFR
TKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIY
NENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAV
CFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKC
TKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVE
VHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEM
TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALH
NHYTQKSLSLSPGK*VLA

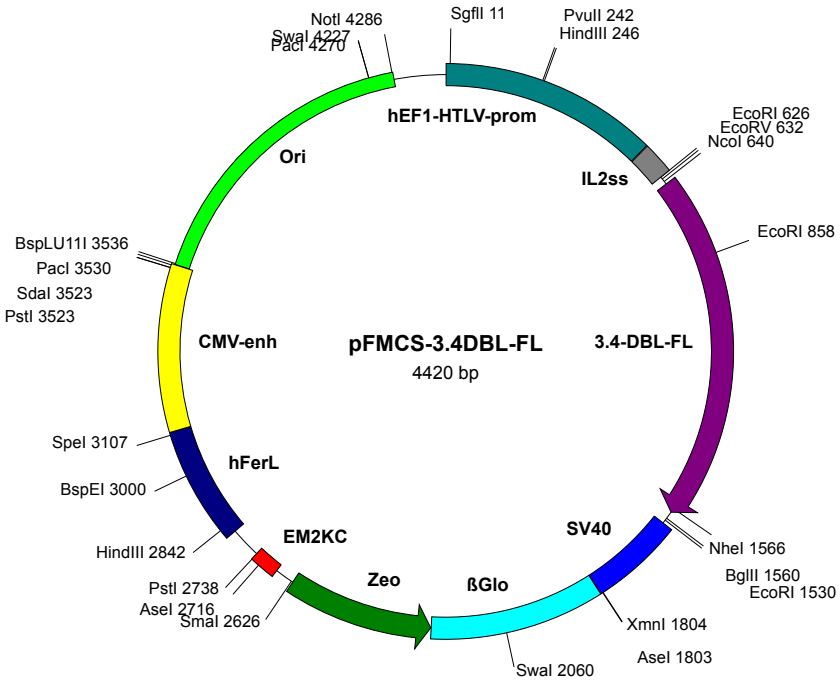
A.7 pFMCS


NotI 3352
SwaI 3344
PacI 3336
SgfII 11
PvuII 242
HindIII 246
hEF1-HTLV-prom
Ori
IL2ss
NheI 661
SV40
HpaI 801
pFMCS
3517 bp
XmnI 899
AseI 898
EcoRI
EcoRV
NcoI
BamHI
XhoI
BglII
BspLU11I 2605
PstI 2592
ßGlo
SwaI 1155
CMV-enh
hFerL
Zeo
SpeI 2173
EM2KC
BspEI 2066
HindIII 1908
PstI 1804
AseI 1782
SmaI 1596

pFMCS

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCCCATGGGGATCCCTCGAGAGATCTGCTAGCTGGCCAGACATGATAAGATACA
TTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAA
CAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATA
GCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTT
CTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAA
ATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGG
AACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCG
CCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACC
CAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGA
GCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaag
gttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgcccca
tctcctgcccaccctttcccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTG
GCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCC
GGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAA
CAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACT
AATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGAT
ACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGG
GCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAA
AGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGG
TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAA
GACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAG
CAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTT
AATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAA
CAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

A.8 pFMCS-3.4DBL-FL

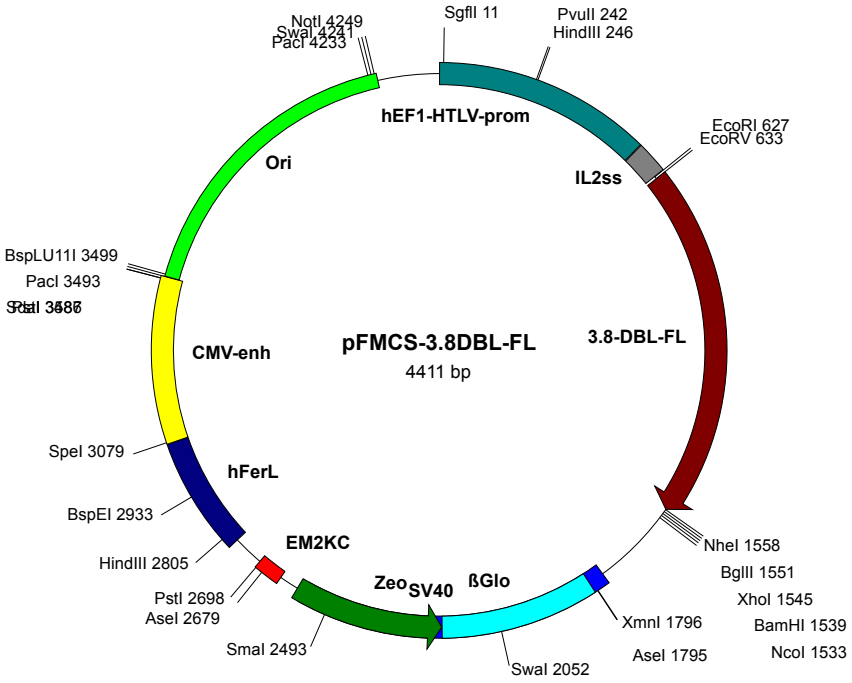


pFMCS-34DBL-FL

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCaagtgcccgacgagaacttctgcaagggcatcaagaacgtgctgagctgcc
ctcccaagaacagcgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccagacggaccaagctgtgcctgcg
gaacatcaacaaagtgtggcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaac
ctgaacgccctgaccgccattaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaaga
tcctgagaaacgaggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaa
caagccctgccccgagcacgacaacatggaccggatccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatc
aagctgtgcaacatccagcagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggc
ggaggcccgagtggaagggccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcga
ctgcaagtacaaggacctggacaacaccttcaaagaattcaaggacCCATGGGGATCCCTCGAGAGATCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAA
ACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTC
ATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCC
AAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAG
ATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCA
TTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATT
GGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGC
CGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTC
CGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAAC
TCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTAT
AGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttt
cccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTT
TATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCT
CAGCCCCCCGCCCCAAAGCAAGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGT
CAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTG
CCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCC
AAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCG
GTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCA
CTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCA
GCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAA
ATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIKCPDENFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKL
CLRNINKVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINR
ALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKN
KFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYL
KEICSECDCKYKDLDNTFKEFKDPWGSLERSASW

A.9 pFMCS-3.8DBL-FL

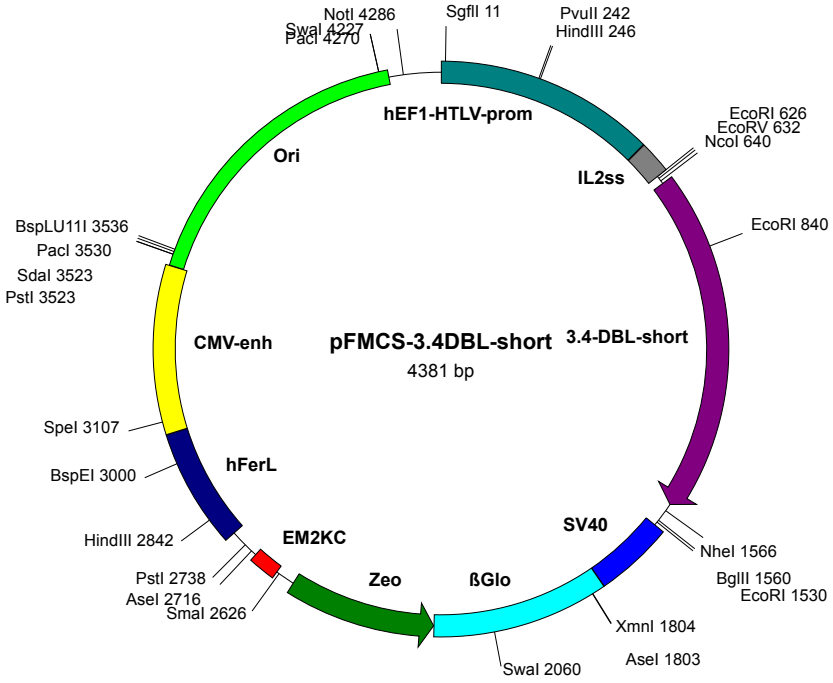


pFMCS-38DBL-FL

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCaagtgccccaccgaggaaatctgcaaggacttcagcaacctgccccagtgccggaaga
acgtgcacgagcggaacaactggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccagacggcagagcctgtgcctgagaatcaccctgca
ggacttccggaccaagaagaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcc
caccaggccatccggtacagcttcgccgacatcggcaacatcatccggggcgacgacatgatggacaccccaccagcaaagagacaatcacctatctggaaaaggtgctgaaga
tctacaatgagaacaacgacaagccccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaacca
gtgcaccggctacggcaacatcgacgatatcccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaatacctgaaggccgtg
tgcttccccaagcagcccagaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccg
agtggaccgagcagagcatcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacctgatcgagaagtgccccgagtgcaagtgcaccaa
gaaaaacctgcaggacgtgttcgagctgaccttcgacCCATGGGGATCCCTCGAGAGATCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACT
AGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGT
TTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGC
CTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGT
ATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGA
AAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAA
GAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGG
TCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCAC
CTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCT
CCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTG
TATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcat
agacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCG
CCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCC
GCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGT
GGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGG
AAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCA
GTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAG
GCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC
TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCA
ATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGT
CCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFKCPTEEICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRI
TLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLE
KVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNM
NTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEK
CPECKCTKKNLQDVFELTFDPWGSLERSASW

A.10 pFMCS-3.4DBL-short

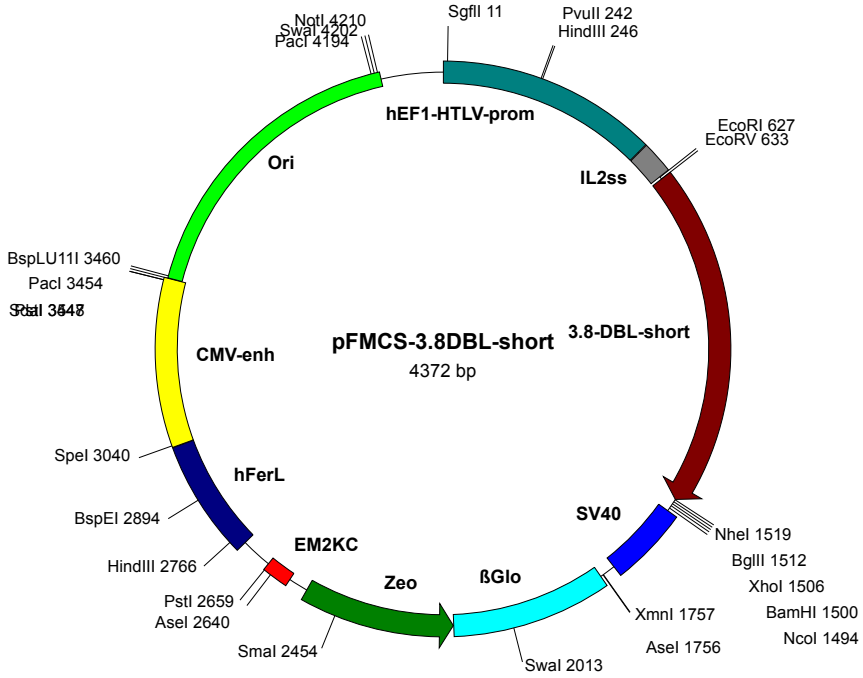


pFMCS-34DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCttctgcaagggcatcaagaacgtgctgagctgccctcccaagaacagcgccg
gcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccagacggaccaagctgtgcctgcggaacatcaacaaagtgtg
gcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaacctgaacgccctgaccgcc
attaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaagatcctgagaaacgaggcca
gcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaacaagccctgccccgagca
cgacaacatggaccggatccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatcaagctgtgcaacatccag
cagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggcggaggcccgagtggaagg
gccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcgactgcaagtacaaggacct
ggacaacCCATGGGGATCCCTCGAGAGATCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTT
TTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAA
TAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCAT
TTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCA
GATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCC
TGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGT
TCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGT
CACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGA
ACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAAT
GATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcatagacagtcagtgacttacCAAACTCACAGG
AGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAG
GCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTG
TAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACC
GCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATA
ATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGT
CAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGC
CTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA
AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGT
GTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATT
TCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNIN
KVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDKIL
RNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKNKFEDVI
KLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSE
CDCKYKDLDNPWGSLERSASW

A.11 pFMCS-3.8DBL-short

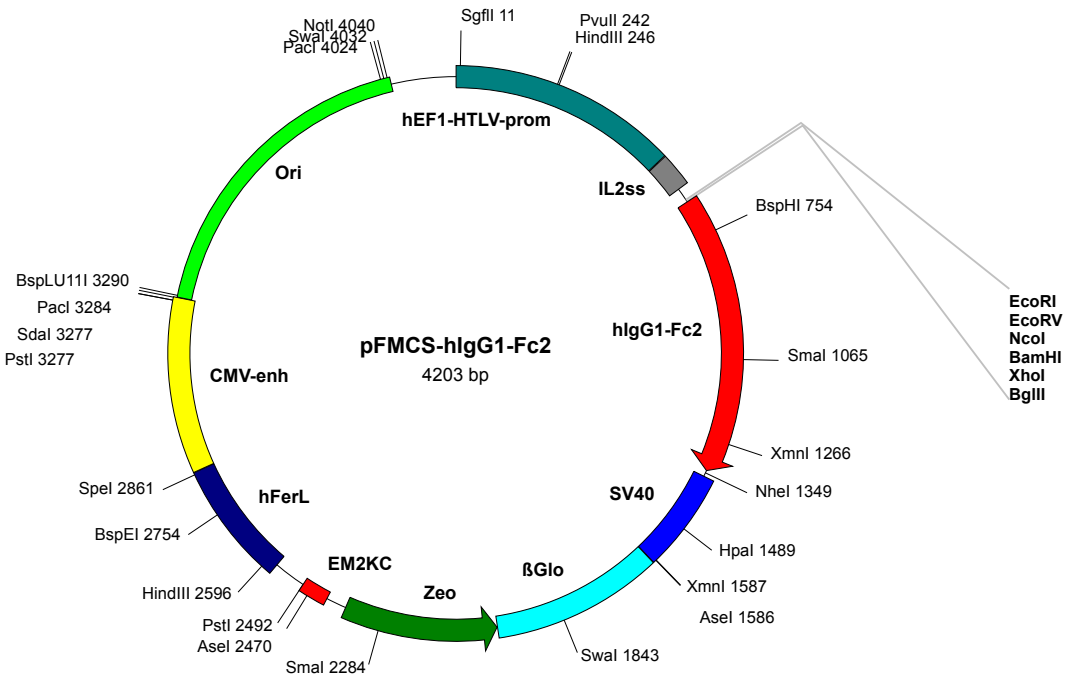


pFMCS-38DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCatctgcaaggacttcagcaacctgccccagtgccggaagaacgtgcacgagcggaaca
actggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccagacggcagagcctgtgcctgagaatcaccctgcaggacttccggaccaagaa
gaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcccaccaggccatccggtac
agcttcgccgacatcggcaacatcatccggggcgacgacatgatggacaccccaccagcaaagagacaatcacctatctggaaaaggtgctgaagatctacaatgagaacaacg
acaagcccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaaccagtgcaccggctacggcaa
catcgacgatatcccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaatacccctgaaggccgtgtgcttcccccaagcagccc
agaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccgagtggaccgagcagagca
tcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacctgatcgagaagtgcccgagtgcaagtgcaccaagaaaaacctgcaggacCC
ATGGGGATCCCTCGAGAGATCTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGT
GATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATA
GGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATG
TTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAA
GGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCT
GCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACA
CGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTC
CCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTG
GTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTG
TCAAACTAGGGCTGCAgggttcatagtgccactttcctgcactgccccatctcctgcccacccttccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAG
GCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGG
CCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAG
CGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCAC
GCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGC
GGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAA
GTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTT
AATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG
CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTT
TTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCG
AA

MYRMQLLSCIALSLALVTEFICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFR
TKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIY
NENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAV
CFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKC
TKKNLQDPWGSLERSASW

A.12 pFMCS-hIgG1-Fc2

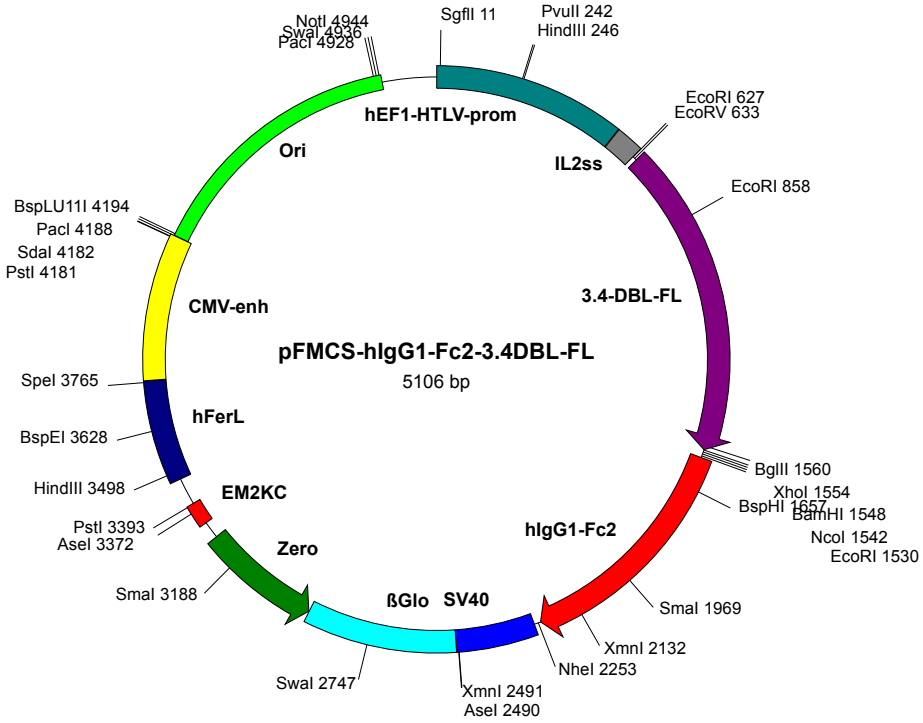


pFMCS-IgG1

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCCCATGGGGATCCCTCGAGAGATCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCA
CGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTC
ACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAG
AACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAG
CAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAAC
GTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGA
TGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAAC
AACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCA
AAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTT
TCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATA
ATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAAC
CTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCC
CCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAG
GCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCC
GGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggtt
agtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatct
cctgcccaccctttcccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCT
AGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGA
GCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAA
ACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAAT
ACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACA
CTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCG
GGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC
CCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTAT
CTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAAT
TAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAA
AACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIPWGSLERSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS
HEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPR
EPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGN
VFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.13 pFMCS-hIgG1-Fc2-3.4DBL-FL

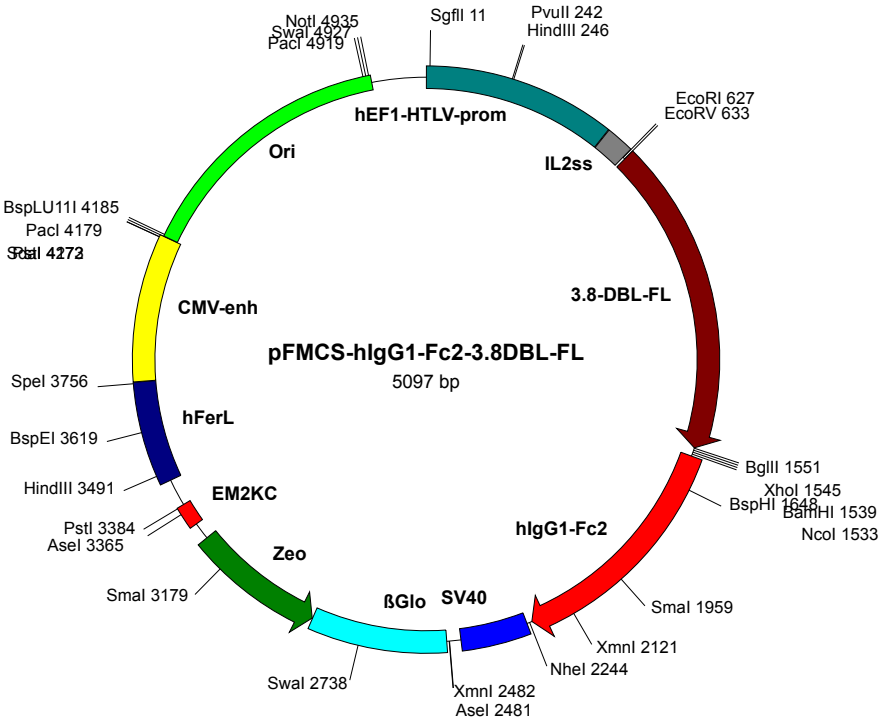


pFMCS-hIgG1-Fc2-34DBL-FL

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCaagtgccccgacgagaacttctgcaagggcatcaagaacgtgctgagctgcc
ctcccaagaacagcgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccagacggaccaagctgtgcctgcg
gaacatcaacaaagtgtggcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaac
ctgaacgccctgaccgccattaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaaga
tcctgagaaacgaggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaa
caagccctgccccgagcacgacaacatggaccggatcccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatc
aagctgtgcaacatccagcagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggc
ggaggcccgagtggaagggccagtgcgatgagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcga
ctgcaagtacaaggacctggacaacaccttcaaagaattcaaggacCCATGGGGATCCCTCGAGAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTG
GGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCA
AGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGA
CTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACC
CTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGA
ACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGT
GATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCA
CAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTT
TATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAAT
CAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATAT
AGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGC
AATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGAC
AGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGAT
CTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGC
ACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGA
CCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTG
AGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccactttcctgcactgccccatctcctgcccaccctttccca
ggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATG
GTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGC
CCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAAT
GGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAA
GTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGT
GGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCA
GCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC
GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAA
AAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGG
CCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAG
GCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIKCPDENFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKL
CLRNINKVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINR
ALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKN
KFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYL
KEICSECDCKYKDLDNTFKEFKDPWGSLERSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVV
DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKA
KGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK
SRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.14 pFMCS-hIgG1-Fc2-3.8DBL-FL

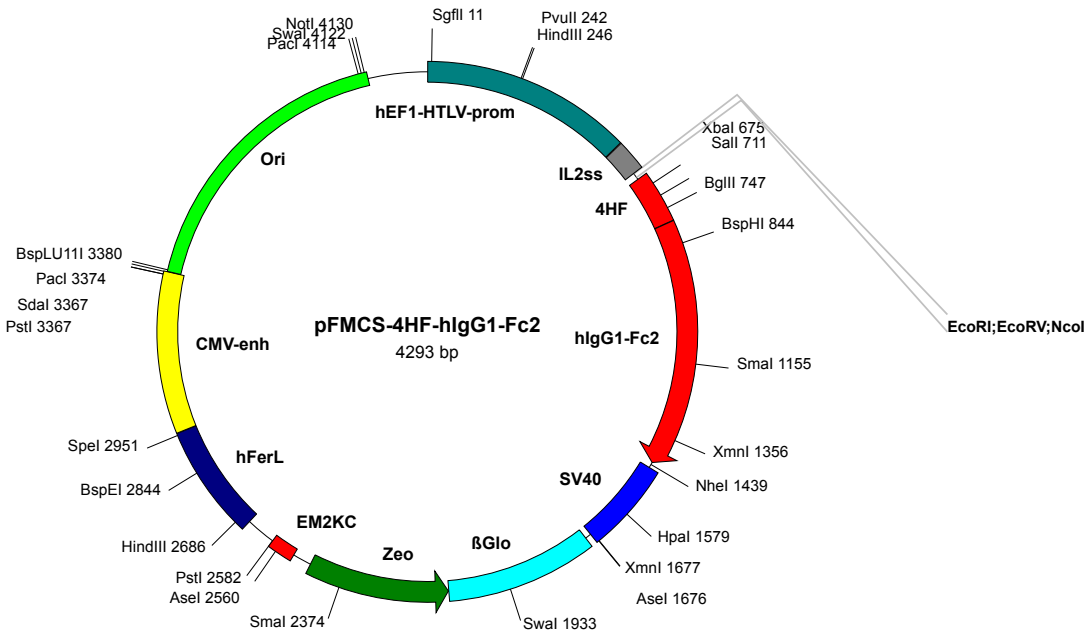


pFMCS-hIgG1-Fc2-38DBL-FL

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCaagtgccccaccgaggaaatctgcaaggacttcagcaacctgccccagtgccggaaga
acgtgcacgagcggaacaactggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccagacggcagagcctgtgcctgagaatcaccctgca
ggacttccggaccaagaagaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcc
caccaggccatccggtacagcttcgccgacatcggcaacatcatccggggcgacgacatgatggacaccccacaggcaaagagacaatcacctatctggaaaaggtgctgaaga
tctacaatgagaacaacgacaagcccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaacca
gtgcaccggctacggcaacatcgacgatatccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaatacccgaaggccgtg
tgcttccccaagcagcccagaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccg
agtggaccgagcagagcatcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacctgatcgagaagtgccccgagtgcaagtgcaccaa
gaaaaacctgcaggacgtgttcgagctgaccttcgacCCATGGGGATCCCTCGAGAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCG
TCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACT
GGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAA
TGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCA
TCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACA
AGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGA
GGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAA
TGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCA
GGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCT
ACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTT
TCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAAT
AAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAA
GCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCAT
GGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGG
TCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGG
CGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATT
ATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcatagac
agtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGG
CCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCC
CAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAG
ACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAG
TCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTT
ACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGG
GCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC
TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCT
CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAA
AATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCC
AGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFKCPTEEICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRI
TLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLE
KVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNM
NTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEK
CPECKCTKKNLQDVFELTFDPWGSLERSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS
HEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQ
PREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW
QQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.15 pFMCS-4HF-hIgG1-Fc2

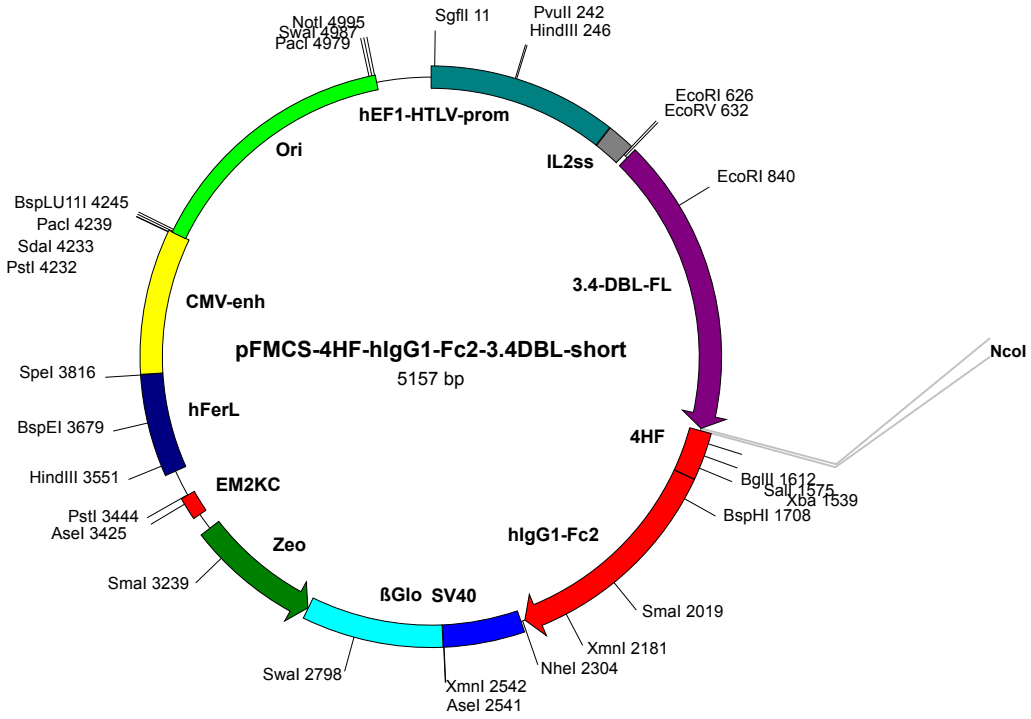


pFMCS-4HF-IgG1

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCCCATGGGACAAAACTCACACAGCCCCACCGGCCCCATCTAGAGACAAAACTC
ACACAGCCCCACCGGCCCCAGTCGACGACAAAACTCACACAGCCCCACCGGCCCCAAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACC
GTCAGTCTTCCTCTTCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAAC
TGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGA
ATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCC
ATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTAC
AAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACG
AGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGA
ATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTC
AGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTC
TACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATT
TTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAA
TAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAA
AGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCA
TGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTG
GTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCG
GCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTAT
TATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcataga
cagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCG
GCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCC
CCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGA
GACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAA
GTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTT
TACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCG
GGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC
CTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATC
TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATA
AAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCC
CAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIPWDKTHTAPPAPSRDKTHTAPPAPVDDKTHTAPPAPRSDKTHTCPPCPAPELL
GGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH
QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG
QPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.16 pFMCS-4HF-hIgG1-Fc2-3.4DBL-short

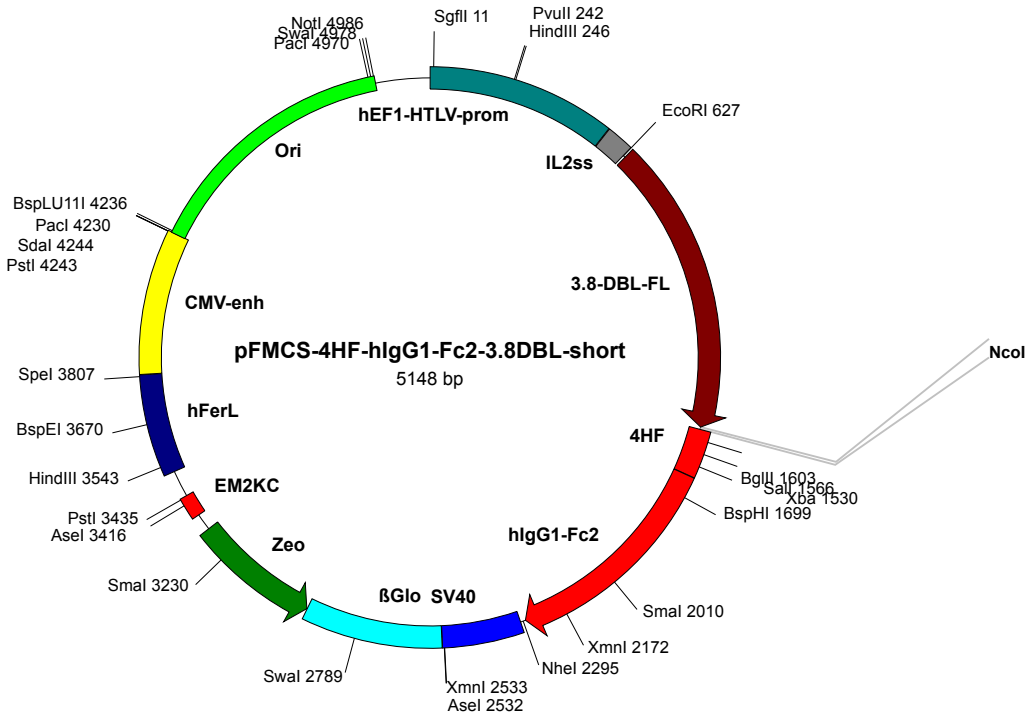


pFMCS-4HF-hIgG1-Fc2-34DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCttctgcaagggcatcaagaacgtgctgagctgccctcccaagaacagcgccg
gcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgcccccagacggaccaagctgtgcctgcggaacatcaacaaagtgtg
gcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaacctgaacgccctgaccgcc
attaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaagatcctgagaaacgaggcca
gcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaacaagccctgcccgagca
cgacaacatggaccggatcccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatcaagctgtgcaacatccag
cagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggcggaggcccgagtggaagg
gccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcgactgcaagtacaaggacct
ggacaacCCATGGGACAAAACTCACACAGCCCCACCGGCCCCATCTAGAGACAAAACTCACACAGCCCCACCGGCCCCAGTCGACGACAAAACTCACACAGCCCCACCGGCCCCA
AGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCC
CTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCA
GTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAG
AAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCT
TCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCT
CACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGA
GTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGT
AACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGT
GGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCA
ATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCC
CACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCC
AGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTT
GCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCG
TACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGT
CGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGC
TCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggt
tcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACC
CGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAG
GCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCT
TGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAA
AACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGA
CGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTAT
GGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAA
GTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT
CACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTA
ACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNIN
KVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDKIL
RNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKNKFEDVI
KLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSE
CDCKYKDLDNPWDKTHTAPPAPSRDKTHTAPPAPVDDKTHTAPPAPRSDKTHTCPPCPAPELLGGPSVFLFPPKP
KDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCK
VSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPV
LDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.17 pFMCS-4HF-hIgG1-Fc2-3.8DBL-short

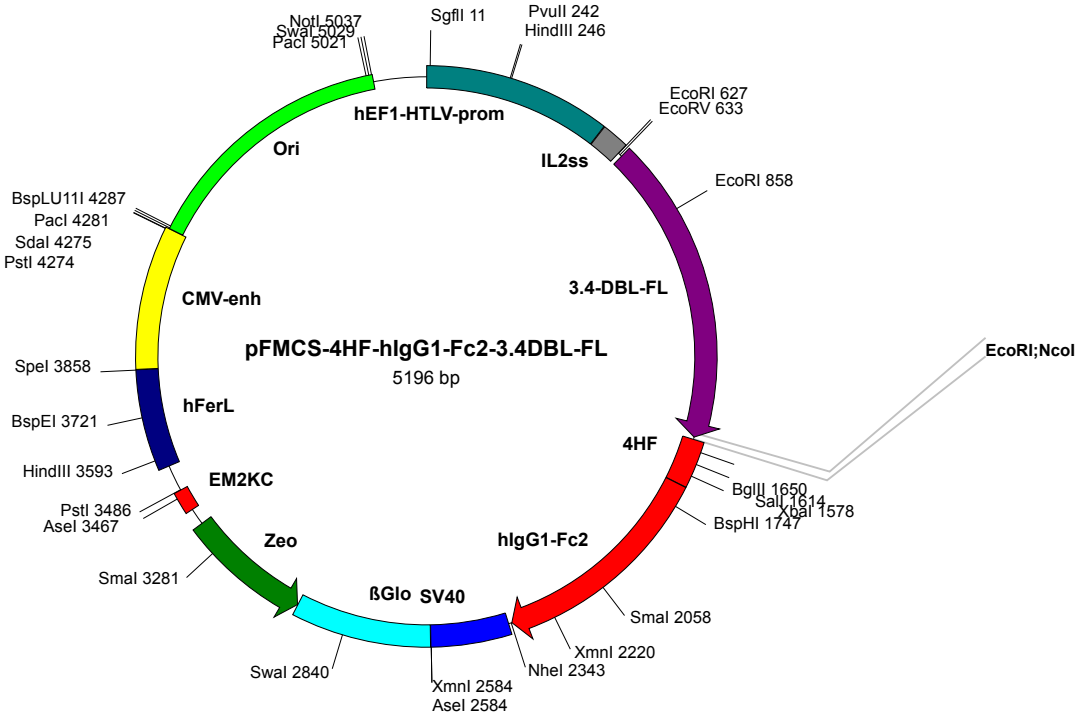


pFMCS-4HF-hIgG1-Fc2-38DBL-short

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCatctgcaaggacttcagcaacctgccccagtgccggaagaacgtgcacgagcggaaca
actggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccagacggcagagcctgtgcctgagaatcaccctgcaggacttccggaccaagaa
gaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcccacaggccatccggtac
agcttcgccgacatcggcaacatcatccggggcgacgacatgatggacaccccaccagcaaagagacaatcacctatctggaaaaggtgctgaagatctacaatgagaacaacg
acaagcccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaaccagtgcaccggctacggcaa
catcgacgatatccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaataccctgaaggccgtgtgcttccccaagcagccc
agaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccgagtggaccgagcagagca
tcaagtataacaacgataagatcaactacgccgacatcaagaccctgagcccagcgagtacctgatcgagaagtgccccgagtgcaagtgcaccaagaaaaacctgcaggacCC
ATGGGACAAAACTCACACAGCCCCACCGGCCCCATCTAGAGACAAAACTCACACAGCCCCACCGGCCCCAGTCGACGACAAAACTCACACAGCCCCACCGGCCCCAAGATCTGAC
AAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCA
CATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAG
CACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATC
TCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCA
GCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGA
CAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCT
GGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAT
AAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAA
TTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATT
AGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCC
TTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTA
GTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGG
GTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCG
TCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCA
CGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgt
caggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgc
cacttttcctgcactgccccatctcctgcccacccctttcccaggcatagacagtcagtgacttacCAAACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACC
GCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCG
AAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGT
TGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATC
ATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAG
GGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATA
CGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAG
CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTAT
CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAA
GGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGC
TCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRITLQDFRTK
KKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENN
DKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEKCPECKCTKKNLQDP
WDKTHTAPPAPSRDKTHTAPPAPVDDKTHTAPPAPRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVT
CVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTIS
KAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK
SRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.18 pFMCS-4HF-hIgG1-Fc2-3.4DBL-FL

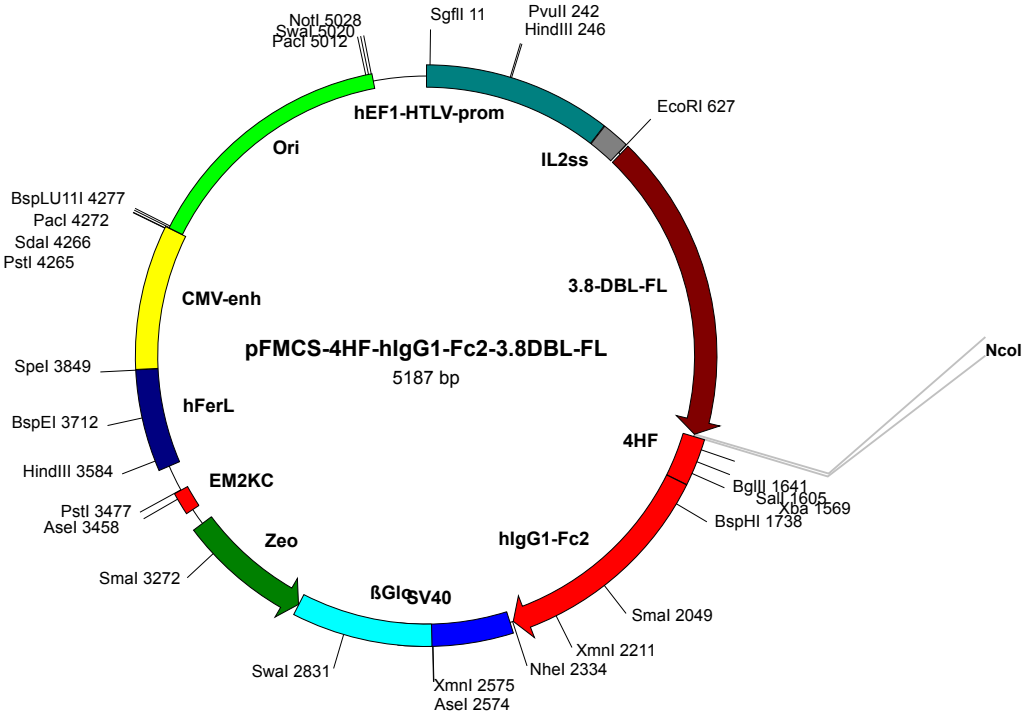


pFMCS-4HF-hIgG1-Fc2-34DBL-FL

GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCGATATCaagtgcccgacgagaacttctgcaagggcatcaagaacgtgctgagctgcc
ctcccaagaacagcgccggcagaaacggcgactggatcagcgtggccgtgaaagagagcagcaccaccaacaagggcgtgctggtgccccccagacggaccaagctgtgcctgcg
gaacatcaacaaagtgtggcaccggatcaaggacgagaagaacttcaaagaggaattcgtcaaggtcgcactgggcgagagcaacgccctgatgaagcactacaaagagaagaac
ctgaacgccctgaccgccattaagtacggcttcagcgacatgggcgacatcatcaagggcaccgacctgatcgattaccagatcaccaagaacatcaaccgggccctggacaaga
tcctgagaaacgaggccagcaacgacaagatcaagaaacgggtggactggtgggaggccaacaaggccgccttctgggacgccttcatgtgcggctacaaggtgcacatcggcaa
caagccctgccccgagcacgacaacatggaccggatcccccagtacctgcggtggtttcgcgagtggggcacctacgtgtgcagcgagtacaagaacaagttcgaggacgtgatc
aagctgtgcaacatccagcagttcaccaaccaggacgacagccagctgctggaaatcagcaagaaagacaagtgcaaagaggccctgaaacactacgaggaatgggtcaaccggc
ggaggcccgagtggaagggccagtgcgataagttcgagaaagagaagtctaagtacgaggacaccaagagcatcaccgccgagaagtacctgaaagaaatctgcagcgagtgcga
ctgcaagtacaaggacctggacaacaccttcaaagaattcaaggacCCATGGGACAAAACTCACACAGCCCCACCGGCCCCATCTAGAGACAAAACTCACACAGCCCCACCGGCC
CCAGTCGACGACAAAACTCACACAGCCCCACCGGCCCCAAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCC
CCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGT
GGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAG
TGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGA
CCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGT
GCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCAC
TACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATG
CTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTG
TGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCT
GAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACT
AGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAG
GCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTA
TCCTCAGTCCTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCG
TCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGA
TGAACAGGGTCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGT
GAGCACCGGAACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATA
CTATGCCAATGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccacttttcctgcactgccccatctcctgcccaccctttcccaggcatagacagtcagtgacttacCA
AACTCACAGGAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGC
GCTCGGGGAGGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGT
CACGCGCCTGTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGT
GAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGT
ACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCAC
CCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGT
TATGTAACGCCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCG
CTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCA
TTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTG
CCAGAACATTTCTCTATCGAA

MYRMQLLSCIALSLALVTEFDIKCPDENFCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCL
RNINKVWHRIKDEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNINRALDK
ILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQYLRWFREWGTYVCSEYKNKFEDVI
KLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVNRRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECD
CKYKDLDNTFKEFKDPWDKTHTAPPAPSRDKTHTAPPAPVDDKTHTAPPAPRSDKTHTCPPCPAPELLGGPSVFLFP
PKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVL
DSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA

A.19 pFMCS-4HF-hIgG1-Fc2-3.8DBL-FL



pFMCS-4HF-hIgG1-Fc2-38DBL-FL

```
GGATCTGCGATCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGG
GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCC
GCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCCTTCACGCGCCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTTCTGCCGCCTCCCGCCTGTG
GTGCCTCCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG
CCTGACCCTGCTTGCTCAACTCTACGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTACCTGAGATCAccggcGAAGGAGGGCCACCATG
TACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACTGAATTCaagtgccccaccgaggaaatctgcaaggacttcagcaacctgccccagtgccggaaga
acgtgcacgagcggaacaactggctgggcagcagcgtgaagaacttcgccagcgacaacaagggcgtgctggtgccccccagacggcagagcctgtgcctgagaatcaccctgca
ggacttccggaccaagaagaagaaagagggcgacttcgagaagttcatctacagctacgccagcagcgaggcccggaagctgcggaccatccacaacaacaacctggaaaaggcc
caccaggccatccggtacagcttcgccgacatcggcaacatcatccggggcgacgacatgatggacaccccaccagcaaagagacaatcacctatctggaaaaggtgctgaaga
tctacaatgagaacaacgacaagccaaggacgccaagaagtggtggaccgagaaccggcaccacgtgtgggaggccatgatgtgcggctaccagagcgcccagaaagacaacca
gtgcaccggctacggcaacatcgacgatatcccccagttcctgcggtggttccgcgagtggggcacctacgtgtgcgaagagtccgagaagaacatgaatacctgaaggccgtg
tgcttcccaagcagcccagaaccgaggccaaccctgccctgaccgtgcacgagaacgagatgtgcagcagcaccctgaagaaatacgaggaatggtacaacaagcgcaagaccg
agtggaccgagcagagcatcaagtataacaacgataagatcaactacgccgacatcaagaccctgagccccagcgagtacctgatcgagaagtgccccgagtgcaagtgcaccaa
gaaaaacctgcaggacgtgttcgagctgaccttcgacCCATGGGACAAAACTCACACAGCCCCACCGGCCCCATCTAGAGACAAAACTCACACAGCCCCACCGGCCCCAGTCGAC
GACAAAACTCACACAGCCCCACCGGCCCCAAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAAC
CCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCA
TAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTC
TCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACC
AGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTC
CGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAG
AAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTG
TGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTT
TTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTAATTCTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGA
ATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAGCTCTTCA
TTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCC
AGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTATCCTCAGTC
CTGCTCCTCTGCCACAAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGGTCATGGCCGGCCCGGAGGCGTCCCGGAAG
TTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGG
TCACGTCGTCCCGGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGG
AACGGCACTGGTCAACTTGGCCATGATGGCTCCTCctgtcaggagaggaaagagaagaaggttagtacaattgCTATAGTGAGTTGTATTATACTATGCAGATATACTATGCCAA
TGATTAATTGTCAAACTAGGGCTGCAgggttcatagtgccactttcctgcactgccccatctcctgcccaccctttcccaggcatagacagtcagtgacttacCAAACTCACAG
GAGGGAGAAGGCAGAAGCTTGAGACAGACCCGCGGGACCGCCGAACTGCGAGGGGACGTGGCTAGGGCGGCTTCTTTTATGGTGCGCCGGCCCTCGGAGGCAGGGCGCTCGGGGA
GGCCTAGCGGCCAATCTGCGGTGGCAGGAGGCGGGGCCGAAGGCCGTGCCTGACCAATCCGGAGCACATAGGAGTCTCAGCCCCCCGCCCCAAAGCAAGGGGAAGTCACGCGCCT
GTAGCGCCAGCGTGTTGTGAAATGGGGGCTTGGGGGGGTTGGGGCCCTGACTAGTCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAAC
CGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCAT
AATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACG
TCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACG
CCTGCAGGTTAATTAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAA
AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTG
TGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACAT
TTCTCTATCGAA
```

```
MYRMQLLSCIALSLALVTEFKCPTEEICKDFSNLPQCRKNVHERNNWLGSSVKNFASDNKGVLVPPRRQSLCLRI
TLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAHQAIRYSFADIGNIIRGDDMMDTPTSKETITYLE
KVLKIYNENNDKPKDAKKWWTENRHHVWEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNM
NTLKAVCFPKQPRTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYLIEK
CPECKCTKKNLQDVFELTFDPWDKTHTAPPAPSRDKTHTAPPAPVDDKTHTAPPAPRSDKTHTCPPCPAPELLGG
PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD
WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQP
ENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*VLA
```

B IgM-binding of the IgG-Fc was not detected by sandwich ELISA

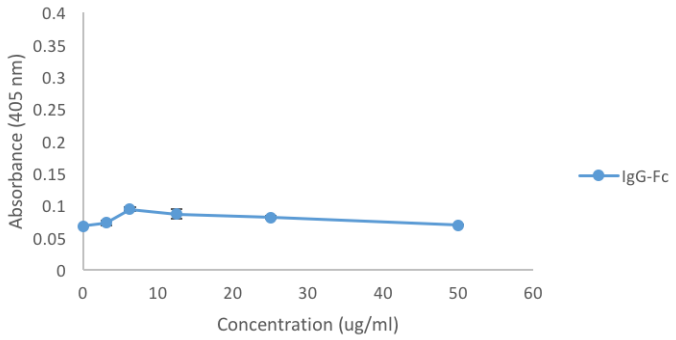


Figure .1: **IgG Fc IgM-binding control ELISA.** IgG Fc does not bind IgM. IgM was captured by αhIgM (Sigma I2386) which was coated on the plate, IgG-Fc was then added and allowed to bind, and the IgG Fc was detected with αIgG-Fc-AP (Sigma A9544). Nothing was detected, therefore the IgG Fc does not bind IgM. This shows that any binding detected in the DBL-Fc fusion proteins is due to the DBL itself and not the IgG-Fc tag.

C Summary of all MSPDBL1 and MSPDBL2 mutant DBL domains

Table .1: **3.8 DBL mutants created.** Mutants selected for protein expression are highlighted in red.

DBL mutant	Mutation	Change in charge	Position
3.8 #4	S210G		h1
	S395C		h6
3.8 #12	L203V		loop (h1a-h1)
	R423L	Y	h5
	T396A		h7
	N424K	Y	h7
3.8 #13	K294R		loop h4
	G324S		h5
	N366S		h6
3.8 #17	V202M		loop (h1a-h1)
	P205S		loop (h1a-h1)
	L211P		loop (h1-h2a)
	K255R		h3
	T335S		loop (h5-h6)
3.8 #23	K200N	Y	loop (h1a-h1)
	D229G	Y	h2
	Y235H	Y	h2
3.8 #24	K232E	Y	h3
	M277L		loop (h3-h4)
	K440E	Y	loop (+h7)
3.8 #27	K225E	Y	loop (h2-h3)
	E401K	Y	h7
3.8 #31	A242T		h3
	K363M	Y	h6
3.8 #33	K255R	Y	h4
	W310R	Y	h5

	M321V		h5
3.8 #34	G228D	Y	loop (h2a-h2)
	K330E	Y	loop (h5-h6)
3.8 #37	W309R	Y	h5
	M321L		h5
	D421V	Y	h7
3.8 #38	N300K	Y	loop (h4-h5)
	K304E	Y	loop (h4-h5)
3.8 #39	K200E	Y	loop (h1a-h1)
	W349R	Y	h6
3.8 #42	E443G	Y	loop (+h7)
3.8 #44	N185S		loop (h1a-h1)
	N339D	Y	loop (h5-h6)
3.8 #46	T430I		loop (+h7)
3.8 #47	K283I	Y	h5
3.8 #50	I270T		h3
	S432N		loop (+h7)
3.8 #51	H388Y	Y	loop (h6-h7)
3.8 #52	T379I		loop (h60h7)
3.8 #53	C177W		loop (h1a-h1)
	L204V		loop (h1a-h1)
3.8 #60	S190G		loop (h1a-h1)
	R207I	Y	h1
	C334Y		loop (h5-h6)
	A387V		loop (h6-h7)
	S394R	Y	h7
	K448E	Y	loop (+h7)
3.8 #65	H257Q	Y	h3
	T396A		h7
3.8 #70	N199Y		loop (h1a-h1)

	Q376H	Y	loop (h6-h7)
3.8 #72	S210G		loop (h1a-h1)
	I248T		h3
	H316L		h5
3.8 #75	M276I		loop (h3-h4)
	N313S		h5
	D341V	Y	loop (h5-h6)
3.8 #78	S240G		h3
3.8 #80	N186K	Y	loop (h1a-1)
	N419S		h7

The mutations in each of the thirty 3.8 DBL mutants are shown in Table C and the eighteen 3.4 DBL mutants are shown in Table .2. Mutants often contained multiple amino acid substitutions, and the table shows whether a mutation changes the charged amino acid or introduces a charge. The location of the mutation on the DBL structure is shown in the final column. Mutants selected for protein expression are highlighted in red.

Table .2: **3.4 DBL mutants created.** Mutants selected for protein expression are highlighted in red.

DBL mutant	Mutation	Change in charge	Position
3.4 #10	G169S		h1
	K194N	Y	h1
	N278D	Y	h4
	C427R	Y	h7
3.4 #22	T360S		h6
3.4 #26	K194T	Y	h1
	Q265H		h4
	E279G	Y	h4
	I285F		loop (h4-h5)
	N388S		h7
3.4 #31	F213S		loop (h1-h2a)
	R270P	Y	h4
3.4 # 36	D323V	Y	loop (h5-h6)
3.4 # 47	I266N		h4
3.4 # 48	S281G		loop (h4-h5)
	A294V		h5
	W300R	Y	h5
3.4 #51	R391K		h7
3.4 # 56	I243V		h3
3.4 # 66	I200V		loop (h1-h2a)
	H310R	Y	h5
	E318V	Y	loop (h5-h6)
3.4 #69	D283N	Y	loop (h4-h5)
3.4 # 74	D170E		h1
	L228P		h2
	N321Y		loop (h5-h6)
3.4 # 77	A175T		loop (h1-h2a)
	N269D	Y	h4
	C340S		h6
	K407M	Y	h7
3.4 # 87	N201Y		h2a
3.4 #89	K266R		h4
	E318D		loop (h5-h6)
3.4 #90	G214R	Y	h5
	K287R		loop (h4-h5)
	I325T		loop (h5-h6)
	D439Y/V350E	Y	h7
3.4 #102	D273Y	Y	h4
3.4 #104	K430R		h7

D Structural analysis of MSPDBL1 and MSPDBL2 mutant DBL domains

3.4 DBL mutant 66 (E318V)

Mutation shown in dark blue (Fig .2) and docking to (light blue) IgM model shown (Fig .3).

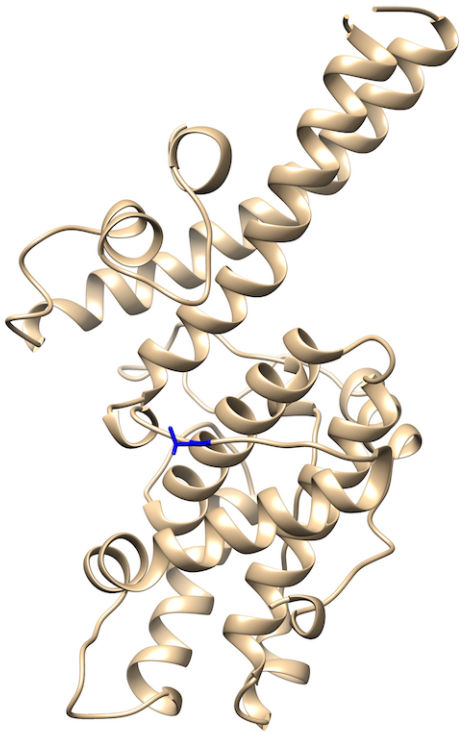


Figure .2: 3.4 DBL mutant 66. E318V mutation is located in the loop between h5 and h6.

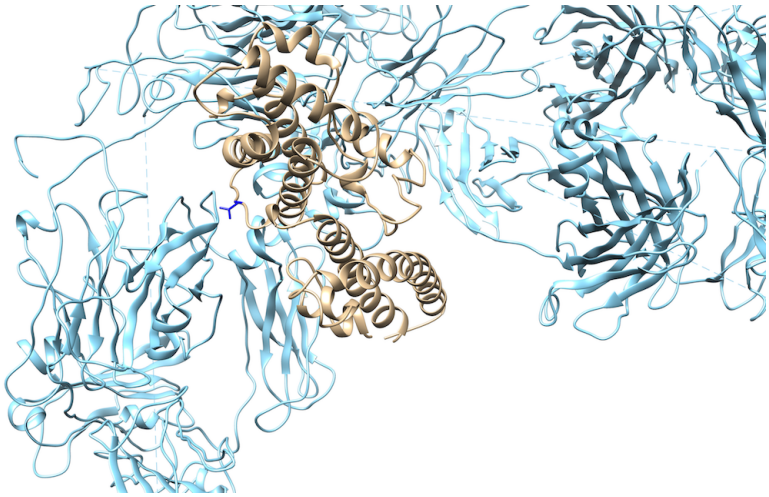


Figure .3: 3.4 DBL mutant 66 E318V mutation (blue) in relation to IgM docking.

3.4 DBL mutant 74 (N321Y)

Mutation shown in blue (Fig .4) and docking to IgM model shown (Fig .5).

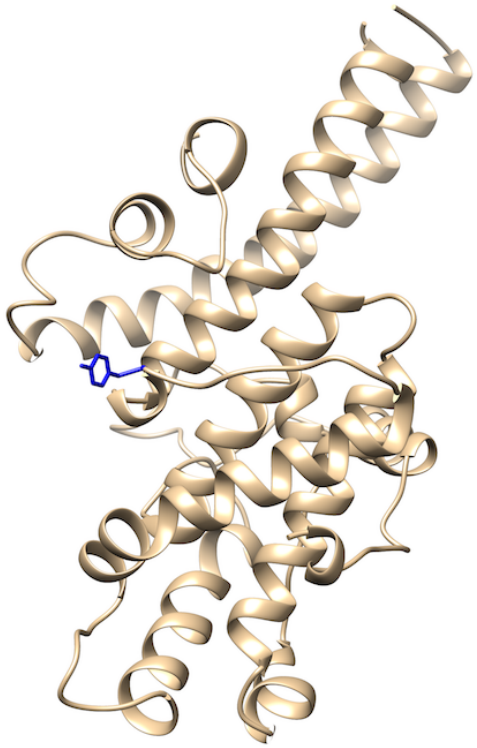


Figure .4: 3.4 DBL mutant 74. N321Y mutation is located in the loop between h5 and h6.

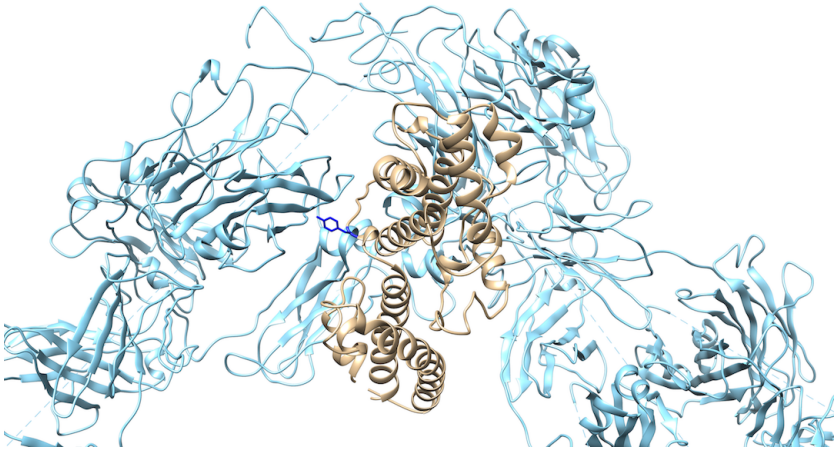


Figure .5: 3.4 DBL mutant 74 321Y mutation in relation to IgM docking.

3.8 DBL mutant 17 (T335S)

Mutation shown in blue (Fig .6) and docking to IgM model shown (Fig .7).

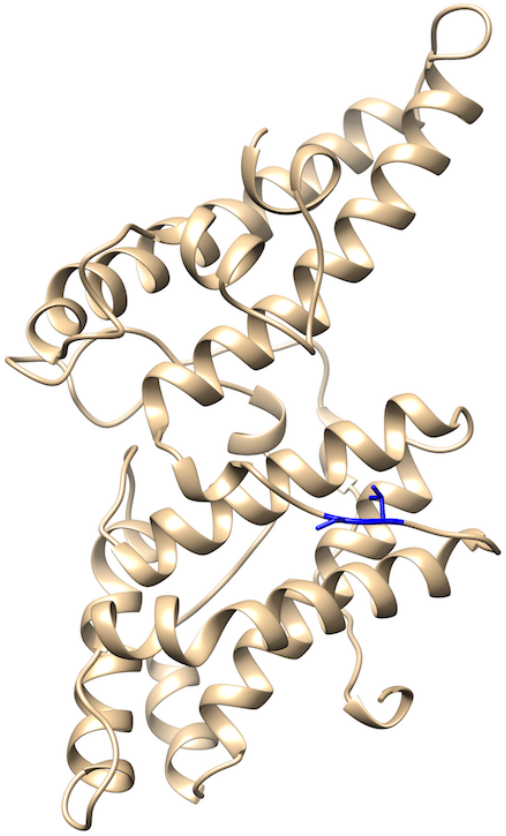


Figure .6: 3.8 DBL mutant 17. T335S mutation is located in the loop between h5 and h6.

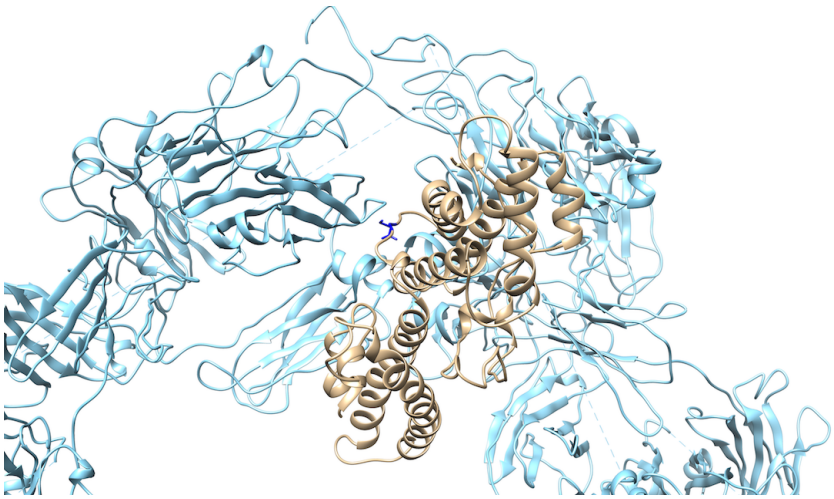


Figure .7: 3.8 DBL mutant 17 T335S mutation in relation to IgM docking.

3.8 DBL mutant 34 (K330E)

Mutation shown in blue (Fig .8) and docking to IgM model shown (Fig .9).

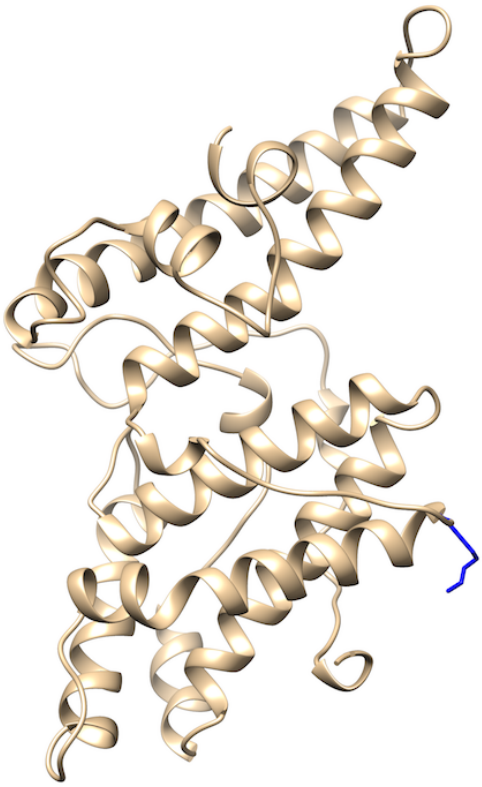


Figure .8: 3.8 DBL mutant 34. K330E mutation is located in the loop between h5 and h6.

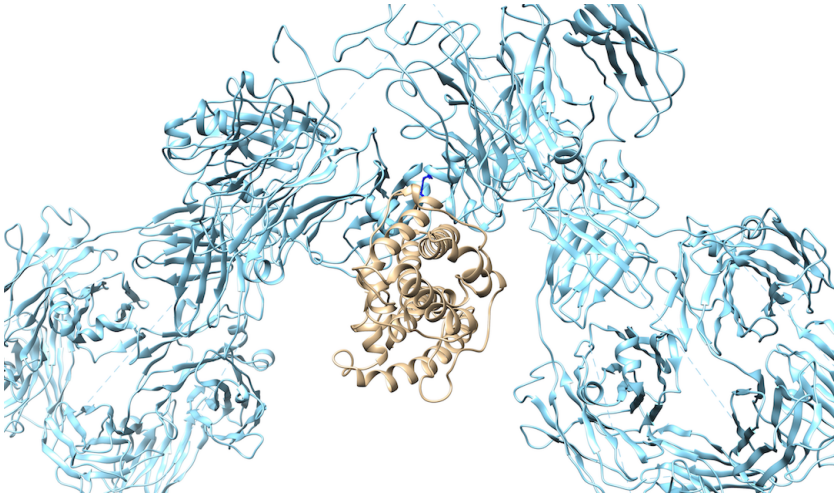


Figure .9: 3.8 DBL mutant 34 K330E mutation in relation to IgM docking.

3.8 DBL mutant 44 (N339D)

Mutation shown in blue (Fig .10) and docking to IgM model shown (Fig .11).

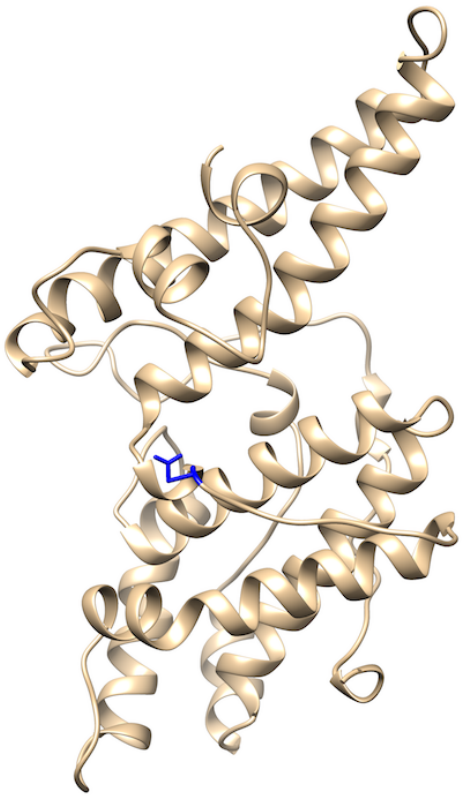


Figure .10: 3.8 DBL mutant 44. N339D mutation is located in the loop between h5 and h6.

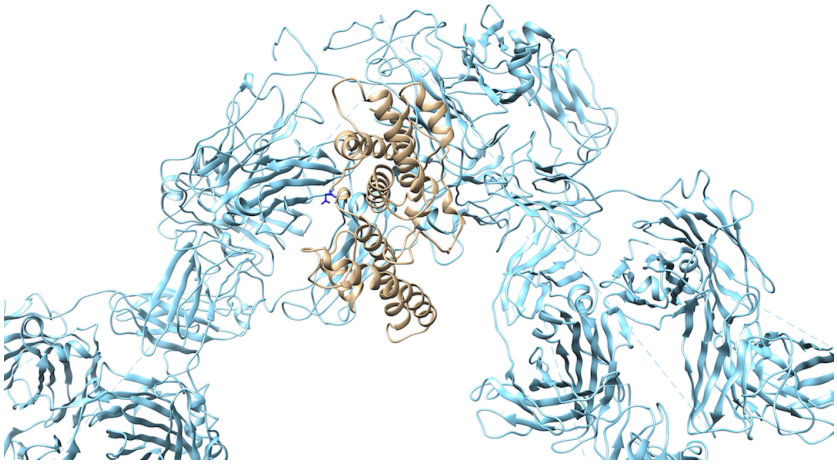


Figure .11: 3.8 DBL mutant 44 N339D mutation in relation to IgM docking.

E Sequence analysis of MSPDBL1 and MSPDBL2 mutant DBL domains

```
MSPDBL1         FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant56        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant47        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant36        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant74        FCKGIKNVLSCPPKNSAGRNGEWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant31        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant48        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant77        FCKGIKNVLSCPPKNSAGRNGDWISVTVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant22        ----------------AGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 44
Mutant69        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
Mutant26        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTTLCLRNINKVWHRIK 60
Mutant10        FCKGIKNVLSCPPKNSAGRNSDWISVAVKESSTTNKGVLVPPRRTNLCLRNINKVWHRIK 60
Mutant66        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNVNKVWHRIK 60
Mutant51        FCKGIKNVLSCPPKNSAGRNGDWISVAVKESSTTNKGVLVPPRRTKLCLRNINKVWHRIK 60
                         ****.:*****:******************.*****:********

MSPDBL1         DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant56        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAVKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant47        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQNTKNI 120
Mutant36        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant74        DEKNFKEEFVKVALGESNAPMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant31        DEKNSKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant48        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant77        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant22        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 104
Mutant69        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant26        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYHITKNI 120
Mutant10        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant66        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
Mutant51        DEKNFKEEFVKVALGESNALMKHYKEKNLNALTAIKYGFSDMGDIIKGTDLIDYQITKNI 120
                **** ************** **************:******************: ****

MSPDBL1         NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant56        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant47        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant36        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMVRIPQY 180
Mutant74        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant31        NPALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant48        NRALDKILRNEAGNDKIKKRVDWWEVNKAAFRDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant77        DRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant22        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 164
Mutant69        NRALDKILRNEASNNKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant26        NRALDKILRNGASNDKFKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant10        NRALDKILRDEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
Mutant66        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVRIGNKPCPVHDNMDRIPQY 180
Mutant51        NRALDKILRNEASNDKIKKRVDWWEANKAAFWDAFMCGYKVHIGNKPCPEHDNMDRIPQY 180
                : *******: *.*:*:*********.***** *********:******* **** *****

MSPDBL1         LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant56        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant47        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant36        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant74        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant31        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant48        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant77        LRWFREWGTYVSSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant22        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFSNQDDSQLLEISKKDKCKEALKHYEEWVN 224
Mutant69        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant26        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVS 240
Mutant10        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant66        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
Mutant51        LRWFREWGTYVCSEYKNKFEDVIKLCNIQQFTNQDDSQLLEISKKDKCKEALKHYEEWVN 240
                ***********.******************:**************************.

MSPDBL1         RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN------------ 288
Mutant56        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant47        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant36        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant74        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant31        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant48        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant77        RRRPEWKGQCDKFEKEKSMYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant22        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDN------------ 272
Mutant69        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant26        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPC- 299
Mutant10        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSERDCKYKDLDNRSDKTHTCPPCP 300
Mutant66        RRRPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
Mutant51        RRKPEWKGQCDKFEKEKSKYEDTKSITAEKYLKEICSECDCKYKDLDNRSDKTHTCPPCP 300
                **:*************** ****************** *********
```

Figure .12: MSPDBL1 mutant DBL sequence alignment.

```
Mutant17        -XXXXXXATMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant12        -XXXXXXXPMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant37        --XXXXXGPHVRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant23        --XXXXXXXMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant70        -XXXXXXXXXXXMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant65        -XXXXXXXTMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant48        -XXXXXRATMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant38        --XXXXXXXXYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant34        XXXXXXXXXMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
Mutant4         --XXXERATMYRMQLLSCIALSLALVTNSICKDFSNLPQCRKNVHERNNWLGSSVKNFAS
MSPDBL2         ----------------------------ICKDFSNLPQCRKNVHERNNWLGSSVKNFAS

Mutant17        DNKGMLVSPRRQSPCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLERAH
Mutant12        DNKGVVVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEALKLRTIHNNNLEKAH
Mutant37        DNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAH
Mutant23        DNNGVLVPPRRQSLCLRITLQDFRTKKKKEGGFEKFIHSYASSEARKLRTIHNNNLEKAH
Mutant70        DYKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAH
Mutant65        DNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAQ
Mutant48        DNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTNHNNNLEKAH
Mutant38        DNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAH
Mutant34        DNKGVLVPPRRQSLCLRITLQDFRTKKKKEDDFEKFIYSYASSEARKLRTIHNNNLEKAH
Mutant4         DNKGVLVPPRRQGLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAH
MSPDBL2         DNKGVLVPPRRQSLCLRITLQDFRTKKKKEGDFEKFIYSYASSEARKLRTIHNNNLEKAH

Mutant17        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant12        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant37        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKRWTENRHHV
Mutant23        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant70        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant65        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant48        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant38        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENKDKPEDAKKWWTENRHHV
Mutant34        QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
Mutant4         QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV
MSPDBL2         QAIRYSFADIGNIIRGDDMMDTPTSKETITYLEKVLKIYNENNDKPKDAKKWWTENRHHV

Mutant17        WEAMMCGYQSAQKDNQCSGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant12        WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant37        WEAMLCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant23        WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant70        WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKHP
Mutant65        WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant48        WEALMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant38        WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant34        WEAMMCGYQSAQEDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
Mutant4         WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP
MSPDBL2         WEAMMCGYQSAQKDNQCTGYGNIDDIPQFLRWFREWGTYVCEESEKNMNTLKAVCFPKQP

Mutant17        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant12        RTEANPALTVHENEMCSSALKKYEEWYNKRKTEWTEQSIKYNNDKIKYADIKTLSPSEYL
Mutant37        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNVKINYADIKTLSPSEYL
Mutant23        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant70        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant65        RTEANPALTVHENEMCSSALKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant48        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant38        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant34        RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
Mutant4         RTEANPALTVHENEMCSCTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL
MSPDBL2         RTEANPALTVHENEMCSSTLKKYEEWYNKRKTEWTEQSIKYNNDKINYADIKTLSPSEYL

Mutant17        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLSPQN-PRHPXDLXXPE-
Mutant12        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFPLPPKTQGHPHDLPDP*G
Mutant37        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLSPQN-PRTPS*SPGPLR
Mutant23        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLSPQN-PRTPS*SPGP*G
Mutant70        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVSS-PPK-PKTPS*SPGPLR
Mutant65        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLFPPN-PRTPS*SPGPLR
Mutant48        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLFPPK-PKDTLMISRTPE
Mutant38        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLFPPK-PKDTLMISRTPE
Mutant34        IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVSSPPKT-QGHPHDLPDP*G
Mutant4         IEKCPECKCTKKNLQDAMVRSDKTHTCPPCPAPELLGGPSVFLSPQN-PRHPHDLPDP*G
MSPDBL2         IEKCPECKCTKKNLQD----------------------------------------------
```

Figure .13: MSPDBL2 mutant DBL sequence alignment.

F Stability analysis of MSPDBL1 and MSPDBL2 mutant DBL domains

The effect of each mutation on protein stability was analysed using the SDM server. The free energy difference between the wild-type and mutant protein gave a stability score for each mutation. This predicts whether the mutation will have an impact on protein structure (Worth et al., 2011). It was predicted that 3.4 DBL mutants 48 and 77 and 3.8 DBL mutant 34 consist entirely destabilising mutations. It was predicted that 3.4 mutant 66 contains two neutral mutations, however mutation H310R is predicted to have a highly destabilising effect. It should also be noted that the mutation I243V in 3.4 mutant 56 has been predicted to have a slightly destabilising effect.

Table .3: **Summary of 3.4 DBL mutants expressed.** Mutants expressed and their corresponding amino acid mutations. Each mutation was analysed for its stabilising effect using the SDM server and given a stability score based on the free energy difference between the wild-type and mutant. A positive stability score was classed as stabilising and a negative destabilising. Scores between -0.5 and 0.5 was classed as having a neutral effect. Scores between 0.5 and 1 (-0.5 and -1) were classed as slightly stabilising (slightly destabilising). Scores between 1 and 2 (-1 and -2) were classed as stabilising (destabilising). Scores above 2 (below -2) were classed as highly stabilising (highly destabilising).

DBL mutant	Mutation	Change in charge	Position	Stability score	Stabilising Effect
3.4 #10	G169S		h1	2.11	+ + +
	K194N	Y	h1	-0.67	−
	N278D	Y	h4	0.89	+
	C427R	Y	h9	-0.30	0
3.4 #26	K194T	Y	h1	1.88	++
	Q265H	Y	h4	-0.91	−
	E279G	Y	h4	-3.59	− − −
	I285F		loop (h4-h5)	0.72	+
	N388S		h7	-0.02	0
3.4 # 36	D323V	Y	loop (h5-h6)	3.72	+ + +
3.4 # 47	I266N		h4	-2.76	− − −
3.4 # 48	S281G		loop (h4-h5)	-2.53	− − −
	A294V		h5	-1.57	−−
	W300R	Y	h5	-3.31	− − −
3.4 # 56	I243V		h3	-0.98	−
3.4 # 66	I200V		loop (h1-h2a)	-0.44	0
	H310R		h5	-2.02	− − −
	E318V	Y	loop (h5-h6)	0.13	0
3.4 # 74	D170E		h1	0.01	0
	L228P		h2	-4.85	− − −
	N321Y		loop (h5-h6)	-0.4	0
3.4 # 77	A175T		loop (h1-h2a)	-2.05	− − −
	N269D	Y	h4	-1.17	−−
	C340S		h6	-5.18	− − −
	K407M	Y	h7	0.52	−
3.4 # 87	N201Y		h4	0.96	+

Table .4: **Summary of 3.8 DBL mutants expressed.** Mutants expressed and their corresponding amino acid mutations. Each mutation was analysed for its stabilising effect using the SDM server and given a stability score based on the free energy difference between the wild-type and mutant. A positive stability score was classed as stabilising and a negative destabilising. Scores between -0.5 and 0.5 was classed as having a neutral effect. Scores between 0.5 and 1 (-0.5 and -1) were classed as slightly stabilising (slightly destabilising). Scores between 1 and 2 (-1 and -2) were classed as stabilising (destabilising). Scores above 2 (below -2) were classed as highly stabilising (highly destabilising).

DBL mutant	Mutation	Change in charge	Position	Stability score	Stabilising Effect
3.8 #4	S210G		h1	-1.31	−−
	S395G		h6	3.51	+++
3.8 #12	L203V		loop (h1a-h1)	0.19	0
	R243L	Y	h5	0.26	0
	T396A		h7	2.29	+++
	N424K	Y	h7	1.57	++
3.8 #17	V202M		loop (h1a-h1)	-0.51	−
	P205S		loop (h1a-h1)	-1.12	−−
	L211P		loop (h1-h2a)	-1.95	−−
	K255R		h3	0.83	+
	T335S		loop (h5-h6)	-0.32	0
3.8 #23	K200N	Y	loop (h1a-h1)	0.26	0
	D229G	Y	h2	-1.15	−−
	Y235H	Y	h2	-0.09	0
3.8 #34	G228D	Y	loop (h2a-h2)	-2.53	−−−
	K330E	Y	loop (h5-h6)	-0.72	−
3.8 #37	W309R	Y	h5	-3.62	−−−
	M321L		h5	-0.01	0
	D421V	Y	h7	2.93	+++
3.8 #38	N300K	Y	loop (h4-h5)	-1.11	−−
	K304E	Y	loop (h4-h5)	0.94	+
3.8 #44	N184S		loop (h1a-h1)	-0.67	−
	N339D	Y	loop (h5-h6)	0.6	+
3.8 #65	H257Q	Y	h3	-1.5	−−
	T396A		h7	2.29	+++
3.8 #70	N199Y		loop (h1a-h1)	1.35	++
	Q376H	Y	loop (h6-h7)	0.79	+

G Protein-protein interaction prediction (PPIP)

Protein-protein interface prediction was undertaken using meta-PPISP tool.

```
Prediction by meta-PPISP : metamethod for Protein-Protein Interaction Site
Prediction
Column 1: AA (Amino Acid code)
Column 2: Ch (Chain ID)
Column 3: AA# (Amino Acid number)
Column 4-7: Prediction scores of cons-PPISP, PINUP, Promate, and meta-PPISP
        Note: PINUP and Promate scores are scaled by 100
Column 8: Prediction of whether the residue is in an interface
         (P = Positive; N = Negative; - = Burried and not predicted)
       (Note: P corresponds to a score > 0.34 in 7th column; user may set a
        different threshold for positive prediction.)
**************************************************************************
R     50    0.003   0.830   0.393   0.463      P
N     51    0.000   0.990   0.396   0.000      -
I     52    0.050   0.990   0.404   0.509      P
N     53    0.005   0.990   0.435   0.568      P
K     54    0.031   0.990   0.400   0.554      P
V     55    0.008   0.660   0.407   0.432      P
W     56    0.006   0.990   0.472   0.586      P
H     57    0.002   0.830   0.386   0.397      P
R     58    0.002   0.990   0.363   0.423      P
I     59    0.002   0.490   0.369   0.382      P
K     60    0.002   0.490   0.407   0.369      P



Y    114    0.039   0.660   0.462   0.361      P
Q    115    0.003   0.660   0.368   0.326      N
I    116    0.002   0.830   0.428   0.455      P
T    117    0.000   0.000   0.380   0.000      -
K    118    0.002   0.330   0.333   0.211      N
N    119    0.002   0.490   0.295   0.375      P
I    120    0.000   0.000   0.408   0.000      -
N    121    0.002   0.830   0.312   0.433      P
R    122    0.002   0.660   0.300   0.296      N
A    123    0.002   0.990   0.376   0.517      P
```

Figure .14: **Meta-PPISP protein interface prediction results for MSPDBL1.** Region in h2a and h4 predicted to be protein-protein interface.

Prediction by meta-PPISP : metamethod for Protein-Protein Interaction Site Prediction
Column 1: AA (Amino Acid code)
Column 2: Ch (Chain ID)
Column 3: AA# (Amino Acid number)
Column 4-7: Prediction scores of cons-PPISP, PINUP, Promate, and meta-PPISP
Note: PINUP and Promate scores are scaled by 100
Column 8: Prediction of whether the residue is in an interface
(P = Positive; N = Negative; - = Burried and not predicted)
(Note: P corresponds to a score > 0.34 in 7th column; user may set a different threshold for positive prediction.)

****

K	13	0.033	0.550	0.321	0.370	P
N	14	0.007	0.660	0.366	0.352	P
V	15	0.082	0.880	0.379	0.488	P
H	16	0.006	0.770	0.350	0.428	P
E	17	0.006	0.660	0.399	0.501	P
R	18	0.011	0.880	0.425	0.567	P
N	19	0.012	0.880	0.375	0.554	P
N	20	0.808	0.880	0.378	0.491	P
W	21	0.928	0.990	0.466	0.710	P
L	22	0.669	0.880	0.435	0.580	P
G	23	0.049	0.880	0.458	0.519	P
S	24	0.039	0.550	0.452	0.364	P
S	25	0.017	0.440	0.410	0.361	P

L	37	0.854	0.550	0.360	0.345	P
V	38	0.959	0.770	0.415	0.493	P
P	39	0.000	0.990	0.440	0.000	-
P	40	0.775	0.660	0.435	0.430	P
R	41	0.000	0.660	0.421	0.000	-
R	42	0.000	0.660	0.329	0.000	-
Q	43	0.865	0.990	0.419	0.592	P
S	44	0.715	0.660	0.387	0.514	P

Figure .15: **Meta-PPISP protein interface prediction results for MSPDBL2.** Region in loop h1a-h1 and loop h1-h2a predicted to be protein-protein interface.

H Homology block analysis

Table .5: Homology blocks commonly found in IgM-binding DBL domains.

Homology block	% occurrence in binders	% occurrence in non-binders	% occurrence in VarDom server
97	50	7	4
108	30	7	3
160	30	7	1
222	20	0	0.9
318	20	0	0.6

Table .6: Homology blocks commonly found in non-binding DBL domains.

Homology block	% occurrence in binders	% occurrence in non-binders	% occurrence in VarDom server
45	10	53	12
48	20	40	10
260	0	20	0.9
304	0	13	0.5
305	0	13	0.5
311	0	13	0.5
326	0	13	0.5
327	0	13	0.5
353	0	13	0.5
380	0	13	0.5
393	0	13	0.5
403	0	13	0.5

Homology blocks (HBs) contained in ten IgM-binding DBL domains and fifteen non-binding DBL domains described by Pleass et al. (2015), were analysed. Homology blocks which frequently occur in either IgM-binding (Table .5) or non-binding DBL domains (Table .6) were identified. These were compared to how often they occur in the VarDom server. HBs are numbered according to their frequency in the database, with lower numbers more common (Rask et al., 2010). Homology blocks 260, 304, 305, 311, 326, 237, 280, 293 and 403 appear only in non-binding DBL domains, and not in IgM-binding DBL domains. Homology blocks 222 and 318 appear in only IgM-binding DBL domains, and not in non-binding domains. These are rare homology blocks in the VarDom server, occurring in less than 1% of the DBLs analysed across seven genomes. When a HB occurs multiple times, it occurs in DBL domains belonging to the same DBL class (e.g. HB 304

is found in two DBL4ε domains and HB 403 is found in two DBLζ domains). These would therefore be expected to have sequence homology.

Using the open source SPMF data-mining library (available at `http://www.philippe-fournier-viger.com/spmf/index.php`), four homology block motifs occurring in multiple IgM-binding DBL domains were identified by the Apriori algorithm (Fournier-Viger et al., 2014). These were (97 108), (76 160), (76 97) and (76 126). However, three of these motifs were also observed in non-binding DBL domains. The motif (76 97) was not observed in a non-binding DBL domain and may therefore be unique to IgM-binding domains. However, 76 is a common HB at the N terminal of both IgM-binding and non-binding DBL domains. The C terminal HB 97, while more common in IgM-binding DBL domains has been observed in the non-binding MAL6P1.4 DBL7ε domain. Further analysis awaits a larger dataset to identify whether the HB motif (76 97) is unique to IgM-binding DBL domains.

I Structural analysis of known IgM-binding and non-binding DBL domains

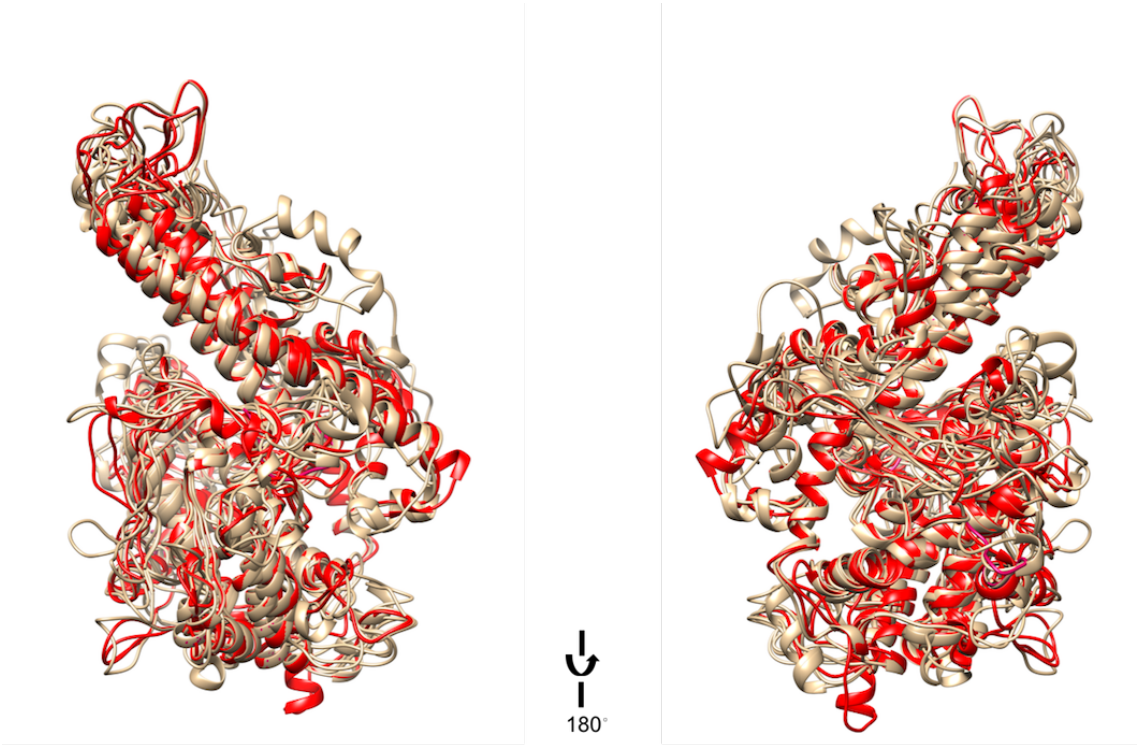


Figure .16: **Structural conservation of all known IgM-binding vs. non-binding DBL domains.** Known IgM-binding DBL domains (red) superimposed on non-binding isolates (cream). Homology models of any DBL domains with unknown structure were produced using the Phyre2 software based on the crystal structure of their most closely related known DBL structure (Kelley et al., 2015) and superimposed to highlight structural differences.

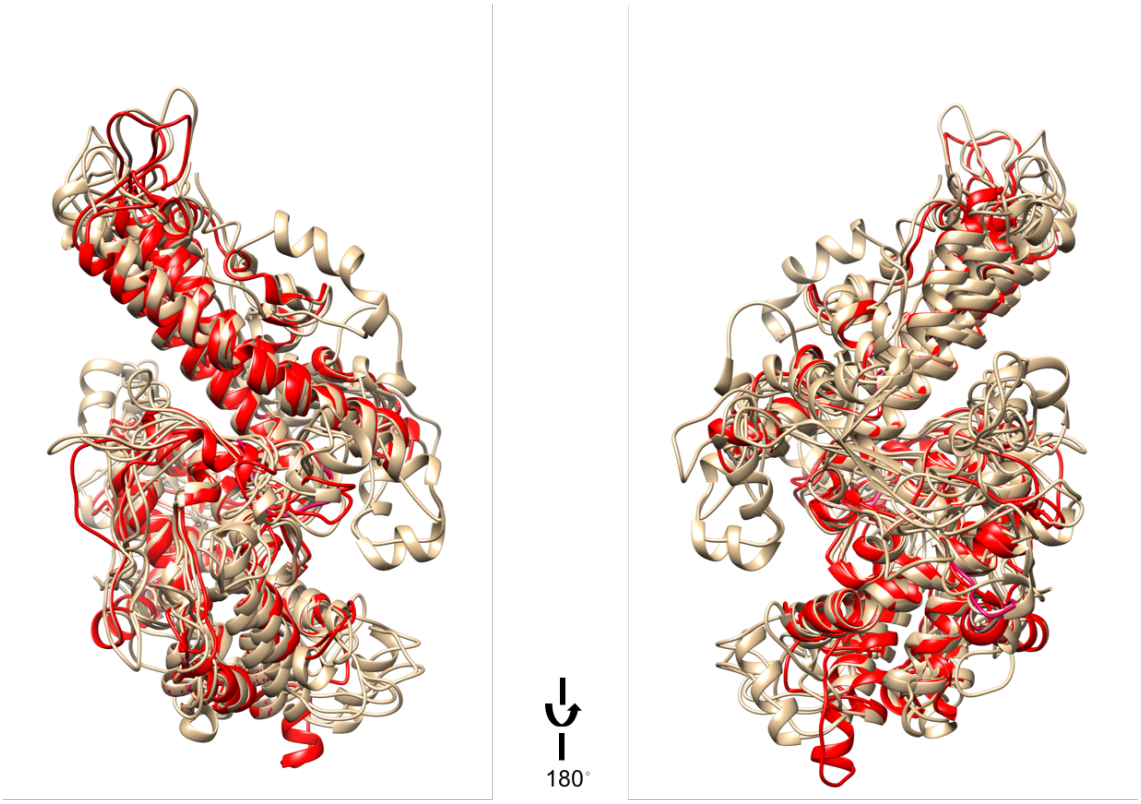


Figure .17: **Structural conservation of all known IgM-binding vs. non-binding DBLε domains.** Known IgM-binding DBLε domains (red) superimposed on non-binding DBLε domains (cream). Homology models of any DBL domains with unknown structure were produced using the Phyre2 software based on the crystal structure of their most closely related known DBL structure (Kelley et al., 2015) and superimposed to highlight structural differences.

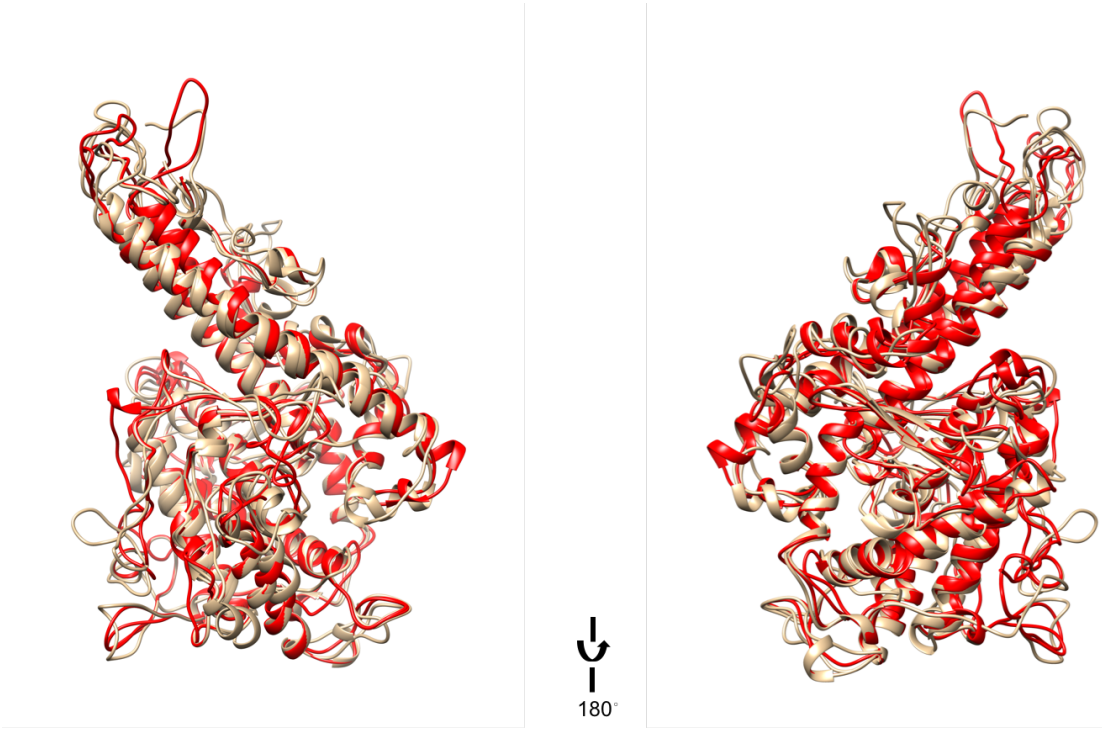


Figure .18: **Structural conservation of all known IgM-binding vs. non-binding DBLζ domains.** Known IgM-binding DBLζ domains (red) superimposed on non-binding DBLζ domains (cream). Homology models of any DBL domains with unknown structure were produced using the Phyre2 software based on the crystal structure of their most closely related known DBL structure (Kelley et al., 2015) and superimposed to highlight structural differences.

J FPLC profiles

Merozoite DBL mutant Fc-fusion proteins

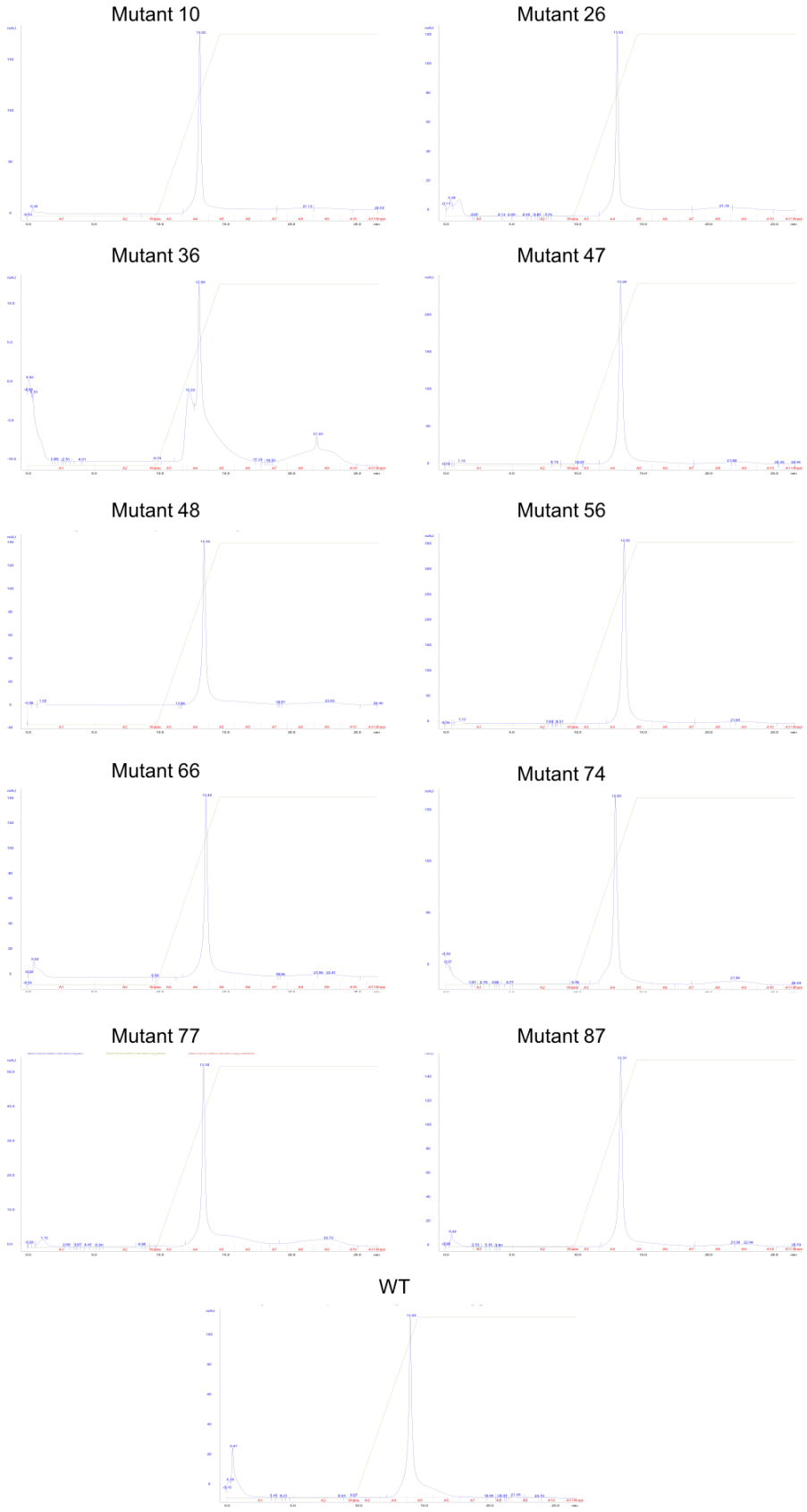


Figure .19: **3.4 DBL mutant purification FPLC profiles.** Expression conditions were kept the same for each mutant, and the purification peaks compared to the wild-type 3.4 DBL domain.

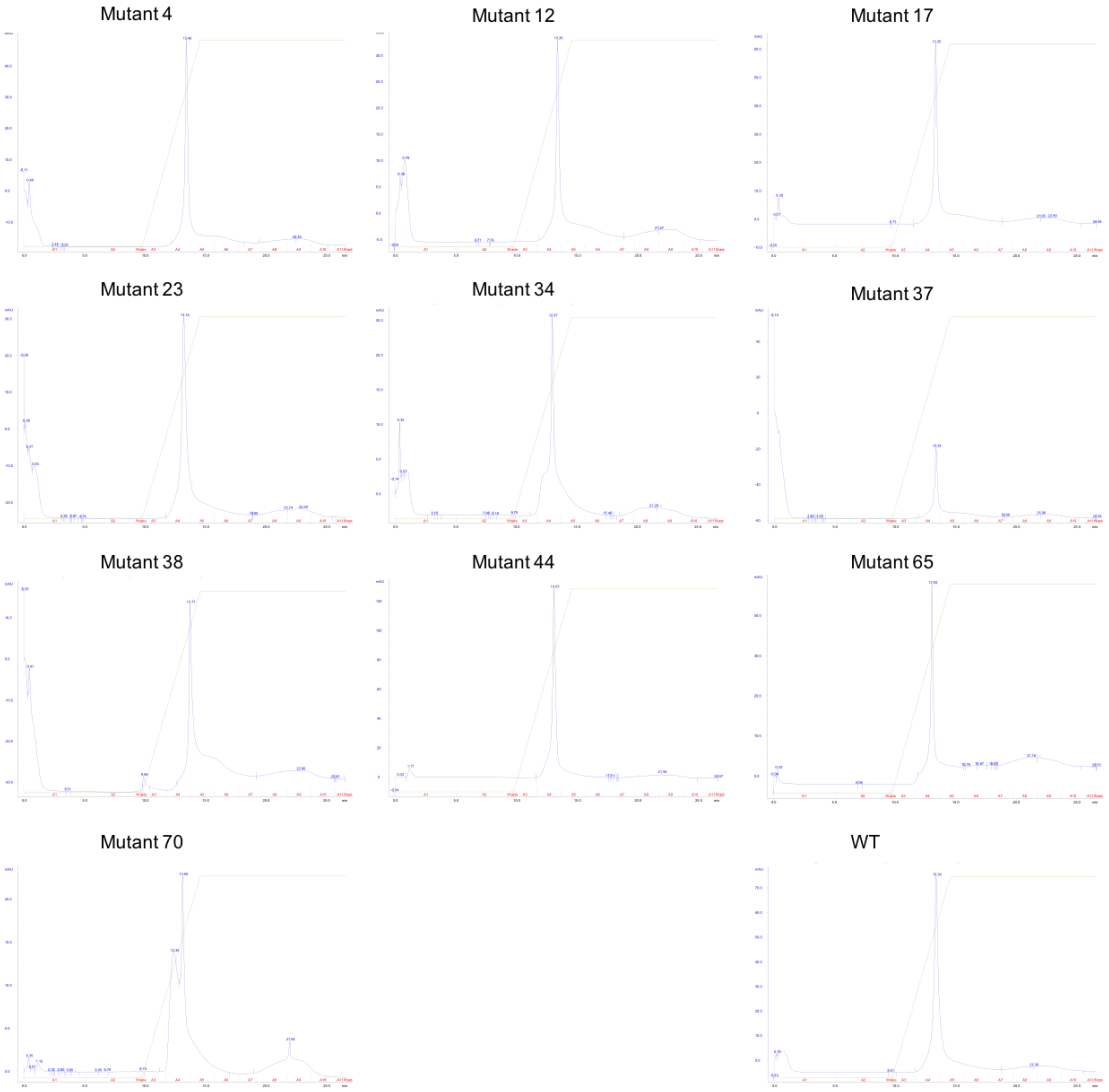
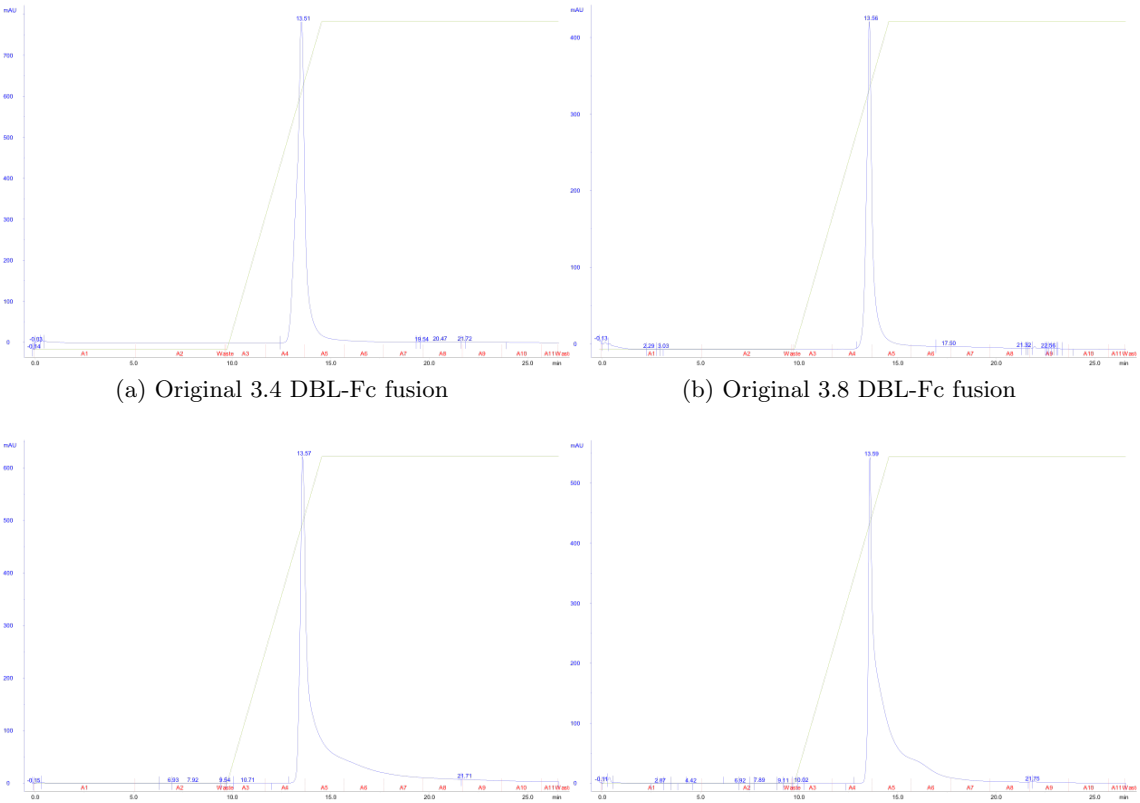


Figure .20: **3.8 DBL mutant purification FPLC profiles.** Expression conditions were kept the same for each mutant, and the purification peaks compared to the wild-type 3.8 DBL domain.

Extended-hinge DBL-Fc fusion proteins



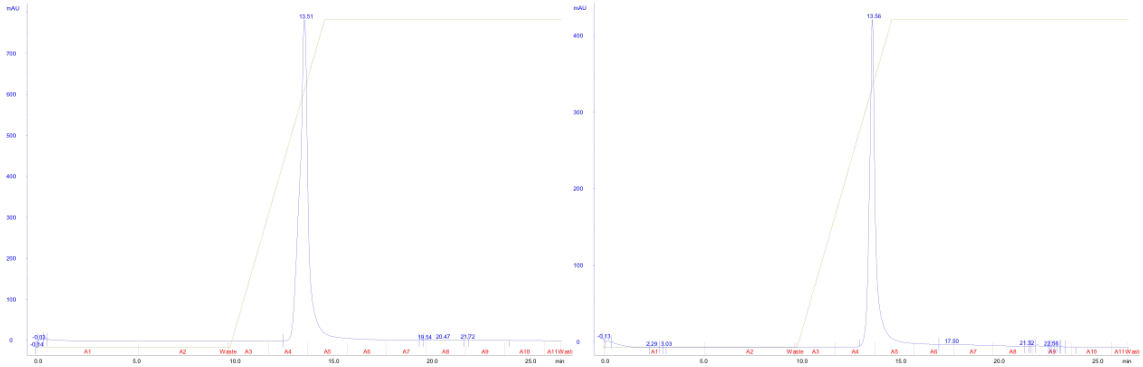
(a) Original 3.4 DBL-Fc fusion

(b) Original 3.8 DBL-Fc fusion

(c) Full-length 3.4 DBL-Fc fusion

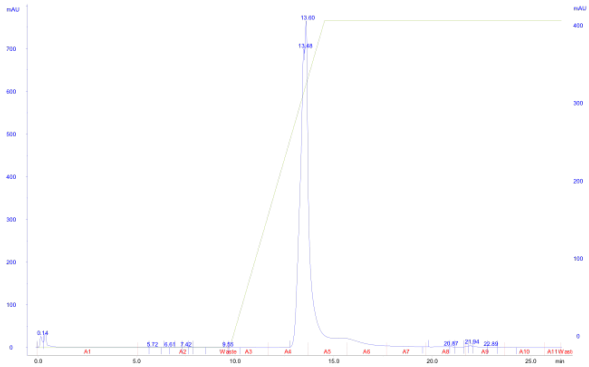
(d) Full-length 3.8 DBL-Fc fusion

Figure .21: **FPLC purification profiles for the original DBL-Fc fusion proteins vs. the new full-length DBL-Fc fusions with the original IgG1 hinge.** Expression conditions were the same for each recombinant protein produced. Purification peaks of the full-length DBL-Fc fusion proteins show no improvement on the original short DBL-Fc fusion proteins.

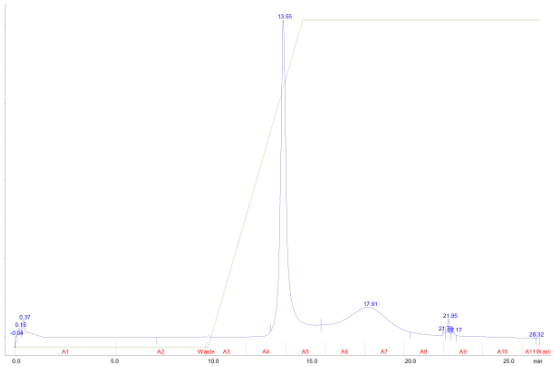


(a) Original 3.4 DBL-Fc fusion

(b) Original 3.8 DBL-Fc fusion

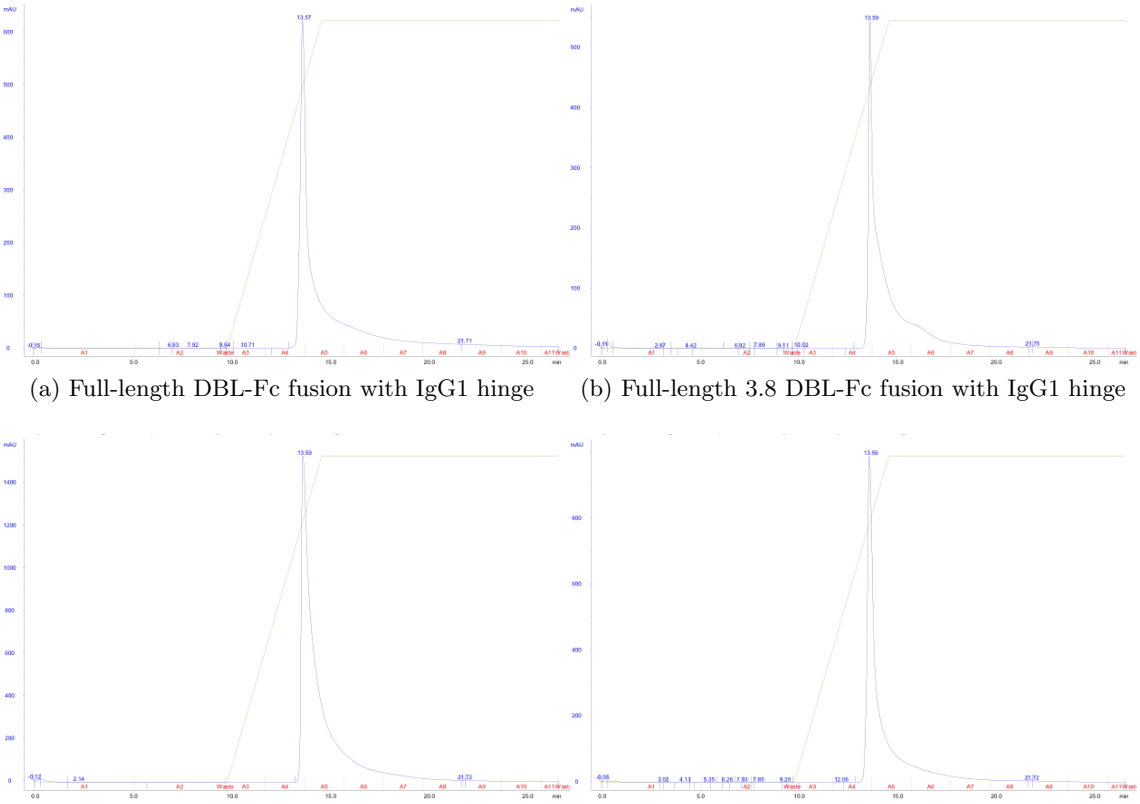


(c) Original 3.4 DBL with extended hinge



(d) Original 3.8 DBL with extended hinge

Figure .22: **FPLC purification profiles for the original DBL-Fc fusion proteins vs. the new extended hinge DBL-Fc fusions with the original short DBL domains.** Expression conditions were the same for each recombinant protein produced. Purification peaks of the extended hinge DBL-Fc fusion proteins show no improvement on the original short DBL-Fc fusion proteins.



(a) Full-length DBL-Fc fusion with IgG1 hinge

(b) Full-length 3.8 DBL-Fc fusion with IgG1 hinge

(c) Full-length 3.4 DBL-Fc fusion with extended hinge

(d) Full-length 3.8 DBL-Fc fusion with extended hinge

Figure .23: **FPLC purification profiles for the full-length DBL-Fc fusion proteins containing the original IgG1 hinge vs. the new extended hinge.** Expression conditions were the same for each recombinant protein produced. The purification peak for the extended hinge version of 3.4 FL DBL-Fc is much bigger than for the same full-length 3.4 DBL produced as an Fc fusion with only the original IgG1 hinge. There is no difference in the purification peaks for the similar FL 3.8 DBL-Fc fusion proteins.