Genetic engineering of recombinant anti-mycolic acid antibody fragments for use in Tuberculosis diagnostics

By

Johannes Loubser Schoombie

Submitted in partial fulfillment of the degree

Magister Scientiae

in the Faculty of Natural and Agricultural sciences

Department of Biochemistry

University of Pretoria

Pretoria

2012

Submission declaration

I, Johannes Loubser Schoombie, declare that this thesis/ dissertation, which I hereby submit for the degree *Magister Scientiae* in the Department of Biochemistry, at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: _____

Date: _____

Plagiarism declaration

UNIVERSITY OF PRETORIA

FACULTY OF NATURAL AND AGRICULTURAL SCIENCES

DEPARTMENT OF BIOCHEMISTRY

Full name: _____ Student number: _____

Title of the work: _____

Declaration

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this _____ (e.g. essay, report, project, assignment, dissertation, thesis etc) is my own, original work. Where someone else's work was used (whether from a printed source, the internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not make use of another student's previous work and submit it as my own.
4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his or her own work.

Signature_____ Date _____

Acknowledgements

My supervisor Dr. M Beukes of the Department of Biochemistry for guidance and support during this project.

Mr W van Wyngaardt for providing the Nkuku phage display library.

Prof E Weiss and Ms. B Linder for donating the pGE20 and pAK400 plasmids respectively.

Ms. J George for support and constructive criticism.

Summary

Mycolic acids are long chain lipids from the cell walls of *Mycobacterium tuberculosis*. The Nkuku phage display library was previously used to obtain monoclonal antibody binders to mycolic acids. In total 11 binders were obtained of which one was selected (MAC10) for further investigation by genetic engineering as presented in this dissertation. The antibodies of the Nkuku phage display library are in the format of single chain variable fragments (scFv). ScFv's constitute only the epitope binding domains of an antibody consisting of the V_H and V_L domains fused into a single chain by a flexible linker protein. The selected anti-mycolic acid scFv is referred to as mycolic acid clone 10 (MAC10). Genes encoding the scFv's of the Nkuku phage display library were cloned into the plasmid pHEN-1, a phage display vector. This vector is not commercially available or ideally suited for expression of scFv proteins. Therefore two vectors were investigated as possible targets for subcloning. The plasmids pGE20 and pAK400 were previously used for the expression of scFv antibody proteins. Subcloning into plasmid pAK400 proved to be the more efficient of the two investigated for subcloning. This subcloning yielded the recombinant plasmid pAKJS. Following the subcloning scFv protein expression was attempted using the plasmids pMAC10 (derived from pHEN-1) and pAKJS (derived from pAK400). Expression of MAC10 using plasmid pMAC10 in both *Escherichia coli* TG-1 and HB2151 was constitutive. This demonstrates that plasmid pHEN-1 is a non ideal vector as expression should not occur unless induced. Expression of MAC10 did not occur when pAKJS and *Escherichia coli* HB2151 were used. This was due to both the vector and expression host producing inhibitor protein for the Lac Z promoter controlling expression of the scFv. The MAC10 gene was subsequently randomized using the directed evolution method, error prone PCR. Sequence analysis of the five selected mutants indicated an average mutation rate of 8.6 mutations per 1000 base pairs. From the combined total of all five mutants, transversions made up the majority of substitutions. The majority of transversion mutations occurred at A-T base pairs. Transition substation mutations that made up the minority of total mutations occurred mostly at G-C base pairs.

Contents

List of figures

List of tables

Abbreviations

Amp	Ampicillin
APS	Ammonium persulfate
ATP	adenosine triphosphate
BSA	Bovine serum albumin
C	domain constant domain
CDR	complementarity determining region
C gene	constant gene
C-terminal	carboxyl terminal
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
D gene	diversity gene
dGTP	deoxyguanosine triphosphate
dNTP	2'-deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EP-PCR	error prone polymerase chain reaction
Fc	crystallizing fragment
g3	gene three
g3p	gene three protein

hER	human oestradiol receptor
His-tag	polyhistidine tag
IPTG	isopropyl- β-D-thiogalactosidase
IG	intergenic region
J gene	joining gene
LB	Luria Bertoni medium
N-terminal	amino terminal
NE	non expression
PCR	polymerase chain reaction
PelB	*Erwinia carotovora* bacterium pectate lyase B leader sequence
RT	room temperature
ScFv	single chain variable fragment
SDS	sodium dodecyl sulfate
SupE	suppressor strain
Taq	*Thermus aquaticus* polymerase
TEMED	N,N,N,N-tetramethylethylenediamine
TetA	tetracycline A
TetR	tetracycline R
V domain	variable domain
V gene	variable gene
X-gal	5-bromo-4-chloro-indolyl- β-D-galactopyranoside

Chapter 1 : Introduction

Single chain variable fragments

Antibodies are a family of glycoproteins found in the serum and lymph of all vertebrates. These proteins occur mostly as free proteins but also membrane bound surface receptors. As proteins they are produced by B-cells as part of the adaptive immune system which reacts and adjusts to immunological challenges. In mammals including humans there are five classes of antibody IgG, IgA, IgM , IgD and IgE. The basic structure of antibodies is represented by IgG, consisting of two identical light chains and two identical heavy chains (figure 1.1). The class as well as IgG and IgA subclass is determined by the heavy chains. Within the light and heavy chains there are intra chain disulphide bonds that divide the chains into domains. Each globular domain contains a central peptide loop. The light chain contains two domains (V_L and C_L) and the heavy chains contain four (V_H and C_H1-3). The V refers to variable while C refers to constant. Inside each V domain are three hypervariable or complementarity determining regions (CDRs). Between the CDR regions are so called framework regions that are not directly involved in antigen binding (Turner *et al*., 2001 and Lesk., 2004).

Antibodies are bifunctional proteins meaning that they have two functional activities. The primary function of antibodies is to bind structural features of antigens termed epitopes, and is determined by the variable domains. Secondary/ effector functions such as complement activation and phagocyte recruitment are determined by the heavy chains (Fc fragments). Proteolytic cleavage of antibodies yields two identical antigen binding Fab fragments and one Fc fragment (Turner *et al*., 2001 and Lesk., 2004).

Antibodies are a diverse family of proteins capable of binding to the vast number of antigen epitopes presented to the vertebrate immune system. The diversity of unique antibodies available is generated by the recombination of several genes and their protein products within each B cell (Turner *et al*., 2001 and Lesk., 2004).

In each B cell an antibody light chain protein is generated by recombination of various variable, joining and constant region gene segments. The number and sequence for these gene segments differ between the two classes of light chain (λ and κ). At the same time diversity is created in the heavy chains by recombination of various variable, joining, diversity and constant region gene segments. The gene segments for either the light chain or heavy chains occur in tandem close to one another. Imprecise recombinations of gene segments as well as the combination of light and heavy chains add further layers of diversification. Somatic hypermutation of the gene sequences of the proteins leads to further diversification of antibodies upon repeated exposure to the specific epitope (Turner *et al*., 2001 and Lesk., 2004).

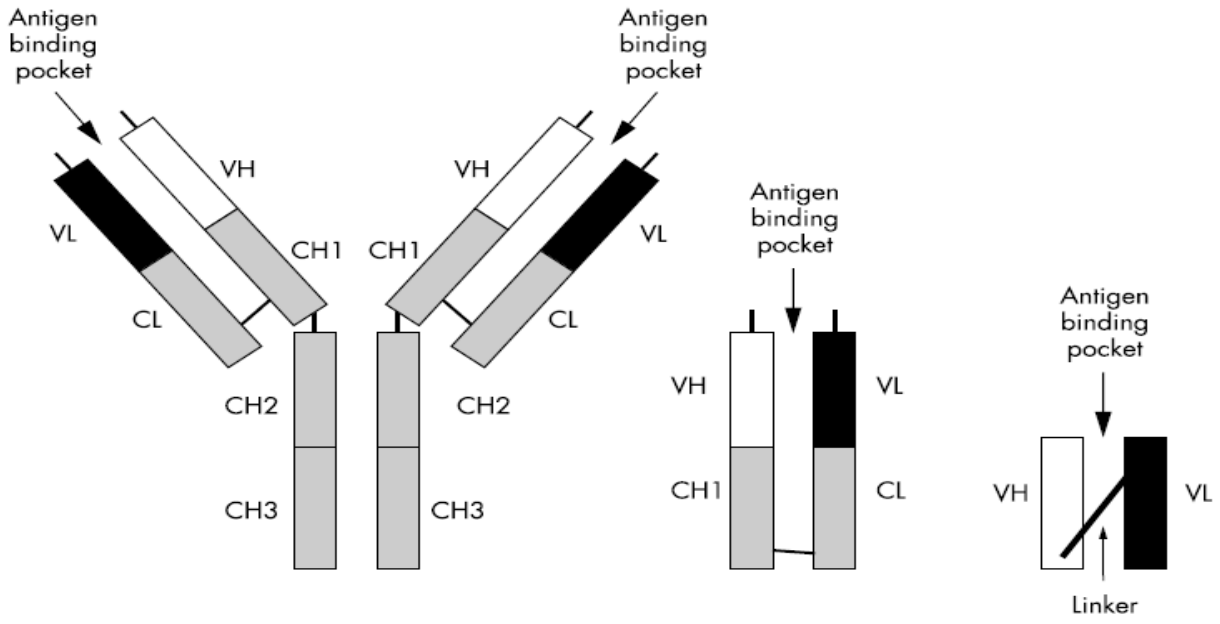


Figure 1.1: Comparison of a human IgG antibody, Fab fragment and scFv structures. VH = variable region of heavy chain, VL = variable region of light chain, CH = constant region of heavy chain and CL = constant region of light chain (Smith *et al*., 2004).

Single chain fragment variables are semi-synthetic constructs obtained by fusing the variable chains from antibodies into a single protein. The variable heavy and light chains are connected together by a flexible linker of 10 to 20 amino acids in length. The most commonly used linker is a 15 amino acid peptide consisting of three Gly_4Ser repeats. The purpose of the peptide linker is to replace the constant region associations as found in Fab fragments. Without which the two variable chains associate poorly and are difficult to express. A scFv thus represents only the part of an antibody that determines antigen binding. The greatest benefit to a scFv only consisting of the antigen binding domain is its size. A standard IgG is 150 kDa in size, a Fab fragment is 50 kDa, while a scFv on the other hand is only 25 kDa in size. Also the only post translational modification required of a scFv is the formation of intra chain disulfide bonds. The small size and lack of post translational modification allow for the expression of scFv's in *Escherichia coli* to be much simpler and easier than Fab fragments. A scFv retains the binding characteristics of a complete antibody and in many cases the same affinity (Bird *et al*., 1988, Huston *et al*., 1988; Gavilondo *et al*., 2000 and Smith *et al*., 2004).

Phage display or recombinant phage technology

Phage display technology was developed soon after the scFv antibody format. The benefit of pairing the two technologies was clearly apparent. In phage display a binding protein is displayed as a fusion to the minor coat protein of M13, fd or fl bacteriophages. Bacteriophages are viruses that infect *Escherichia coli* (figure 1.2). The five coat proteins of these viruses are encoded in their single stranded DNA genome. After infection the genome is replicated and bacteriophage particles are assembled. Following assembly the new bacteriophages are secreted from the *Escherichia coli* host without lysis (McCafferty *et al*., 1990, Kang *et al* and Siegel *et al*., 2002).

Phage display technology is derived from the biology of Ff bacteriophages/coliphages. These filamentous bacteriophages (f1, fd and M13) infect the Gram negative enteric bacteria like *Escherichia coli* by way of the F^+ pilus. An important part of the bacteriophage genome is the 500 base pair intergenic (IG) region responsible for replication and packaging of the single stranded genome. Two important components of phage display are phagemid plasmids and helper phages. Phagemids are plasmids which contain the IG region of a bacteriophage. Helper phages are derivates of wild type bacteriophages which contain defective replication and packaging signals in their genomes. However the genes and elements required for the production of bacteriophage proteins are unchanged. The result is that bacteriophages produced by co-infection of *Escherichia coli* with a phagemid plasmid and helper phage result in the preferential packaging of the plasmid into the virus. Helper phage genomes also code for wild type gene three protein (g3p) so that some copies of the protein are not fusions to ensure infectivity is maintained in bacteriophage produced by co-infection (Zagursky *et al*., 1984, Zinder *et al*., 1985, Bird *et al*., 1988, Kang *et al*., 1991 and Chasteen *et al*., 2006).

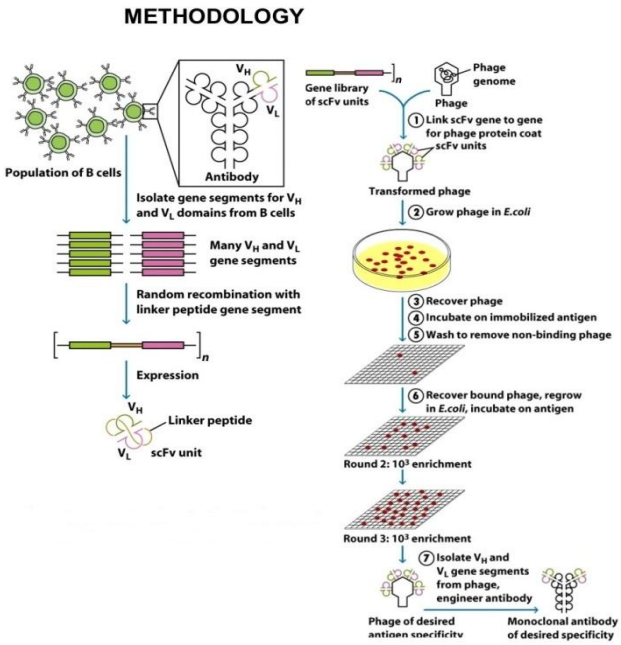


Figure 1.2: Methodology for the construction and application of a phage display library (Osborne *et al*., 2007).

Phage display of antibody fragments as a method aims to immortalize the gene that codes for an antibody. This is in contrast to the tissue culture method of hybridoma formation that aims to immortalize the cells which produce antibodies. Phage display is an RNA based technology, which therefore allows access to any B-cell compartment in which there is mRNA coding for immunoglobulins. This mRNA can therefore be obtained from resting B-cells, activated B-cells as well as plasma cells; the latter cannot be used for hybridoma formation (Siegel *et al*., 2002).

To produce phage displayed antibody fragments the mRNA of rearranged Ig genes is obtained from donor B-cells. From the mRNA, cDNA is prepared from which the heavy and light variable chain gene segments are amplified by PCR and cloned into a phage display plasmid vector. Plasmid and M13 helper phage co-transfected *Escherichia coli* cells produce phagemid particles. Inside these bacteriophages is the DNA for the antibody fragment and on their outside is the antibody fragment protein fused to a minor coat protein. By a process of ‘panning’ libraries of such bacteriophages are incubated with antigen to obtain specific binders. Repeated rounds of panning are used to obtain binders with increased specificity to the target antigen (McCafferty *et al*., 1990, Kang *et al* and Siegel *et al*., 2002).

The antibody genes required for a phage display library can be obtained from three different kinds of sources. Immunized human or animal sources can be used to obtain antigen specific phages. Non immunized ‘naive’ are thought to provide antibodies with a wider range of antigen specificities. Synthetic libraries are B-cell independent as they use genes generated by PCR randomization of germline CDRs. Phage display is the oldest of the display technologies available today. Newly developed techniques include RNA and ribosome display. In the cell surface display technologies the proteins are displayed as cell surface molecules of either *Escherichia coli* or yeast cells (Siegel *et al*., 2002).

Application of single chain variable fragments

Due to their characteristic as binding molecules monoclonal antibodies are finding more clinical (diagnostic and therapeutic) applications in addition to their established roles as research tools (Holliger *et al*., 2005). The usefulness of monoclonal antibodies as diagnostic and therapeutic reagents can be attributed to their high directional binding to the respective antigens. This directional binding, results in a reagent which is guaranteed to localise to its target. A benefit to phage displayed scFv's is that adaptation for further application requires only subcloning into the appropriate vector. Some diagnostic procedures such as enzyme linked immunosorbent assay and immunocytochemistry require a high affinity from the scFv. This is to prevent loss during washing steps since scFv's consist of only one binding site (Smith *et al*., 2003).

For clinical applications scFv's are only one of the many antibody formats that are in use. ScFv's however are advantaged in their smaller size, recombinant expression in microbial systems as well as their improved pharmacokinetics as compared to whole antibodies. In addition, in some clinical applications the effector functions as induced by the Fc domain of a whole antibody is unnecessary or unwanted. The size, shape and high affinity on intact whole antibodies may even slow down entry into targets such as solid tumors. Intact antibodies may also be limited by their long serum half lives caused by the difficulty in clearing them from circulation. This may be detrimental to surrounding tissues during radioimmunotherapy of cancerous tumors (Hudson *et al*., 2003). However the short serum half life of scFv's may also be detrimental in other therapeutic applications. In therapeutics a prolonged serum half life results in longer bioavailability and fewer treatments required. A simple method of replacing serum half life increasing properties of the Fc domain is by linkage of polyethylene glycol to the scFv (Hudson *et al*., 2003, Kubetzko *et al*., 2006 and Krinner *et al*., 2006).

Bivalency of intact IgG antibodies may increase binding strength due to avidity, but it can also cause the binding of two separate antigens by the same antibody (crosslinking). Thus for cell surface molecules this could lead to internalization, proliferation or apoptosis depending on the target molecule. Thus instead of blocking ligand binding the antibody may replace it. Producing antibodies in monovalent fragment formats such as scFv's are the simplest way of avoiding such complications (Labrijn *et al*., 2008).

Simple proteolysis of whole antibodies does yield smaller Fab fragments but no smaller and cannot compete with recombinant expression in terms of production. The scFv format can be expressed by all expression systems, but since it requires no glycosylation bacteria are favored. Intact whole antibodies however require more complex mammalian or plant cell expression systems (Hudson *et al*., 2003).

Molecular biological techniques and recombinant microbial expression give scFv's a distinct advantage when it comes to their use as chimeric immunoconjugates (Hudson *et al*., 2003, Holliger *et al*., 2005 and de Marco., 2011). ScFv's have been fused to a number of molecules providing secondary functions. These include radionuclides, cytotoxic proteins, liposomes, proteins, enzymes and viruses for gene therapy (Holliger *et al*., 2005, Adams *et al*., 2004, Krauss *et al*., 2004, Park *et al*., 2002, van Beusechem *et al*., 2002, Rippmann *et al*., 2000, Kousparou *et al*., 2002 and Hudson *et al*., 2003). Molecular biology also allows for the often observed lower affinities of scFv's to be overcome by display technologies. If required the monovalency of a scFv may also be overcome by constructing bi, tri or tetrabodies, which are fusions of single scFv's (Hudson *et al*., 2003 and Holliger *et al*., 2005). Another development is that of intrabodies which are scFv's that are expressed inside patient cells to deliver therapy. The technology is however currently limited by the delivery of the genetic material that codes for the scFv to the cell (Gavilondo *et al*., 2000 and Greenman *et al*., 1996).

An important factor to consider in the use of antibody fragments in therapeutics is the immunogenicity of the protein. Monoclonal antibodies derived from murine sources require humanization in which the CDR domains are grafted onto human antibody structures (Knight *et al*., 1995). An alternative is the use of transgenic murine donors. ScFv's have a clear advantage to whole antibodies and Fab fragments due to the absence of constant domains in scFv's. However scFv's derived from non human sources may still require humanization. Fully human antibodies or fragments may still be able to illicit an immunogenic response. This should all be taken into account during the design and development of an antibody or fragment for use in therapy. An immunogenic response against the therapeutic agent will clearly affect its usefulness in treatment (Gavilondo *et al*., 2000, Schroff *et al*., 1985, Schneider *et al*., 1993 and Iwahashi *et al*., 1999).

Unlike immunotherapy there is not necessarily a need for human(ized) antibodies in immunodiagnostics. The possible immunogenicity of an antibody or fragment is not a factor in diagnostics as it is in therapy. As a result hybridoma technology is sufficient for assay developments. However as with therapeutics antibody engineering and display technologies can contribute in assay development by affinity and specificity adjustment as well as expression of antibody-marker fusions (Borrebaeck., 2000). In addition the display technologies allow for the generation of high affinity scFv's against possibly any target molecule *in vitro* (Hanes *et al*., 1997, Dillon *et al*., 2003 and Beukes 2010). ScFv's have the advantage over Fab's that they are better expressed in *Escherichia coli* and can be multimerized to in order to obtain avidity for an antigen (O'Kennedy *et al*., 2010). In non assay diagnostic applications scFv-quantum dot fusions were used for imaging of single protein molecules extracellularly (Iyer *et al*., 2008).

The ease with which scFv's are manipulated has proven them superior to intact IgG in immunoassays requiring immobilized antigen receptors. ScFv's can easily be expressed with tags that allow for immobilization on surfaces. Uniform directionality and denser packing of immobilized scFv's was shown to lead to a significant increase in the sensitivity of assays (Backmann *et al*., 2005, Arntz *et al*., 2003, Peluso *et al*., 2003, Karyakin *et al*., 2000, Kanno *et al*., 1999, Wingren *et al*., 2005). Purification tags usually fused to scFv's to remove the protein from crude lysates post expression, have also been used for immobilization on biosensor surfaces (Conroy *et al*., 2009, Scibek *et al*., 2002, Lori *et al*., 2005, Mersich *et al*., 2007, Piervincenzi *et al*., 1997, O'Kennedy *et al*., 2010).

Biosensor based immunodiagnostic assays using scFv's against a variety of analytes have been developed. From chemicals, drug metabolites, and infectious diseases have been detected by biosensor assays using scFv's (Grennan *et al*., 2003, Horacek *et al*., 1998, Dunne *et al*., 2004, Dillon *et al*., 2003, Hu *et al*., 2004, Encarnacao *et al*., 2007, Love *et al*., 2008 and Nanduri *et al*., 2007).

Polyclonal and hybridoma derived monoclonal antibodies are the traditional sources of reagents for immunodiagnostic devices. Both of these are however limited in their diagnostic usefulness. Polyclonal antibodies need to be continually isolated from immunized animals and heterogeneity between batches is a real possibility. The polyclonal nature of these antibodies also increases the chances of cross-reactivity. Hybridomas are derived from mice and so it may be difficult in raising antibodies towards conserved mammalian proteins. The performance of an immunodiagnostic device or assay is dictated by the specificity and affinity of the applied antibody. It has been argued that if the affinity is sufficiently high, specificity towards the target antigen is increased. Consequently the assay for a particular antigen is simplified. Another important point is that antibodies that are commercially available cannot necessarily be forced into an assay. Rather an antibody or fragment should be designed and developed according to the requirements of the assay application (O'Kennedy 2010).

The Nkuku phage display library

Chicken immunoglobulin genes are organized in a manner that allows for simple exploitation in the construction of scFv's. The light chain of the chicken Ig gene is generated by rearrangement of a single V gene and a single J-C unit. The heavy chain arrangement is the same except that it has multiple D genes in addition to the V, J and C genes. Once rearrangement of these single genes has occurred pre-immune diversity is obtained by the insertion of segments from pseudogenes found upstream of the single V gene. These segments replace 10 to 120 bp sequences within the rearranged immunoglobulin gene. Unlike in mammals such as mice and humans rearrangement only occurs during early development in chickens. Rearrangement occurs while stem cells colonize the Bursa of Fabricius while pseudo gene insertion operates while B cell proliferation occurs. Therefore in chickens the 5' and 3' termini of both the heavy and light chain genes are conserved. The result is that only a single primer pair is required for reverse transcriptase PCR of mRNA V regions of either heavy or light chains (Horton *et al*., 2001 and van Wyngaardt *et al*., 2004).

The Nkuku phage display library was constructed as a source of scFv's for *in vitro* diagnostic methods. The scFv's were constructed from natural chicken H and L chains as well as selected synthetically randomized CDR3s in some of the H chains. From this library scFv's that recognize haptens, proteins and virus particles were obtained (van Wyngaardt *et al*., 2004).

Single chain variable fragments and protein engineering

The development of scFv's was hailed as a feat of protein engineering (Bird *et al*., 1988 and Huston *et al*., 1988). This was because it consisted of taking what was provided by nature (antibody binding domains) and adjusted it to a format that was more agreeable to manipulation. To date the protein engineering of antibodies can refer to humanization by CDR grafting (Kuegler et al., 2009 and Presta., 2006). As well as the combination of directed evolution and display technologies to obtain functionally improved mutants. The directed evolution mutates the protein sequence creating a library which is then displayed screening for improved function antigen binders (Miyazaki *et al*., 1999, Daugherty *et al*., 2000 and Drummond *et al*., 2005).

The engineering of antibodies has been a strong motivator for the development of many protein engineering techniques. Despite protein engineering having its origins in the study of enzyme structure and function, complementarity determining region (CDR) grafting and phage display of antibody fragments are considered important milestones. From the start mutagenesis, specifically site directed mutagenesis was considered a primary tool in the study of protein function. This is because the technique allows for the manipulation of proteins at the amino acid residue level. In turn this allowed for the study of the effects of certain residues on the structure and function of proteins. However in order to use site directed mutagenesis effectively detailed crystallographic and kinetic data of the protein in question as well as the location of contacts with ligands was required. Combinatorial or directed evolution in combination with high through put selection methods would however be more agreeable to the designing of properties and novel functions (Brannigan *et al*., 2002).

Monoclonal antibodies are increasingly becoming important biopharmaceutical reagents. Substantial molecular modification has gone into the development of those antibodies that have found clinical application (Smith *et al*., 2004).

There are two separate approaches to the generation mutant libraries. The first approach seeks to insert point mutations at random positions because many function modifying mutations occur away from active sites. The alternative approach is focused on the randomization of amino acid residues within the active site that is in direct contact with the ligand/substrate (Bershtein *et al*., 2008).

The former of the two approaches is referred to as random mutagenesis. Random mutagenesis is thought of as a directed evolution method either on its own or in combination with a selection/screening method. Natural evolution generates a wide spectrum of protein variants, however only for a physiological function. The aim of random mutagenesis is to act as an accelerated synthetic evolution in order to sequence diversity in a timeframe more suited for biotechnology. Subsequently the properties sought by biotechnology from a particular protein are more easily accessible (Wong *et al*., 2006, Neylon, 2004 and Labrou, 2010).

Mycolic acids

Mycolic acids are a major lipid component in the cell walls of the *Mycobacterium tuberculosis* bacterium and related genera. These long chain, high molecular weight, lipids are structurally related consisting of a hydrophobic merochain and a more polar mycolic motif. Mycolic acids are classified according to the functional groups present on the meromycolate chain. In *Mycobacterium tuberculosis* the subclasses of mycolic acids are alpha, keto and methoxy referring to the functional groups on the respective merochains. Mycolic acids have previously been identified as being immunogenic despite their hydrophobicity. In addition 11 anti-mycolic acid scFv antibodies have been obtained by phage display from the Nkuku phage display library (Beukes et al., 2010, Pan et al., 1999 and Fujiwara *et al*., 1999).

Aims

In chapter 2, plasmids pGE20 and pAK400 were investigated as subcloning vectors for the MAC10 scFv gene.

In chapter 3, a limited study was undertaken into expression of the MAC10 scFv gene from plasmids pMAC10 and pAKJS using the *Escherichia coli* TG-1 and HB2151 strains.

In chapter 4, error prone PCR was investigated as a directed evolution method to randomize the MAC10 scFv gene sequence.

Chapter 2 : Sub-cloning of anti-mycolic acid single chain fragment variable gene from pHEN-1 to pGE20 and pAK plasmid vectors.

Introduction

The Nkuku phage display library was constructed to be a source of monoclonal scFv's, specifically for diagnostic purposes. For the construction of the phage display library, the chicken derived scFv genes were cloned into the pHEN-1 plasmid. Plasmid vector pHEN-1 (figure 2.1A) was developed by inserting the fd-gene 3 coding region from the fd-tet—DOG1 plasmid vector (figure 2.1B) into pUC119. The signal sequence gene 3 protein coding region was replaced with the *Erwinia carotovora* bacterium pectate lyase B *(pelB)* leader sequence. The *pel*B leader sequence is a 22 amino acid residue peptide responsible for the translocation of *pel*B proteins into the periplasmic space when expressed in *Escherichia coli*. Bacterial proteins capable of crossing membranes commonly possess such leader sequences. Transport to the periplasmic space is required for the correct folding and association of the antibody fragment chains. Thus fusions of scFv proteins with the *pel*B leader sequence peptide allow their transport into the periplasmic space. A *c-myc* peptide tag sequence was inserted after the *Sfi* I-*Not* I cloning site. This is followed by an amber stop codon which allows for the cloned scFv to be expressed as either a phage surface displayed or a soluble protein. The plasmid pHEN-1 also possesses an ampicillin resistance gene, an *Escherichia coli* origin of replication, M13 bacteriophage origin of replication and a β-galactosidase gene promoter (van Wyngaardt *et al*., 2004 and Hoogenboom *et al*., 1991, Short *et al*., 1988 and Lei *et al*., 1987).

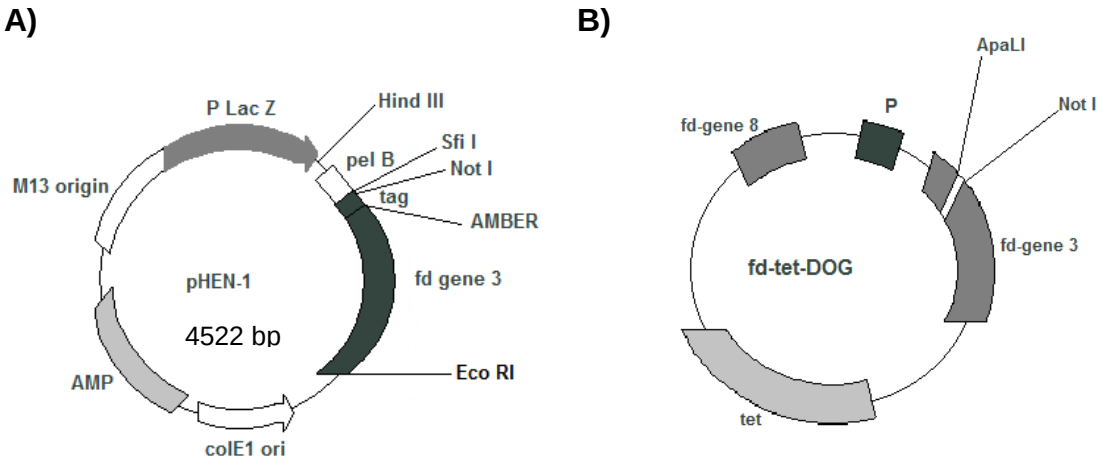


Figure 2.1: Plasmid map of the phage display vectors pHEN-1 and fd-tet-DOG. A) Plasmid pHEN-1, ColE1 ori, origin of replication; AMP, ampicillin resistance gene; M13 origin, origin of replication for phage display; P lacZ, β-galactosidase gene promoter; Hind III, *Sfi* I and *Not* I, restriction sites; PelB, pectate lyase signal peptide sequence; tag, *c-myc* peptide; AMBER, amber (TAG) stop codon; fd-gene III, fd phage coat protein 3 gene; *Eco* RI, restriction site. B) Plasmid fd-tet-DOG, tet, tetracycline resistance gene; fd-gene 8, fd bacteriophage coat protein 8; P, promoter; *Apa*LI and *Not* I, restriction sites; fd-gene 3, fd bacteriophage coat protein 3 (Adapted from Hoogenboom *et al*., 1991).

The pGE20 plasmid was designed as an expression vector for Fab fragments. This plasmid was derived from a construct obtained by fusing the lambda bacteriophage vectors λHC2 and λLC1. The lambda bacteriophage vectors λHC2 and λLC1 in turn are derivatives of the lambda bacteriophage vector λZapII. The λHC2 and λLC1 vectors were each for the cloning of either antibody variable heavy (V_H) and variable light (V_L) chain libraries respectively. Once they were fused a combinatorial construct was obtained that contained all the factors required for the expression of Fab fragments. The bacteriophage lambda vector λZapII contains within it the sequence for the pBluescript SK- vector. It is from this pBluescript SK- (2958 bp) that pGE20 is essentially derived. Consequently the two vectors share many characteristics including an *Escherichia coli* origin of replication, a β-lactamase gene that grants resistance to ampicillin, a bacteriophage f1 intergenic region/origin of replication and a β-galactosidase (Lac Z) gene (Huse *et al*., 1989, Short *et al*., 1988 and Orfanoudakis *et al*., 1993).

The alpha subunit of β-galactosidase or Lac Z gene allows for what is termed blue-white screening in which recombinant plasmids can be differentiated from non recombinants. The cloning sites for foreign DNA are located within the Lac Z gene and therefore cloning would result in disruption of the expression of the alpha subunit β-galactosidase enzyme. Growing transformed *Escherichia coli* cells on media containing isopropyl- β-D-thiogalactosidase (IPTG) and 5-bromo-4-chloro-indolyl- β-D-galactopyranoside (X-Gal) allows for this screening method. The β-galactosidase enzyme catalyses the hydrolysis of the bond between the O1 of β-D-galactose sugar and a substituent. The X-Gal compound is colorless, but after hydrolysis by the β-galactosidase enzyme β-D-galactose and 5-bromo-4-chloro-3-hydroxyindole are produced. With the latter being responsible for the blue color observed for non recombinant colonies. The IPTG molecule acts as an inducer of the Lac Z promoter and will be introduced in chapter 3 (Short *et al*., 1988 and Voet *et al*., 2004).

Plasmid pGE20 is different from pBluescript SK- in that two pelB leader sequences had been inserted upstream of the *Xho* I – *Spe* I and *Sac* I – *Xba* I cloning sites. These pelB leader sequences originated from λHC2 and λLC1. Also downstream of the *Xho* I – *Spe* I cloning site is a sequence that codes for a 15 amino acid peptide. This peptide is part of the human oestradiol receptor (hER), and is recognized by a monoclonal antibody B10. This allows for the peptide to act as a purification tag (Orfanoudakis *et al*., 1993).

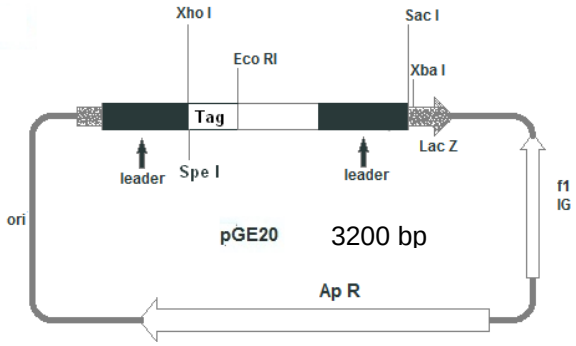


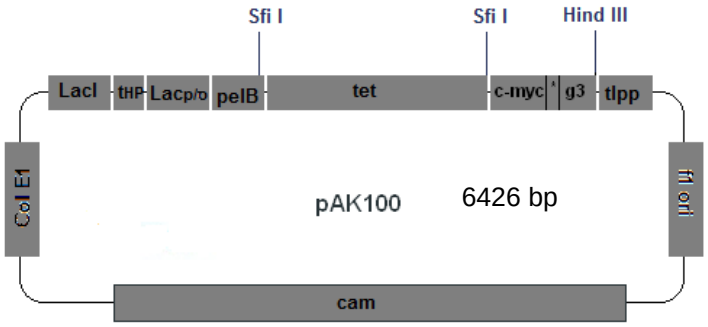
Figure 2.2: Vector map of plasmid pGE20. ApR, ampicillin resistance gene; ori, ColE1 origin of replication; LacZ, β-galactosidase alpha subunit gene; leader, pectate lyase signal peptide sequence; *Xho* I, *Spe* I, *Eco* RI, *Sac* I and *Xba* I restriction sites; tag, B10 tag peptide; f1 IG, intergenic region (origin of replication) of phage f1 (Adapted from Orfanoudakis *et al*., 1993).

The pAK series of plasmid vectors were designed to be highly regulatable vectors for scFv phage display and soluble expression. The main plasmid from which all others in the series were derived is pAK100. Plasmid pAK100 is a phage display vector containing many elements for the efficient display and expression of scFv proteins. These include an *Escherichia coli* origin of replication, a *Lac* repressor gene, a *LacZ* promoter, a *pelB* leader sequence, a *Sfi I-Sfi I* cloning site which encloses a tetracycline resistance cassette, a *c-myc* tag, an amber stop codon, an M13 gene 3, a bacteriophage f1 origin of replication and a chloramphenicol resistance gene (Krebber *et al*., 1997).

The tetracycline resistance cassette (tetA and tetR) as well as the *Sfi* I sites that enclose it allow for a simple cloning and selection process for the pAK plasmids. Firstly the *Sfi* I restriction enzyme is unique in that it recognizes only 8 out of 13 of the nucleotide bases in the restriction sites. The 5 bases in the middle of the restriction site sequence can therefore differ thus ensuring that directional cloning occurs. Also only a single restriction enzyme is required to completely digest the pAK plasmids. Secondly the 2101 base pair tetracycline cassette allows for simple identification of *Sfi* I restriction digested pAK plasmids due to the large size of the cassette. A simple agarose gel electrophoresis analysis would allow for the separation of digested from undigested plasmid. In addition clones can be confirmed as recombinant by screening them for tetracycline resistance. This is because recombinant plasmids would have lost resistance to the tetracycline antibiotic with the cutting out of the cassette (Krebber *et al*., 1997).

The pAK400 plasmid (figure 2.3) is a derivative pAK100 that has been enhanced for the soluble expression of scFv's. Since the expressed protein will not be used for phage display the gene 3 sequence has been removed. Also the *c-myc* fusion peptide sequence has been replaced by a sequence coding for a polyhistidine tag. The features of pAK400 will be introduced in chapter 3 as they relate to protein expression (Krebber *et al*., 1997).

A)



B)

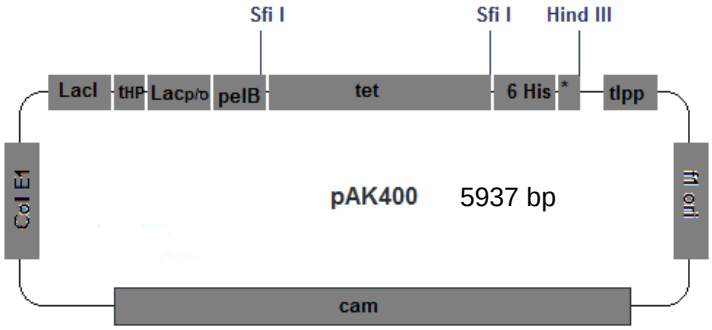


Figure 2.3: Vector maps of the plasmids pAK100 and 400. A) pAK100 plasmid vector; cam, chloramphenicol resistance gene; Col E1, origin of replication; LacI, repressor gene; t_{HP} ,strong upstream terminator; $Lac_{p/o}$, *lac* promoter/ operator; *pelB*, pectate lyase leader sequence; *tet*, tetracycline resistance cassette; *Sfi* I and *Eco* RI, restriction sites; *myc*, *c-myc* peptide; *, amber stop codon; g3, phage coat protein 3 gene; *Hind* III, restriction site; t_{lpp}, downstream terminator; f1 ori, phage f1 origin. B) Changes to pAK100 plasmid for pAK400; 6 His, His-tag peptide gene and *Hind* III, *Hind* III restriction site (Adapted from Krebber *et al*., 1997).

For some time now many expression systems have been available that allows for the inducible expression of a desired protein with an affinity tag for purification, and more are always being developed (Sheibani, 1999). Polyhistidine tags (His-tags) consist of six histidine residues attached to either the N- or C-terminal of the expressed protein. Of all affinity tags these His-tags are currently the most widely used (Waugh, 2005 and Arnau *et al*., 2006). Affinity tags allow for the purification of the expressed protein from extracts of the host cell due to their affinities for certain ligands. The benefit to using such an affinity tag to obtain the expressed protein is that little previous knowledge of the biochemical properties of the expressed protein is required (Arnau *et al*., 2006).

Nested PCR is a variation of the standard reaction in which two sets of primers are used in the amplification of a desired section of DNA. The reaction is divided in two with 15 to 30 cycles performed with the first set of primers. Then another 15 to 30 cycles is performed with the second set of primers with the DNA amplified by the first round as template. The second set of primers is internal, that is downstream in the 5' to 3' direction of the first set in the sequence to be amplified. The desired sequence is subsequently obtained with high yield and the amplification of secondary PCR products is minimized. Using nested primers during PCR grants another level of specificity to the reaction as well as efficiency by reducing the chance of non specific primer annealing (Haqqi *et al*., 1988 and Bej *et al*., 1991).

In this chapter the subcloning of the MAC10 scFv gene from one plasmid vector (pMAC10, 5245 bp) into two others (pGE20 and pAK400) was investigated. The plasmid vector pMAC10 is the recombinant derivative of pHEN-1 containing the MAC10 scFv gene sequence. Also methods were investigated to ensure an efficient and fast subcloning procedure.

Materials and Methods

Flow diagram of subcloning methods

Summarized below into flow diagrams and figures (2.4 to 2.7) are the alternative procedures followed to subclone the MAC10 scFv gene into the pGE20 and pAK400 plasmids. At first the MAC10 gene and His-tag gene oligonucleotide were to be cloned separately into pGE20 plasmid (figure 2.4). Then only the His-tag gene oligonucleotide was to be cloned into pGE20 (figure 2.5). Lastly for pGE20, the MAC10 and His-tag genes were to be ligated together before being cloned into the plasmid (figure 2.6). Alternatively the MAC10 gene was subcloned into the pAK400 (figure 2.7). The procedure for subcloning into pAK400 is also applicable for cloning into pAK100.

Subcloning of separate mycolic acid clone 10 single chain fragment variable and polyhistidine-tag genes into pGE20

Xho I restriction digestion of pGE20 → Klenow fragment blunting of *Xho* I site

↓

Ligation of MAC10 PCR amplicon ← *Eco* RI restriction enzyme digestion of pGE20

↓

Eco RI restriction enzyme digestion of pGE20 → Ligation of His-tag oligonucleotide to pGEJS

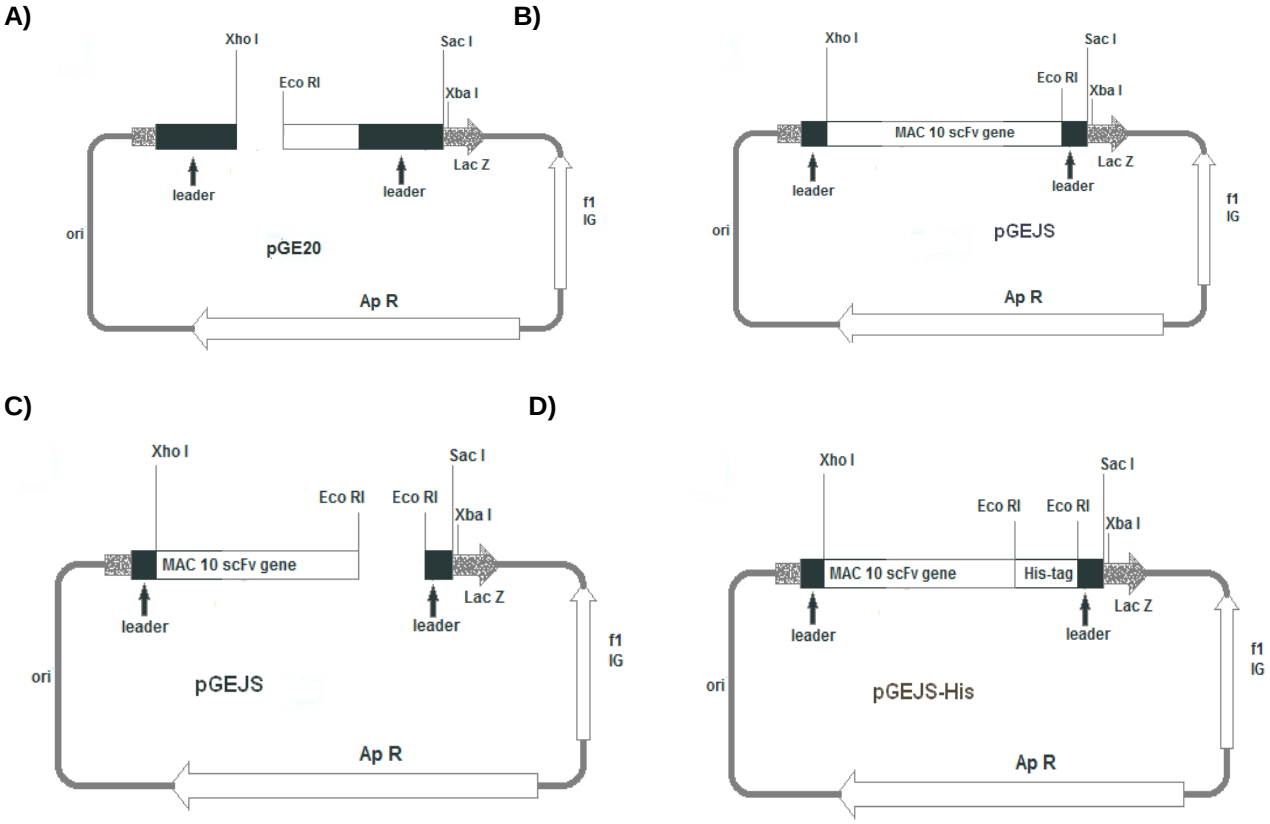


Figure 2.4: Ligation of the MAC10 scFv gene and His-tag oligonucleotide into the pGE20 expression plasmid vector. A) *Xho* I and *Eco* RI restriction enzyme digestion of pGE20. B) Recombinant pGE20 (pGEJS), produced by ligation of the MAC10 scFv gene into pGE20. C) *Eco* RI restriction enzyme digested pGEJS. D) The pGEJS-His plasmid vector containing both the MAC10 scFv gene and His-tag oligonucleotide.

Cloning of polyhistidine-tag gene oligonucleotide into pGE20

Eco RI and *Xba* I restriction enzyme digestion of pGE20 → Ligation of His-tag oligonucleotide

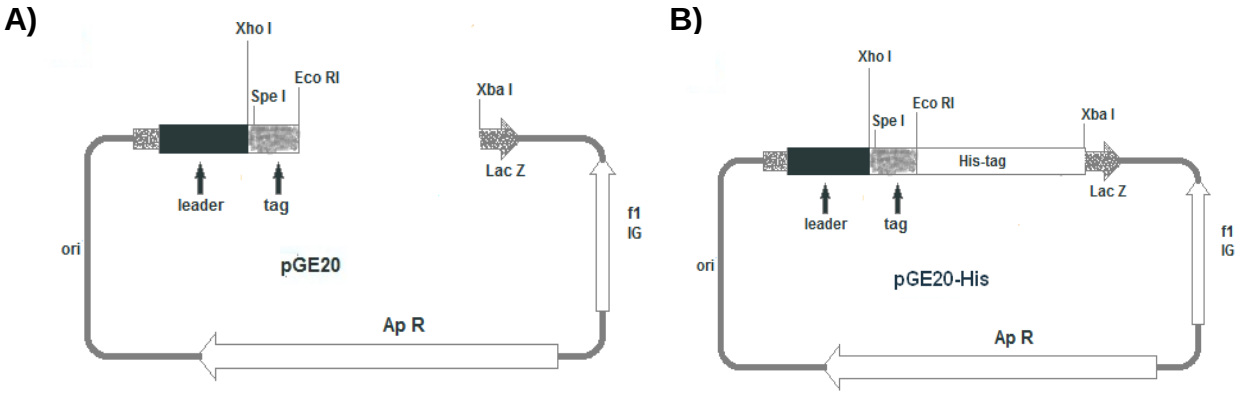


Figure 2.5: Construction of the plasmid vector pGE20-His for His-tagged scFv expression. A) *Eco* RI and Xba I restriction enzyme digested pGE20. B) The pGE20-His plasmid obtained by ligating the His-tag oligonucleotide to pGE20.

Cloning of ligated mycolic acid clone 10 single chain fragment variable and His-tag gene oligonucleotide into pGE20

Ligation of MAC10 PCR amplicon and His-tag oligonucleotide → *Xho* I and *Xba* I

Digestion of pGE20

↓

Ligation of insert into pGE20

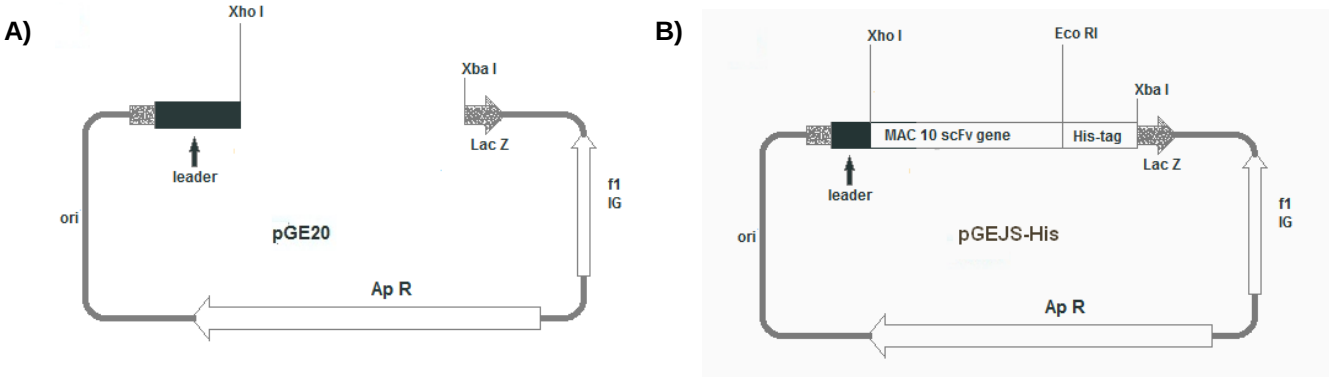


Figure 2.6: Ligation of pre-ligated MAC10 scFv gene and His-tag oligonucleotide into *Xho* I and *Xba* I restriction enzyme digested pGE20 to obtain pGEJS-His in fewer steps. A) *Xho* I and *Xba* I restriction enzyme digested pGE20. B) pGEJS-His after ligation of pre-ligated MAC10 scFv gene and His-tag oligonucleotide into pGE20.

Subcloning of mycolic acid clone 10 single chain fragment variable gene into pAK400

Double digestion of pAK400 with *Sfi* I → Ligation of MAC10 scFv gene into pAK400

↓

pAKJS

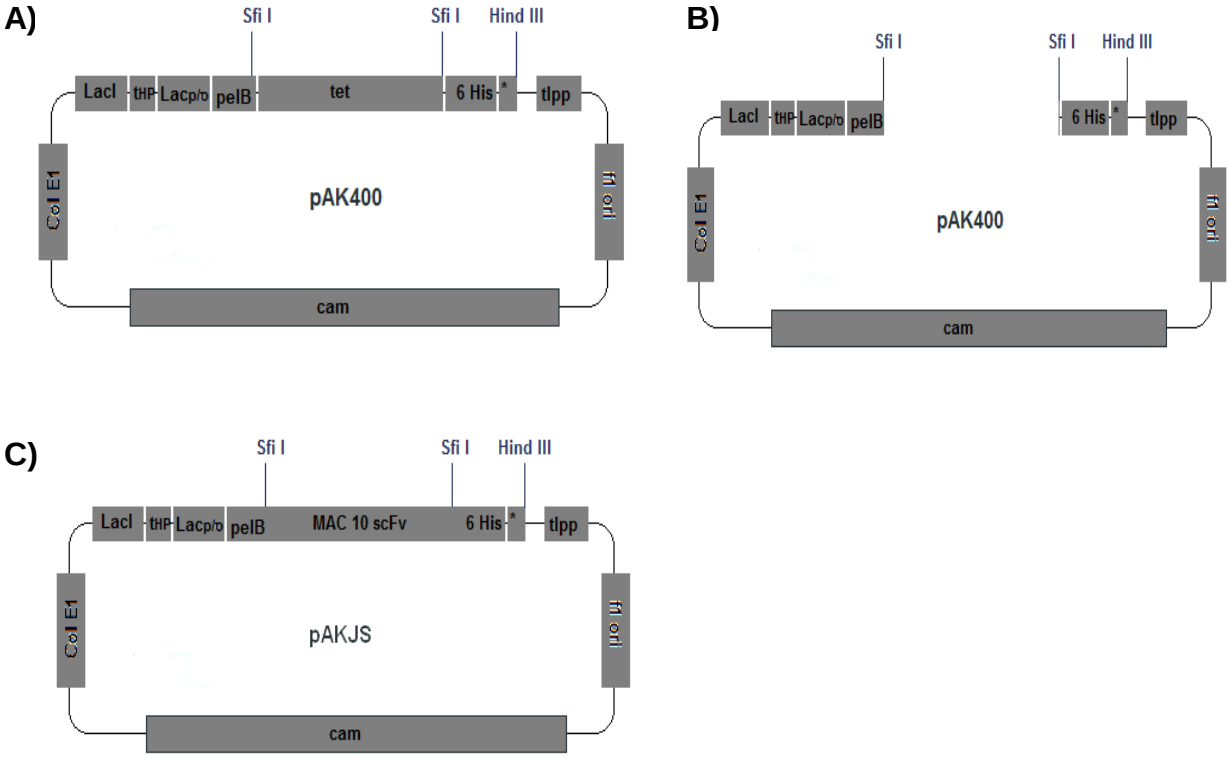


Figure 2.7: Ligation of the MAC10 scFv gene into the pAK400 plasmid vector between the two distinct *Sfi* I sites. A) Undigested pAK400 plasmid vector. B) *Sfi* I double digested pAK400. C) The plasmid pAKJS obtained by ligating the MAC10 scFv gene into the *Sfi* I double digested pAK400.

Preparation of chemically competent *Escherichia coli*

For the preparation of chemically ($CaCl_2$) competent *Escherichia coli* a previously described method was followed (Ausubel *et al*., 2002). Briefly, competent *Escherichia coli* cells (TG1, HB2151 and XL-1 blue) were prepared from an overnight culture (16 h at 37°C) shaking at 200 rpm. This was obtained from a single colony picked off an LB agar plate (1% tryptone, 0.5% yeast extract, 1% NaCl and 1.4% (w/v) agar) into liquid LB (10 ml). Of this culture 1 ml was diluted 100 x and grown with shaking (200 rpm) until an OD_{550} of approximately 0.6 was reached. The 100 ml culture was split into two 50 ml tubes and centrifuged (11 900 rcf, 20 – 25 °C for 8 min). The cell pellets were re-suspended and combined in ice cold 100 mM $CaCl_2$ in a single tube. The cells were again pelleted by centrifugation (11 900 rcf, 20 – 25 °C for 8 min). The resultant pellet was re-suspended in 50 ml ice cold 100 mM $CaCl_2$ and left on ice overnight at 4°C. The cells were subsequently pelleted by centrifugation (11 900 rcf, 4°C for 8 min). The cell pellet was then re-suspended in 86% (v/v) ice cold $CaCl_2$ and 14% (v/v) glycerol. This solution was aliquoted into 0.2 ml volumes in pre-cooled (4°C) eppendorf tubes and stored at -70°C until needed.

Transformation of chemically competent *Escherichia coli*

To transform competent *Escherichia coli* cells (TG1, HB2151 and XL-1 blue strains) approximately 100 ng of plasmid DNA was added to 50 µl of competent cells. The cell and DNA mixture was incubated on ice for 30 min. Following this, the cells were heat shocked (42 °C for 2 min). Pre- warmed (37 °C) LB medium (1 ml) was added immediately to the cells which were then incubated in a water bath (37 °C) for 1 h. After incubation, 100 µl aliquots of transformed cells was plated on either LB-ampicillin (100 µg/ml ampicillin) agar or non expression (NE) (1.6% tryptone, 1% yeast extract, 0.5% NaCl, 1% glucose and 25 µg/ml chloramphenicol) plates, which were incubated overnight (37°C) depending on the plasmid antibiotic resistance gene.

Overnight culture of transformed *Escherichia coli*

Cultures were obtained by selecting single *Escherichia coli* colonies from LB-ampicillin (100 µg/ml) or NE (25 µg/ml chloramphenicol) agar plates and inoculating them into liquid LB-ampicillin (100 µg/ml) or NE (25 µg/ml chloramphenicol) medium. The cells were incubated (37°C) for 16 h with shaking at 200 rpm.

Plasmid isolation using the Macherey-Nagel NucleoSpin® Plasmid kit

For isolation of both non recombinant and recombinant plasmid DNA (pHEN-1, pGE20 and the pAK plasmids) the specifications of the Macherey-Nagel NucleoSpin® Plasmid kit were followed. In short, 5 ml (5 x 1 ml) of an overnight culture was centrifuged for 30 s the pellet was kept and the supernatant was discarded. Thereafter 250 µl of buffer A1 was added to the pellet which was mixed by finger tapping until a uniform solution was obtained. To this 250 µl of buffer A2 was added. The sample tubes were inverted 8 times and the sample incubated at room temperature (RT) for a maximum of 5 min. To this 300 µl of buffer A3 was added, the solution was again mixed by inverting the sample tube 8 times. The cell debris was pelleted by centrifuging at 11 000 rcf for 5 min in a benchtop centrifuge. For the binding step the supernatant was loaded onto a NucleoSpin® plasmid column and centrifuged for 1 min at 11 000 rcf. After the flow-through had been discarded a two part washing step was followed. In the first part, 500 µl of pre-warmed buffer AW (50°C for 30 min) was added to the column followed by centrifugation at 11 000 rcf. In part two 600 µl of buffer A4 was added to the column which was centrifuged for 1 min at 11 000 rcf to wash the membrane. The column was centrifuged again (5 min at 11 000 rcf) to remove any remaining ethanol. In the elution step bound plasmid DNA was eluted 2x with 30 µl of AE buffer which was incubated on the membrane for 3 min followed by centrifugation for 1 min at 11 000 rcf.

Plasmid isolation using the Zyppy™ plasmid miniprep kit

The specifications of the Zyppy™ plasmid miniprep kit were followed. From an overnight (16 h) culture (5 ml) 4 ml was pelleted. The pellet was then re-suspended in 600 µl of culture or 5 mM Tris-HCl buffer (pH 7.5). To this, 100 µl of 7 x lysis buffer (blue color) was added and mixed with the cell solution by inverting the sample tube 6 times. To this 350 µl of cold neutralization buffer (yellow color) was added and mixed until a thick yellow precipitate had formed and all blue color had disappeared. After neutralization the samples were centrifuged for 4 min at 11 000 rcf to pellet the cell debris. The 900 µl of supernatant was loaded onto a Zymo-Spin™ IIN column and centrifuged for 20 s at 11 000 rcf, the flow-through was discarded. For the washing step 200 µl of Endo-wash buffer was added to the column followed by 20 s of centrifugation at 11 000 rcf. This was followed by adding 400 µl of Zyppy™ wash buffer and 30 s of centrifugation at 11 000 rcf. During the elution step, the plasmid DNA bound to the column was eluted with 30 µl of Zyppy elution buffer and 30 s of centrifugation at 11 000 rcf after a 1 min RT incubation.

Restriction enzyme digestion of plasmid pGE20

The plasmid pGE20 was digested with the restriction enzymes *Xho* I, *Eco* RI and *Xba* I, either as single digests or as *Eco* RI/ *Xba* I double digest. For single digests 3 µg of plasmid DNA was digested by 2 Units of FastDigest® enzyme with 10 x FastDigest® buffer (2 µl) made up to 20 µl with triple distilled water (dddH_2O). For double digests 2 Units of each enzyme were used, but it was ensured that the combined volumes did not exceed 1/10 that of the reaction volume. Approximately 1 µg of plasmid was digested in a 40 µl volume containing 10 x FastDigest® buffer (4 µl) made to volume with dddH_2O. The reactions were incubated for 1 h, except those with *Eco* RI which were incubated for a maximum of 20 min to prevent star activity, at 37°C.

Table 2.1: **Restriction digest of pGE20 with one FastDigest® enzyme.**

Component	Amount
Plasmid DNA	3 µg
10 x FastDigest® buffer	2 µl
FastDigest® enzyme	2 U
$ddd H_2O$	Reaction volume up to 20 µl

Table 2.2: Restriction digest of pGE20 with two FastDigest® enzymes.

Component	Amount
Plasmid DNA	1 µg
10 x FastDigest® buffer	4 µl
FastDigest® enzyme	2 U
FastDigest® enzyme	2 U
$ddd H_2O$	Reaction volume up to 40 µl

Blunting of the *Xho* I 3'- overhang in plasmid pGE20 with Klenow fragment

To fill in the 3'-overhang produced by *Xho* I digestion of pGE20; a reaction was set up containing the digested plasmid, 60 µM of each dNTP and 5 units of Klenow fragment enzyme in 1 x reaction buffer (50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$ and 1 mM DTT) and $dddH_2O$ made up to a volume of 20 µl. The reaction was incubated for 2 h at 37°C. The enzyme was removed by inactivation at 75°C for 10 min and followed by column purification.

Table 2.3: Composition of Klenow blunting reaction of *Xho* I restriction digested pGE20.

Component	Amount
digested plasmid	1 µg
dNTPs	60 µM of each
1 x reaction buffer	2 µl
Klenow fragment enzyme	5 units
$dddH_2O$	Reaction volume up to 20 µl

Polymerase chain reaction

The amplicon of the mycolic acid clone 10 (MAC10) scFv gene was obtained using a previously designed protocol (van Wyngaardt *et al*., 2004). Two separate primer sets were used for the amplification.

Table 2.4: Oligonucleotide primers used for the amplification of the mycolic acid clone10 scFv gene from plasmid pMAC10 template.

Primer	Sequence	Restriction enzyme site	Tm (°C)[a]	Tm (°C)[b]
MB1	TGA TGG TG**G CGG CCG C**AT TGG GCT G	*Not* I	84	71.2
MB2	GTC CTC GCA ACT GC**G GCC CAG CCG GCC** CTG ATG GCG GCC GTA ACG	*Sfi* I	156	85
HenuniF1	CTG ATG GCG GCC GTG ACG	N/A	62	62.7
HenuniR1	**GAA TTC** ATT GGG CTG GCC TAG	*Eco* RI	64	60

a) Tm = 4x (G+C) + 2x (A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

The reactions consisted of PCR buffer (20 mM Tris-HCl, 50 mM KCl, 1.5 mM, $MgCl_2$, pH 8.5) (Jena Bioscience, Germany), dNTPs (200 µM each) (Jena Bioscience, Germany), 400 nM of each primer (Inqaba Biotech, South Africa), approximately 100 ng of plasmid template DNA, 2.5 U Taq DNA polymerase (2.5 U Taq, 200 µM Tris-HCl, 1 mM KCl, 1 µM EDTA, 10 µM DTT, 0.005% (v/v) Tween-20, 0.005% (v/v) Nonidet P-40, 0.5% (v/v) Glycerol, pH 8.0 (25°C)) (Jena Bioscience, Germany) in $dddH_2O$ to a final volume of 50 µl. Thermal cycling was carried out as follows: 1 min at 94°C after that 30 cycles of 94°C for 1 min 60, 65, 68 or 70°C for 1 min then 72°C for 2 min and a final extension cycle of 72°C for 5 min.

Nested Polymerase chain reaction

For the Nested PCR the same protocol was used as for PCR but, with the MB primer pair PCR fragment as template with the following primers.

Table 2.5: Nested oligonucleotide primers used in the amplification of the mycolic acid clone 10 scFv gene from the MB primer template.

Primer	Sequence	Restriction enzyme site	Tm (°C)[a]	Tm (°C)[b]
HenuniF1	CTG ATG GCG GCC GTG ACG	N/A	62	62.7
HenuniR1	**GAA TTC** ATT GGG CTG GCC TAG	*Eco R*I	64	60
MB2	GTC CTC GCA ACT GC**G GCC CAG CCG GCC** CTG ATG GCG GCC GTA ACG	*Sfi* I	156	85
pAKreverse	TG**G GCC CCC GAG GCC** GCA TTG GGC TG	*Sfi* I	92	76

a) Tm = 4x (G+C) + 2x (A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

The reactions consisted of PCR buffer (20 mM Tris-HCl, 50 mM KCl, 1.5 mM, $MgCl_2$, pH 8.5) (Jena Bioscience, Germany), dNTPs (200 µM each) (Jena Bioscience, Germany), 400 nM of each primer (Inqaba Biotech, South Africa), approximately 95 ng of PCR amplicon template, 5 U Taq DNA polymerase (5 U Taq, 400 µM Tris-HCl, 2 mM KCl, 2 µM EDTA, 20 µM DTT, 0.01% (v/v) Tween-20, 0.01% (v/v) Nonidet P-40, 1% (v/v) Glycerol, pH 8.0 (25°C)) (Jena Bioscience, Germany) in $dddH_2O$ to a final volume of 50 µl. Thermal cycling was carried out as follows: one cycle of 1 min at 94°C after that 30 cycles of 94°C for 1 min 68°C (Henuni primers) or 65°C (pAK primers) for 1 min then 72°C for 2 min and a final extension cycle of 72°C for 5 min.

Polymerase chain reaction clean up and gel extraction

For PCR clean up the specifications of the Macherey-Nagel NucleoSpin® Extract II kit were followed. Two volumes (200 µl) of buffer NT was added to one volume of sample (100 µl). This solution was then loaded onto a NucleoSpin® Extract II column which was centrifuged at 11 000 rcf for 1 min. The flow-through was discarded followed by a washing step with 700 µl buffer NT3 and centrifuged at 11 000 rcf for 1 min. The flow-through was again discarded. As a drying step the column centrifuged for 2 min at 11 000 rcf to remove any traces of ethanol. In the elution step the bound DNA was eluted from the silica membrane by adding 30 µl of buffer NE and incubated at room temperature for 4 min, followed by centrifugation at 11 000 rcf for 1 min. A second elution step was then performed.

For extraction of DNA from an agarose gel using the NucleoSpin® Extract II kit, the desired band was excised from the gel using a clean scalpel blade. For each 100 mg of agarose gel 200 µl of NT buffer was added. The gel samples were incubated at 50°C in a water bath until the agarose had completely dissolved in the buffer, usually within 20 min, with infrequent vortexing. The sample was then loaded onto a NucleoSpin® Extract II column and centrifuged at 11 000 rcf for 1 min. The flow-through was discarded. During the washing step 700 µl of buffer NT3 was added and the column centrifuged for 1 min at 11 000 rcf. The eluate was discarded, after which the column was centrifuged again at 11 000 rcf for 1 min to remove all ethanol. After a 2 min room temperature incubation, DNA bound to the silica membrane of the column was eluted with 60 µl of buffer NE and centrifugation at 11 000 rcf for 1 min.

Eco RI restriction enzyme digestion of polymerase chain reaction amplicons

For digestion of the PCR amplicon, obtained from the primers Henuni F1 and R, a reaction consisting of 500 ng of DNA and 3 Units of enzyme in 1 x FastDigest® buffer and $ddd H_2O$ to a volume of 30 µl. The reaction was incubated for 30 min at 37°C in a water bath.

Table 2.6: Restriction enzyme digest of polymerase chain reaction amplicon with *Eco* RI.

Component	Amount
PCR amplicon DNA	500 ng
10 x FastDigest® buffer	3 µl
FastDigest® enzyme	3 U
$ddd H_2O$	Reaction volume up to 30 µl

Annealing of complementary single stranded oligonucleotides

Attempts were made to anneal a forward and reverse primer pair that code for a polyhistidine tag. The resultant double stranded oligonucleotide would then have been ligated into either the recombinant or non recombinant plasmid downstream of the scFv gene site. Or the hybridized oligonucleotide is ligated to the scFv gene sequence amplicon before ligation to the plasmid pGE20.

Table 2.7: Oligonucleotide sequences used in annealing reactions used to obtain double stranded polyhistidine tag oligonucleotide.

Primer	Sequence	Restriction site	Tm (°C)[a]	Tm (°C)[b]
His1	**AAT TC**C ATC ATC ATC ATC ATC ATT AA**G**	*Eco* RI/ *Eco* RI	70	57
His-Xba	**CTA GA**C TTA ATG ATG ATG ATG ATG ATG **G**	*Xba* I/ *Eco* RI	76	61

a) Tm = 4x (G+C) + 2x (A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

Annealing/ hybridization was attempted with two different methods: Equimolar (1 µM or 10 µM) amounts of each oligonucleotide were added to 1 x PCR buffer (10 mM Tris-HCl (pH 8.3), 50 mM KCl and 1.5 mM $MgCl_2$) (TaKaRa, Japan) in $dddH_2O$. The oligonucleotides were incubated at 94°C for 2 min followed by 57 or 37°C for 90 min.

Alternatively 10 µM of each oligonucleotide were combined in 1 x annealing buffer (10 mM Tris-HCl, pH 8, 50 mM NaCl and 1 mM EDTA) and incubated at 94°C for 2 min. In a step wise manner the annealing reaction temperature was taken from 93°C down to 25°C, with a 1 min incubation period at each temperature.

Polymerase chain reaction amplicon and polyhistidine-tag coding oligonucleotide ligation with T4 DNA ligase

In order to obtain an insert for ligation into pGE20, which had previously been cut with *Xho* I and then blunt ended with Klenow fragment followed by restriction digestion with *Xba* I, the following was done. The purified *Eco* RI cut Henuni primer PCR amplicon (approximately 100 ng) and the entire annealing reaction (10 µl, 10 µM) were combined with 1x Ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM dithiothrytol and 1 mM adenosine triphosphate) and 3 U T4 DNA ligase (3 U enzyme, 1 mM Tris-HCl (pH 7.4), 5 mM KCl, 100 µM DTT, 10 µM EDTA and 5% glycerol) and $dddH_2O$ to a final volume of 13 or 30 µl. The ligation reaction was incubated overnight at 4°C.

Vector- Insert ligation with T4 DNA ligase

Plasmid (100 ng) and insert (PCR fragment or PCR fragment-His tag ligation reaction) were ligated with a 1:3 or 1:5 molar ratio with the insert being in excess. These were added together with 1x Ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM dithiotreitol [DTT] and 1 mM adenosine triphosphate) and 3 U T4 DNA ligase (3 U enzyme, 1 mM Tris-HCl (pH 7.4), 5 mM KCl, 100 µM DTT, 10 µM EDTA and 5% glycerol) and $dddH_2O$ to a final volume of 30 µl. The ligation reaction was then incubated at 4°C overnight.

Blue- white screening for clones with recombinant pGE20

The pGEJS plasmid obtained by ligation of the Henuni primer amplicon into pGE20 was used to transform *Escherichia coli* cells. Recombinant pGE20 clones were distinguished from non-recombinants by growing transformed *Escherichia coli* XL-1 blue on LB-Amp/IPTG/X-gal plates (1% tryptone, 0.5% yeast extract, 1% NaCl, 1.4% (w/v) agar, 200 µM isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) 40 µg/ml). Blue colonies indicated non-recombinant, white colonies indicate recombinant pGE20.

Nucleotide sequencing

Plasmid and PCR amplicon DNA needed to be sequenced to determine the order of nucleotides in and around the MAC10 gene. Sequencing was conducted using the BigDye 3.1 chemistry. Template (approximately 100 ng), primers (3.2 pmol), 5 x dilution buffer and BigDye 3.1 Terminator Ready Reaction Mix were combined in $ddd H_2O$ to a final volume of 10 or 20 µl. The sequencing reaction was subjected to an initial denaturing cycle of 96°C followed by 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s and elongation at 60°C for 4 min. The product of the sequencing reaction was precipitated by addition of cold 100% ethanol (64 µl) to the sequencing reaction and 270 mM sodium acetate (pH 4.8) made up to a final volume of 22 µl with $ddd H_2O$. This mixture was centrifuged at 11 000 rcf at 4°C for 30 min. The supernatant was discarded and the pellet washed twice with 60 µl of 70% ethanol and centrifugation at 11 000 rcf at 4°C for 15 min. The following primers were used for sequencing depending on the template.

Table 2.8: Oligonucleotide primers used for BigDye 3.1 sequencing of the mycolic acid clone 10 scFv gene cloned into plasmids or as polymerase chain reaction amplicons.

Primer	Sequence	Restriction enzyme site	Tm (°C)[a]	Tm (°C)[b]
HenuniF1	CTG ATG GCG GCC GTG ACG	N/A	62	62.7
HenuniR1	**GAA TTC** ATT GGG CTG GCC TAG	*Eco* RI	64	60
MB1	TGA TGG TG**G CGG CCG C**AT TGG GCT G	*Not* I	84	71.2
MB2	GTC CTC GCA ACT GC**G GCC CAG CCG GCC** CTG ATG GCG GCC GTA ACG	*Sfi* I	156	85
pAKreverse	TG**G GCC CCC GAG GCC** GCA TTG GGC TG	*Sfi* I	92	76
M13R	CAG GAA ACA GCT ATG AC	N/A	50	50.4

a) Tm = 4x (G+C) + 2x (A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

Samples were submitted for sequencing at the Sequencing facility at the University of Pretoria on an ABI 3130 sequencer.

Sequence analysis

Results were obtained as ABI data and chromatograms. Sequences were corrected using ClustalW multiple alignments in the BioEdit sequence alignment editor program. Knowledge of the 5' and 3' end sequences of the MAC10 gene as well as restriction sites were used as reference points.

Restriction enzyme digestion of plasmid pAK400 with *Sfi* I

To digest plasmid pAK400 (2 µg) the DNA was combined with 1 x buffer G (1 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 5 mM NaCl and 0.01 mg/ml BSA) (Fermentas), 10 U of enzyme (0.5 mM Tris-HCl (pH 7.4 at 25°C), 15 mM NaCl, 500 µM $MgCl_2$, 50 µM DTT, 0.0075% Triton X-100, 10 µg/ml BSA and 2.5% glycerol) (Fermentas) and $dddH_2O$ up to a volume of 20 µl. The reaction was then incubated at 50°C in a water bath for 3h.

Table 2.9: Restriction digest of plasmid pAK400 with *Sfi* I.

Component	Amount
Template DNA	2 µg
Buffer G	3 µl
restriction enzyme	10 U
$dddH_2O$ *Sfi* I	Reaction volume up to 20 µl

Ligation of amplicon from pAK primers into pAK400 with T4 DNA ligase

Two types of ligation reactions were set up, the first according to literature (Krebber *et al*., 1997) and the second according to the T4 DNA ligase manufacturers specifications. In the first method, plasmid pAK400 (100 ng) and insert (13 ng) were combined in 1x ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP) (Promega, USA) with 3 U of enzyme (0.5 mM Tris-HCl (pH 7.4), 2.5 mM KCl, 50 µM DTT, 5 µM EDTA and 2.5% glycerol) (Promega, USA) and $dddH_2O$ up to a volume of 20 µl at a ratio of 1:1.5.

Table 2.10: Ligation of pAK primer set amplicon into pAK400 at a molar ratio of 1: 1.5 ratio

Component	Amount
Vector	100 ng
Insert	13 ng
Ligase buffer	2 µl
T4 DNA ligase	3 U
$dddH_2O$	Reaction volume up to 20 µl

In the second method pAK400 (100 ng) and insert (58 ng) were combined in 1x ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP) (Promega, USA) with 3 U of enzyme (0.5 mM Tris-HCl (pH 7.4), 2.5 mM KCl, 50 µM DTT, 5 µM EDTA and 2.5% glycerol) (Promega, USA) and $dddH_2O$ up to a volume of 20 µl at a ratio of 1:3.

Table 2.11: Ligation of pAK primer set amplicon into pAK400 at a molar ratio of 3: 1 ratio

Component	Amount
Vector	100 ng
Insert	58 ng
Ligase buffer	2 µl
T4 DNA ligase	3 U
$dddH_2O$	Reaction volume up to 30 µl

Results

Plasmid isolation

The plasmid vector pGE20 was constructed as a derivative of pBluescript for the expression of scFv's (Orfanoudakis *et al*., 1993). Purified pGE20 (~3 kb) was obtained from overnight cultures (2 x TY or LB) using Zyppy™ and Macherey-Nagel NucleoSpin® Plasmid kits (figure 2.8). The plasmid was eluted in 30 and 50 µl of elution buffer respectively. The maximum yields obtained, as measured by Nanodrop spectrometry were 100 ng/µl for Zyppy™ kits and 300 ng/µl for the Macherey-Nagel NucleoSpin® Plasmid kit respectively. Maximum yields expected were 10 µg for the Zyppy™ kit and 25 µg for the Macherey-Nagel NucleoSpin® Plasmid kit. The presence of DNA in each sample was identified by agarose gel electrophoresis (110 V, 1 h) of a 5 µl aliquot depending on DNA concentration. As seen in figure 2.8, in both cases a prominent band at the middle of the gel was obtained.

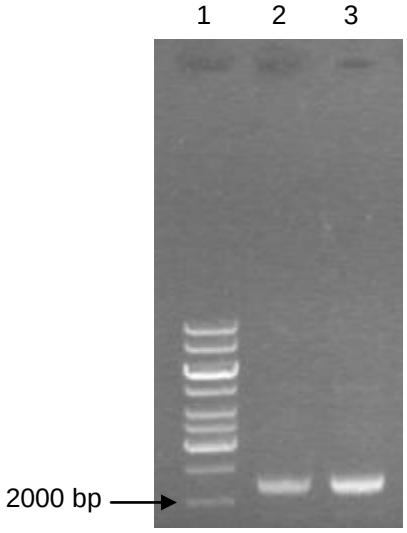


Figure 2.8: Agarose gel electrophoresis (1% agarose) of non recombinant pGE20 isolated from an overnight culture of *E*scherichia *coli* XL-1 blue in 2xTY or LB media with ampicillin. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 and 3 = plasmid pGE20.

Sequencing with the M13R primer allowed for the identification of the region between the first of two pelB leader sequence and the Not I restriction site as can be seen in figure 2.9.

ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGATATTACTCGCTGCCCAACCAGCCATGGCC
CAGGTGAAACTGCTCGAGATTTCTGACTAGTCGTCCGAACTCCGATAATCGCCGTCAGGGCGGTCG
CGAACGTTTATAATAGAATTCTAAACTAGCTAGTCGCCAAGGAGACAGTCATA**ATGAAA**TACCTATTG
CCTACGGCAGCCGCTGGATTGTTATTACTCGCTGCCCAACCAGCCATGGCCGAGCTCGTCAGTTCT
AGAGTTAAGCGGCCGC

Figure 2.9: Sequence of cloning region of plasmid pGE20. Indicated in bold are the first two codons of the *PelB* leader gene. Underlined are the *Xho* I, *Spe* I, *Eco* RI, *Sac* I, *Xba* I and *Not* I restriction sites.

The pAK series of plasmid vectors were constructed for either phage display or scFv expression. To obtain an efficient expression system (pAK400), the gene III was removed, the *c-myc* tag was replaced by a His-tag and the Shine-Dalgarno sequence of pAK100 was replaced by SDT7g10 (Krebber *et al*., 1997). The plasmids pAK100 and 400 were isolated from an overnight culture of NE medium. High yields were obtained for both plasmids as was determined by Nanodrop spectrometry, for pAK100 498.6 ng/µl (23.9 µg) and for pAK400 247.2 ng/µl (10.6 µg) was obtained. Figure 2.10 shows that for both plasmids a single band was obtained with pAK100 (6426 bp) being larger than pAK400 (5937 bp).

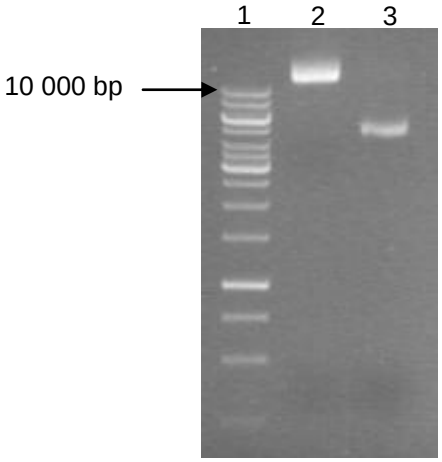


Figure 2.10: Agarose gel electrophoresis (1% agarose) of plasmids pAK100 and pAK400, isolated from an overnight culture of *Escherichia coli* HB2151 in NE medium. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = plasmid pAK100; 3 = plasmid pAK400.

Preparation of plasmid pGE20 for ligation

In order for the MAC10 scFv amplicon to be ligated into pGE20, the plasmid has to be restriction enzyme digested. However because there is an Xho I restriction site inside the MAC10 gene, after digestion the Xho I site has to be filled in for blunt end ligation. For ligation of the MAC 10 scFv gene into plasmid pGE20, the plasmid was cut first with *Xho* I. Two samples of plasmid (1.5 and 3 µg each) were digested and analyzed by agarose gel electrophoresis. A single band was observed between the 3500 and 3000 bp fragments of the O'GeneRuler™ 1 kb DNA ladder for both samples. The digested band was slightly closer to the origin than the prominent undigested band as can be seen in figure 2.11. After purification the maximum amount of DNA recovered (1% agarose gel) was 1.3 µg of digested plasmid from an initial 3 µg of undigested plasmid. Concentrations for the plasmids were determined by Nanodrop spectrometry. The purified digested plasmid was then used for Klenow fragment blunting.

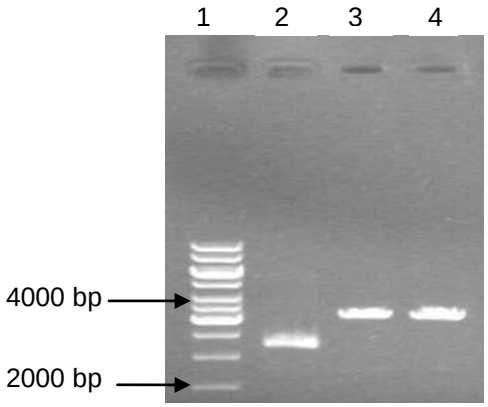


Figure 2.11: Agarose gel electrophoresis (1%) of Xho I restriction enzyme digested non recombinant pGE20 plasmid. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = undigested plasmid pGE20, 3 and 4 = *Xho* I digested plasmid pGE20.

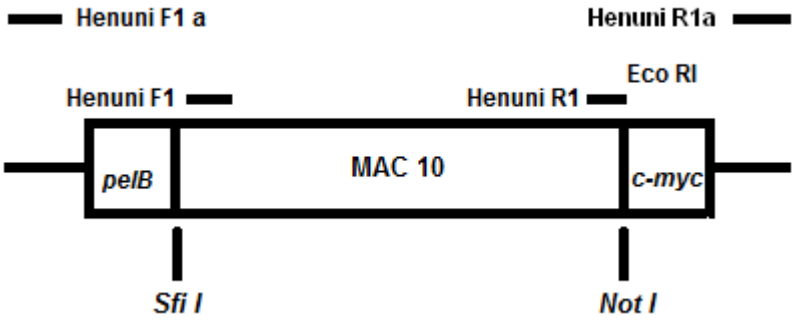
Polymerase chain reaction with Henuni primer set and pMAC10 as template

The Henuni primers were designed to amplify the MAC10 scFv gene from pMAC10 and allow for ligation into pGE20. The forward primer (F1) was blunt ended at the 5'-end while the reverse primer (R1) added an *Eco* RI to the 3'-end of the MAC10 gene. Polymerase chain reaction amplification of the MAC10 gene was performed using the Henuni primers and the recombinant pHEN-1 (pMAC10) plasmid template. The cycling parameters used were previously defined (van Wyngaardt *et al*., 2004). The amplicon was expected to be 732 bp in length starting at the 3'- CTG ATG – 5', which is just downstream of the *Sfi* I restriction site in pMAC10, and ending in the added Eco RI site. Sequencing of the amplicons obtained from the Henuni primers, at different annealing temperatures, indicated that the forward primer annealed incorrectly as indicated by figures 2.12 A and B. At an annealing temperature of 70°C the shortest upstream sequence was obtained with 43 bases from pHEN-1 and the *pelB* leader sequence were amplified along with the gene. The amplicon obtained was run on a 1% agarose gel as can be in figure 2.12 C; the bands obtained were at the expected position yet thick and slightly diffuse.

A)

ATTCCCAGCCCCTTTACCAGGCGCCTGTAGCACGCAAACCATG**ATGAAATACCTATTGCCTACGGC**
AGCCGCTGGATTACTCGCGGCCCAGCCGGCCCTGATGGCGGCCGTGACGTTGGACGAGTCCGGG
GGCGGCCTCCAGACGCCCGGAGGAGGGCTCAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAG
CAGTTACAGCATGGGTTGGGTGCGACAGGCGCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTG
ATGATGATGGTAGTTTCACACACTACGGGTCGGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACA
ACGGGCAGAGCACAGTGAGGCTGCAGCTGAACAACCTCAGGGCTGAGGACACCGGCACCTACTAC
TGCGCCAAAAAGATTATTATATTATACGATTATCGACGCATGGGGCCAGGGACCGAAGTCATCGTCT
CCTCCGGTGGAGGCGGTTCAGGTGGAGGTGGCTCTGGCGGAGGCGGATCGGCGCTGACTCAGCC
GTCCTCGGTGTCAGCGAACCCGGGAGAAATCGTCGAGATCACCTGCTCCGGGAGTGACAGCAGCA
ACAACTATGGCTGGTATCAGCAGAAGGCACCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACAC
CAACAGACCCTCGAACATCCCTTCACGATTCTCCGGTTCCACATCCGGCTCCACAGCCACATTAACC
ATCACTGGGGTCCAAGCCGAGGACGAGGCTGCCTATTTCTGTGCGAGTACAGACAGCATATTTGGG
GCCGGGACAACCCTGACCGTCCTAGGCCAGCCCAATGAATTC(A)$_m$

B)



C)

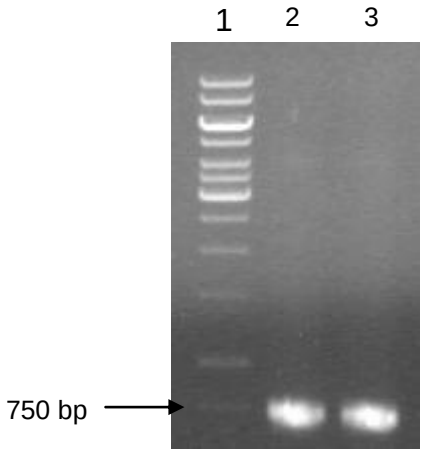
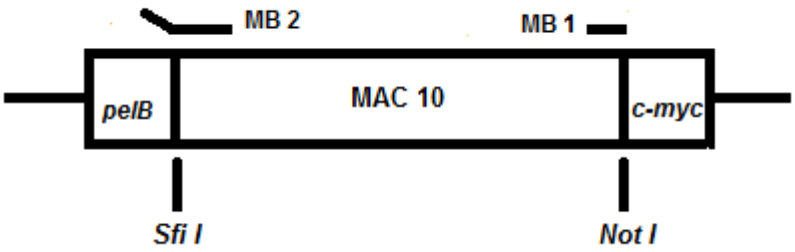


Figure 2.12: PCR amplification of the MAC10 scFv gene using Henuni primers and pMAC10 template. A) Sequence obtained for PCR amplicon from Henuni primers and pMAC10 template with annealing temperature 70°C. *PelB* leader sequence in bold. *Sfi* I and *Eco* RI restriction sites as well as internal *Xho* I site underlined. Sequence upstream of *PelB* leader not corrected. m = uncertain number of nucleotides. B) Diagram of Henuni primer annealing positions on pMAC10, *pelB*, pectate lyase signal peptide sequence; *Sfi* I and *Not* I restriction site; MAC10, mycolic acid clone 10 scFv gene; *c-myc*, *c-myc* signal peptide; Henuni F1/R1, designed positions for primer annealing; Henuni F/R1a,non specific annealing position of Henuni F/R1; *Eco* RI, *Eco* RI restriction site of Henuni R1. C) Agarose gel electrophoresis (1%) of PCR amplicon obtained from the Henuni primers and pMAC10 template. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 and 3 = amplicon obtained by PCR of MAC10 gene with Henuni primers.

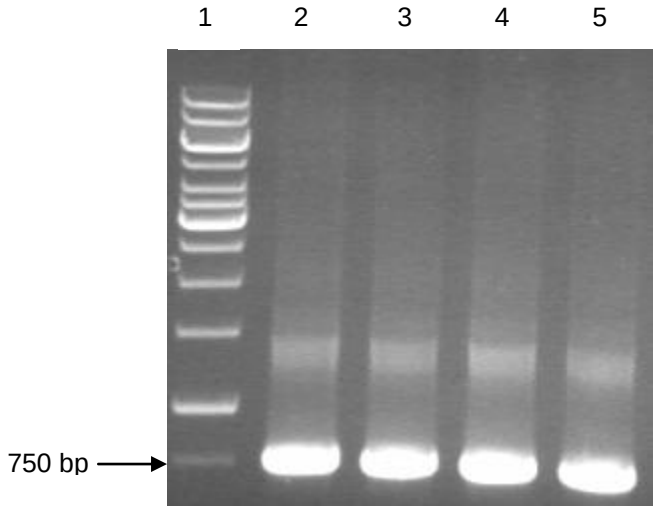
Polymerase chain reaction with MB primers and pMAC10 template

The MAC10 scFv gene was amplified from the recombinant pHEN-1 (pMAC10) plasmid using the MB1/2 primers and a previously described protocol (van Wyngaardt *et al*., 2004). A 769 bp fragment was expected to be obtained from the reaction including the scFv gene and the short sequences upstream of the *Sfi* I site and downstream of the *Not* I site. Figure 2.13 A) illustrates the annealing positions the primers were designed for by van Wyngaardt et al. Agarose gel electrophoresis of a 10 µl of each of the repeats done indicated that a very prominent band of ~ 750 bp had been produced (figure 2.13 B). Which were confirmed by Nanodrop spectrometry to have concentrations of 189 ng/µl. Some faint smearing due to over amplification is visible above this band particularly concentrated at a region of ~ 1.5 kbp. Sequencing results as shown in figure 2.13 C indicated that the exact expected region had been amplified and that no non specific annealing of the forward primer had occurred. This was a concern since both MB2 and Henuni F1 share the GGCGGCCG sequence.

A)



B)



C)

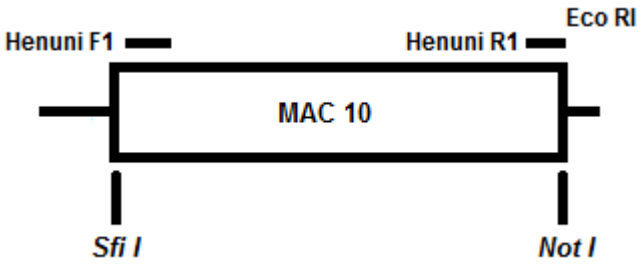
TCGCAACTGCGGCCCAGCCGGCCCTGATGGCGGCCGTGACGTTGGACGAGTCCGGGGGCGGCCT
CCAGACGCCCGGAGGAGGGCTCAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAGCAGTTACA
GCATGGGTTGGGTGCGACAGGCGCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTGATGATGAT
GGTAGTTTCACACACTACGGGTCGGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACAACGGGCA
GAGCACAGTGAGGCTGCAGCTGAACAACCTCAGGGCTGAGGACACCGGCACCTACTACTGCGCCA
AAAAGATTATTATATTATACGATTATCGACGCATGGGGCCAGGGACCGAAGTCATCGTCTCCTCCGG
TGGAGGCGGTTCAGGTGGAGGTGGCTCTGGCGGAGGCGGATCGGCGCTGACTCAGCCGTCCTCG
GTGTCAGCGAACCCGGGAGAAATCGTCGAGATCACCTGCTCCGGGAGTGACAGCAGCAACAACTAT
GGCTGGTATCAGCAGAAGGCACCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACACCAACAGA
CCCTCGAACATCCCTTCACGATTCTCCGGTTCCACATCCGGCTCCACAGCCACATTAACCATCACTG
GGGTCCAAGCCGAGGACGAGGCTGCCTATTTCTGTGCGAGTACAGACAGCATATTTGGGGCCGGG
ACAACCCTGACCGTCCTAGGCCAGCCCAATGCGGCCGCCACCATCA

Figure 2.13: Polymerase chain reaction with MB primers and pMAC10 template. A) Diagram of MB primers annealing positions on pMAC10, *pelB*, pectate lyase signal peptide sequence; *Sfi* I and *Not* I restriction site; MAC10, mycolic acid clone 10 scFv gene; *c-myc*, *c-myc* signal peptide; MB 1/2, positions for primer annealing (van Wyngaardt *et al*., 2004). B) Agarose gel electrophoresis (1%) of 769 bp PCR amplicon obtained from the MB primers and pMAC10 template. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 to 5 = amplicon produced by PCR of MAC10 gene with MB primers. C) Sequence of 769 bp amplicon obtained from MB primers and pMAC10 template. *Sfi* I and Not I restriction sites are underlined.

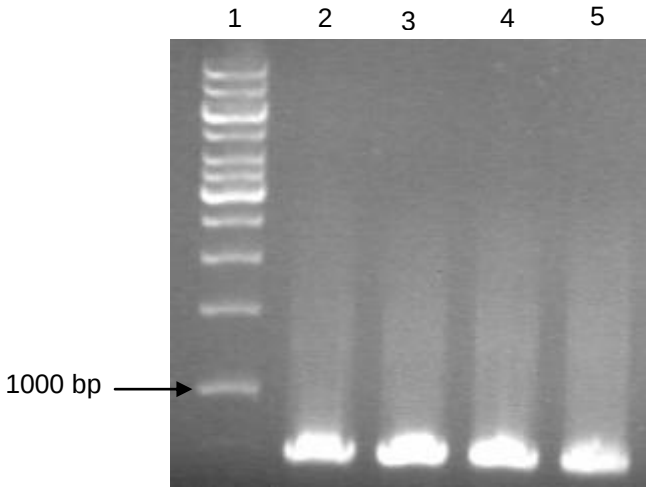
Nested Polymerase chain reaction with Henuni primer set

Using the Henuni primers directly to amplify the MAC10 scFv gene from pMAC10 resulted in non specific annealing of the primers. In order to eliminate the sequences responsible for the non specific annealing nested PCR was used. Nested PCR was performed using the MB1/2 primer PCR amplicon as template and the Henuni primer set. With the annealing positions of the Henuni primers on the amplicon as illustrated in figure 2.14 A. The cycling parameters were as before except that the annealing temperature had been increased to 68°C. The amplification product was expected to be 732 bp in length consisting of only the MAC10 scFv gene and the added *Eco* RI site to the 3' end of the gene. Figure 2.14 B shows that a prominent band at ~ 750 bp was obtained from a 10 µl aliquot of each of the four repeats run on agarose gel electrophoresis. Over amplification was somewhat less for the Nested PCR as compared to the PCR with the MB primers. Sequencing of the amplicon obtained at an annealing temperature of 68°C indicated that the gene was amplified exactly as expected beginning and ending where it should as can be seen in figure 2.14 C. Except for a variable number of thymidines at the 5' end and adenines at the 3' of the gene no bases originating from pHEN-1 were amplified along with the gene. It is also noted that the *Eco* RI site had been added to the 3' end of the gene.

A)



B)



C)

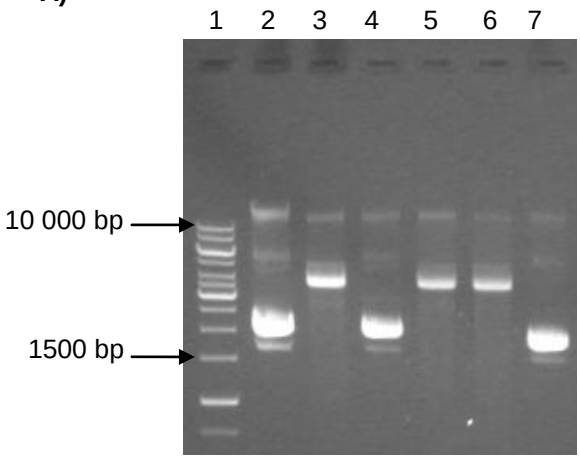
$(T)_n$CTGATGGCGGCCGTGACGTTGGACGAGTCCGGGGGCGGCCTCCAGACGCCCGGAGGAGGGCT
CAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAGCAGTTACAGCATGGGTTGGGTGCGACAGGC
GCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTGATGATGATGGTAGTTTCACACACTACGGGTC
GGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACAACGGGCAGAGCACAGTGAGGCTGCAGCTG
AACAACCTCAGGGCTGAGGACACCGGCACCTACTACTGCGCCAAAAAGATTATTATATTATACGATT
ATCGACGCATGGGGCCAGGGACCGAAGTCATCGTCTCCTCCGGTGGAGGCGGTTCAGGTGGAGGT
GGCTCTGGCGGAGGCGGATCGGCGCTGACTCAGCCGTCCTCGGTGTCAGCGAACCCGGGAGAAAT
CGTCGAGATCACCTGCTCCGGGAGTGACAGCAGCAACAACTATGGCTGGTATCAGCAGAAGGCACC
TGGCAGTGCCCCTGTCACTGTGATCTATGACAACACCAACAGACCCTCGAACATCCCTTCACGATTC
TCCGGTTCCACATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCAAGCCGAGGACGAGGCT
GCCTATTTCTGTGCGAGTACAGACAGCATATTTGGGGCCGGGACAACCCTGACCGTCCTAGGCCAG
CCCAATGAATTC$(A)_m$

Figure 2.14: Nested PCR amplification of MAC10 scFv gene with Henuni primer set. A) Diagram of annealing positions of Henuni primers on MB primers amplicon, *Sfi* I and *Not* I restriction sites; MAC10, mycolic acid clone 10 scFv gene; Henuni F1/R1, annealing positions of primers; *Eco* RI, restriction site of Henuni R1. B) Agarose gel electrophoresis (1%) of 732 bp Nested PCR amplicon obtained from the Henuni primers with MB primer amplicon as template. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder; 2 to 5 = amplicon produced by PCR of MB primer amplicon with Henuni primers. C) Sequence of the 732 bp amplicon obtained by Nested PCR from Henuni primers and MB primer amplicon template. n/m = uncertain number of nucleotides. Underlined is the *Eco* RI site added on by Henuni R1.

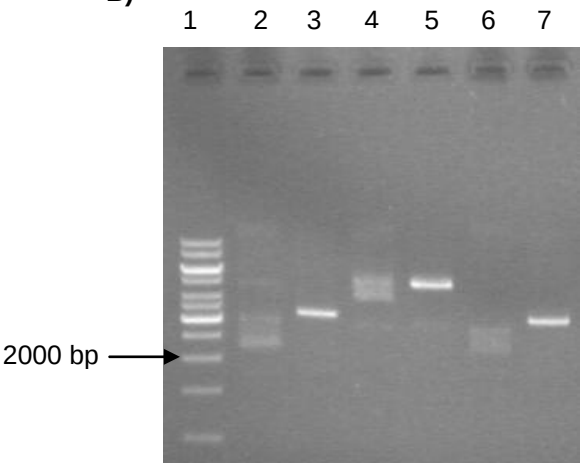
Ligation of Polymerase chain reaction amplicon to plasmid pGE20

The MAC10 scFv PCR amplicon obtained using the Henuni primers and the pMAC10 template was to be cloned into the pGE20 plasmid. Due to an *Xho* I restriction site inside the MAC10 gene, the 5'-end of the amplicon and the pGE20 plasmid were ligated by blunt end cloning. While the 3'-end of the amplicon with the *Eco* RI site from Henuni R1 was to be ligated to the *Eco* RI site of pGE20. Following the Klenow blunting, the plasmid was digested with *Eco* RI and purified by gel extraction. From this 276 ng (9.2 ng/ µl) and 330 ng (11 ng/ µl) were recovered respectively as determined by Nanodrop spectrometry. The *Eco* RI digested plasmid and PCR fragment were consequently ligated together and transformed into competent *Escherichia coli* XL1-blue. The transformation efficiency of the non recombinant pGE20 transformation was calculated to be approximately $1x10^7$ cfu/µg (colony forming units per microgram of plasmid). Blue-white screening allowed for the identification of 3 recombinant (white) and 3 light blue colonies. Plasmid isolated from all these colonies run on 1% agarose gel electrophoresis (figure 2.15 A) showed a markedly different banding pattern to non recombinant pGE20. Most noticeable is the increase in the number of bands for all the clones, having increased from 1 in pGE20 to 3 or 5 for the clones. Also the six clones can be divided into two groups based on their banding patterns. The first group (I, 1.2 and 2.4) contains four faint bands and one very prominent band. The second group (1.1, 2.1 and 2.2) has 2 faint bands and a slightly more prominent band. The change in banding was assumed to indicate ligation of the scFv gene had occurred. Digestion of the pGEJS colonies with *Eco* RI yielded either a single band or a number of faint bands close in size as shown in figure 2.15 B. The single bands of 1.1 and 2.4 are close in size while the band of 2.1 was larger. The faint bands seen for I, 1.2 and 2.2 are probably due to star activity of *Eco* RI. Sequence analysis identified 1.1 and 2.1 as recombinant. From the sequences obtained (figure 2.15 C) it was also observed that non-specific annealing of the Henuni primer set had occurred. As can be seen not only was the gene amplified but also upstream and downstream sequences of the pHEN-1 vector as exemplified by the *Sfi* I site and *c-myc* tag gene respectively. Also the *Xho* I and *Eco* RI restriction sites of the pGE20 plasmid vector could not be found flanking the sequence also indicating that a larger fragment than the 732 bp scFv gene had been amplified.

A)



B)



C)

ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTACTCGCGGCCCAGCCGGCCCTGATGGCGGC
CGTGACGTTGGACGAGTCCGGGGGCGGCCTCCAGACGCCCGGAGGAGGGCTCAGCCTCGTCTGC
AAGGCCTCCGGGTTCACCTTAAGCAGTTACAGCATGGGTTGGGTGCGACAGGCGCCTGGCAAAGG
GCTGGAATGGGTCGCTGGTATTGATGATGATGGTAGTTTCACACACTACGGGTCGGCGGTGAAGGG
CCGTGCCACCATCTCGAGGGACAACGGGCAGAGCACAGTGAGGCTGCAGCTGAACAACCTCAGGG
CTGAGGACACCGGCACCTACTACTGCGCCAAAAATTATTATATTATACGATTATCGACGCATGGGGC
CAGGGACCGAAGTCATCGTCTCCTCCGGTGGAGGCGGTTCAGGTGGAGGTGGCTCTGGCGGAGGC
GGATCGGCGCTGACTCAGCCGTCCTCGGTGTCAGCGAACCCGGGAGAAATCGTCGAGATCACCTG
CTCCGGGAGTGACAGCAGCAACAACTATGGCTGGTATCAGCAGAAGGCACCTGGCAGTGCCCCTGT
CACTGTGATCTATGACAACACCAACAGACCCTCGAACATCCCTTCACGATTCTCCGGTTCCACATCC
GGCTCCACAGCCACATTAACCATCACTGGGGTCCAAGCCGAGGACGAGGCTGCCTATTTCTGTGCG
AGTACAGACAGCATATTTGGGGCCGGGACAACCCTGACCGTCCTAGGCCAGCCCAATGCGGCCGC
AGAACAAAAACTCATCTCAGAAGAGGATCTGAAT

D)

GTGCCTTTGCATTCAAGTCTATTTCAGGAGACAGTCATA

CTCCTTGCATGCAAATTCTATTTCAGGAGACAGTCATA

E)

ACTGTTGAAAGTTGTTTAGCAAAACCTCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAA
AATTTTAGATCGTTACGCTAACTATGAGGCTGTCTGTGGAATGCTACAGGCGTTGTGTTTTGTACTGT
GACGAAAACTCAATGTTTACGGTACATGGGTTCCTTATGGGCCTTGCTAATCCTGAAAATGAGGGTG
GTGGCTTCTGAGGTGGCTGTTTCTGAAGGGTGTGCCGTTCTGAAGGTGGCGATACATAACCCTCTT
GGAATACCGGTGAATACACCTAATTTCCGGGACTTATTTA

ACTGTTGAAAGTTGTTTAGCAAAACCTCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAA
AACTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGGTGGAATGCTACAGGCGTTGGTGGTTTGTA
CTGGTGACGAAAACTCAGTGTTACGGTACATGGGTTTCCTTATTGGGGCTTGCTATCCCCTGAAAAT
GGAGGGTGGGTGGCTCTGAGGGTGGCCGGTTTGAGGGGTGGGCGGTTCCTGGAGGGTGGGCGGT
ACTTAACCTCCTGGAGTACCGGTGATACACCTATTTCCGGGGCCATAATAC

Figure 2.15: Ligation of PCR amplicon obtained by non specific annealing of Henuni primers and pMAC10 template into *Xho* I and *Eco* RI restriction enzyme digested pGE20. A) Agarose gel electrophoresis (1%) of plasmid DNA obtained from recombinant pGEJS colonies. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pGEJSI, 3 = pGEJS1.1, 4 = pGEJS1.2, 5 = pGEJS2.1, 6 = pGEJS2.2, 7 = pGEJS2.4. B) Agarose gel electrophoresis (1%) for *Eco* RI restriction enzyme digests of recombinant clones. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = uncut I, 3 = cut I, 4 = uncut 1.1, 5 = cut 1.1, 6 = uncut 2.1, 7 = cut 2.1. C) Sequence of pGEJS colony obtained from ligation of Henuni amplicon into pGE20. D) Sequences upstream of pelB leader gene from pGEJS1.1 and 2.1 respectively. E) Sequences downstream of c-myc tag gene from pGEJS1.1 and 2.1 respectively.

Ligation of His–tag oligonucleotide to pGEJS

Once the MAC10 scFv gene had been cloned into the pGE20 plasmid, the oligonucleotide coding for the Histidine expression tag had to ligated into the newly recombinant pGEJS plasmid. In order to ligate the oligonucleotide into pGEJS the plasmid was digested with *Eco* RI (figure 2.16). The oligonucleotide was to be obtained by annealing the single stranded oligonucleotides His1 and 2. The His-tag oligonucleotide annealing reaction was used in a ligation reaction with *Eco* RI digested recombinant pGE20. Restriction enzyme digestion of the recombinant clones 1.1 and 2.1 yielded bands of between 5000 and 6000 bp which is much larger than expected. The parental pGE20 is between 3000 and 3500 bp and digestion with *Xho* I and *Eco* RI would have result in the loss of 183 bp. The scFv gene would then have added 730 bp and therefore the expected size would be ~4000 bp.

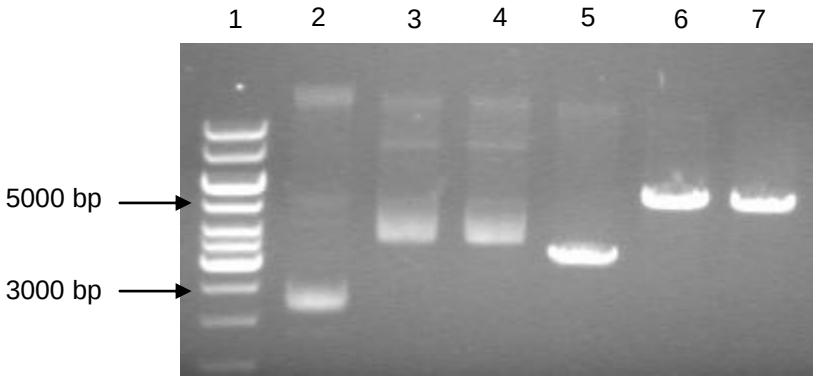


Figure 2.16: Agarose gel electrophoresis (1%) of *Eco* RI restriction enzyme digested recombinant pGE20 (pGEJS). 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = uncut pGE20, 3 = uncut pGEJS1.1, 4 = uncut pGEJS2.1, 5 = cut pGE20, 6 = cut pGEJS1.1, 7 = cut pGEJS2.1. Cut = *Eco* RI digested plasmid.

Selected clones of the ligation of the His-tag to pGE20 and 1.1 and 2.1 yielded the plasmids seen in figure 2.17. The position of the pGE20 clone is unchanged with respect to the parental plasmid as expected as the change in the number of base pairs is too small to see on a gel. Also as expected is that the clones picked for the recombinant pGE20 1.1 and 2.1, were unchanged in their size for the same reason as before. All the recombinant plasmids were shown to be of approximately the same size.

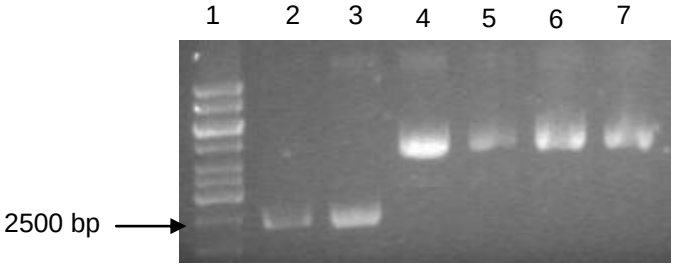


Figure 2.17: Agarose gel electrophoresis (1%) of plasmid obtained from ligation His-tag oligonucleotide to non recombinant pGE20 and recombinant pGEJS. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pGE20, 3 = pGE20-His, 4 = pGEJS1.1, 5 = pGEJS1.1-His, 6 = pGEJS2.1, 7 = pGEJS2.1-His.

To confirm the presence of the MAC10 and His-tag genes in the recombinant plasmid, a set of PCRs was performed. Figure 18 is the 1% agarose gel electrophoresis of the amplicons obtained. The first two lanes are duplicate PCRs for the MAC10 gene alone; the last two lanes are for MAC10 and the His-tag. For all four samples four non-specific bands were obtained at the same position. Only one of the PCRs to confirm the MAC10 genes presence showed any amplification at the expected region of approximately 750 bp.

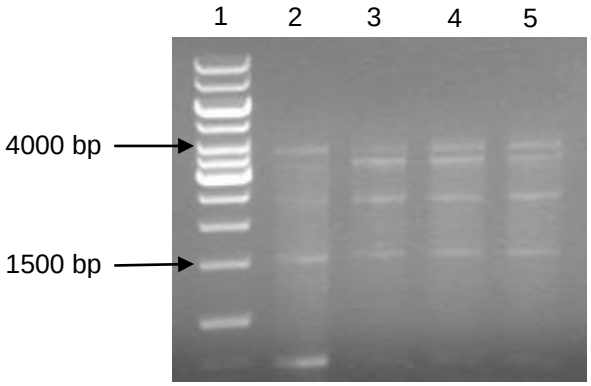


Figure 2.18: Agarose gel electrophoresis (1%) of PCR amplicons obtained from Henuni F1 and His-Xba primers with recombinant pGEJS as template to determine ligation of His-tag gene. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 and 3 = amplicon(s) obtained by PCR of ligation product with Henuni primers, 4 and 5 = amplicon(s) obtained by PCR of ligation product using Henuni F1 and His-Xba primers.

To further confirm the ligation of both genes and the retention of the restriction sites, pGE20, 1.1 and 2.1 were digested with *Eco* RI and *Xba* I. Digestion of pGE20 with either enzyme yielded a band at approximately 3000 bp as expected. Digestion of 1.1 and 2.1 by *Eco* RI again yielded bands of approximately 5000 bp. However digestion with *Xba* I did not occur for either clone (figure 2.19).

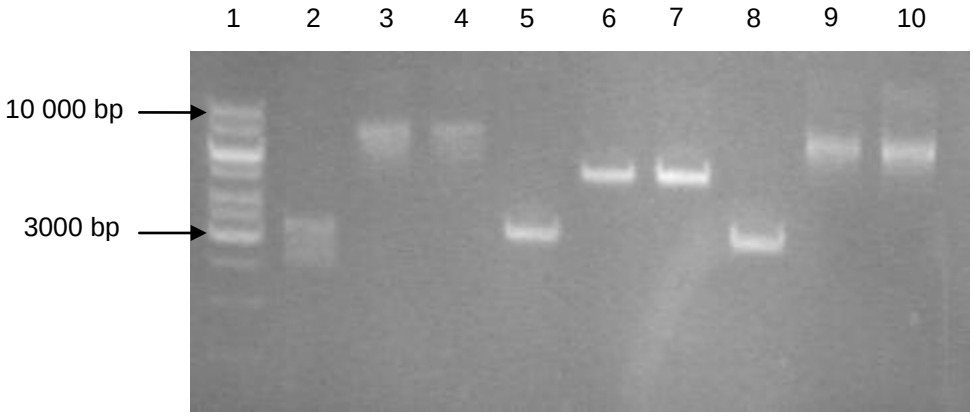


Figure 2.19: Agarose gel electrophoresis (1%) of uncut, *Eco* RI and *Xba* I restriction enzyme digested pGE20 and recombinant pGEJS clones. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pGE20, 3 = pGEJS1.1, 4 = pGEJS2.1, 5 = *Eco* RI cut pGE20, 6 = *Eco* RI cut pGEJS1.1, 7 = *Eco* RI cut pGEJS2.1, 8 = *Xba* I cut pGE20, 9 = *Xba* I cut pGEJS1.1, 10 = pGEJS2.1.

Sequences were identical to those of figure 15 C above, confirming that no His-tag gene had been ligated into the recombinant plasmid pGEJS clones.

Ligation of His-tag oligonucleotide to non-recombinant pGE20

The ability of the annealed His-tag oligonucleotides to ligate to *Eco* RI and *Xba* I restriction enzyme digested pGE20 was analyzed. In figure 2.20 below *Eco* RI, *Xba* I and double digested pGE20 were compared with undigested pGE20 plasmid. For both *Eco* RI and the double digested plasmid, there does not seem to be any visible plasmid degradation due to *Eco* RI star activity. In figure 2.20 digestion with the single enzymes is confirmed by the single bands between 3000 and 3500 bp. This compares well with the 2961 bp of pBluescript II from which pGE20 was derived plus the 46 bp of the pelB sequence. As expected the *Eco* RI/*Xba* I double digested bands were not noticeably smaller in size than the single cut bands as only 110 bp were cut out.

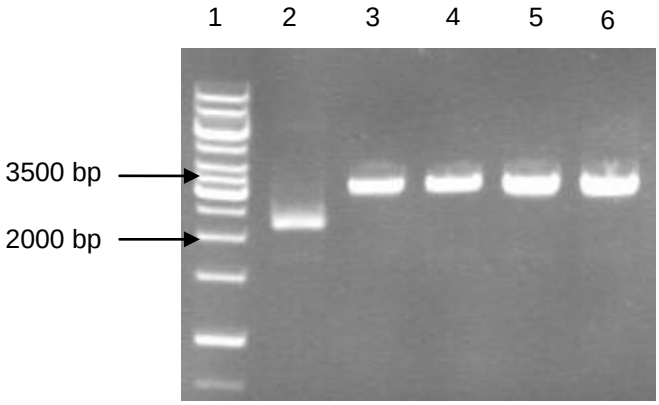


Figure 2.20: Agarose gel electrophoresis (1%) of *Eco* RI and *Xba* I restriction enzyme double digested pGE20. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = uncut pGE20, 3 = *Eco* RI cut pGE20, 4 = *Xba* I cut pGE20, 5 and 6 = *Eco* RI and *Xba* I cut pGE20.

Blue white screening of ligated double digested pGE20 and the His-tag oligonucleotide annealing reactions yielded only blue colonies. Minipreps of selected colonies resulted in plasmid bands with sizes closer to that of parental uncut pGE20 than expected (figure 2.21).

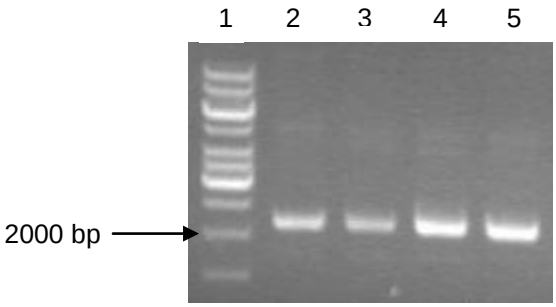


Figure 2.21: Agarose gel electrophoresis (1%) of plasmid isolated after ligation of His-tag annealing reaction to pGE20. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 and 3 = pGE20, 4 and 5 = pGE20-His.

Sequencing of the selected clones showed that they were identical to parental pGE20 (Figure 2.22) as can be identified from the two pelB sequences, *Xho* I, *Eco* R I, *Xba* I and *Not* I sites.

ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGATATTACTCGCTGCCCAACCAGCCATGGCC
CAGGTGAAACTGCTCGAGATTTCTGACTAGTCGTCCGAACTCCGATAATCGCCGTCAGGGCGGTCG
CGAACGTTTATAATAGAATTCTAAACTAGCTAGTCGCCAAGGAGACAGTCATA**ATGAAA**TACCTATTG
CCTACGGCAGCCGCTGGATTGTTATTACTCGCTGCCCAACCAGCCATGGCCGAGCTCGTCAGTTCT
AGAGTTAAGCGGCCGC

Figure 2.22: Sequence of cloning region of plasmid pGE20. Indicated in bold are the first two codons of the *PelB* leader gene. Underlined are the *Xho* I, *Eco* RI, *Xba* I and *Not* I restriction sites**.**

Ligation of His-tag oligonucleotide to Polymerase chain reaction amplicon

Before being ligated to the pGE20 plasmid the Nested PCR amplicon of the MAC10 scFv gene and the His-tag oligonucleotide annealing reaction were ligated together. The resultant amplicon-His-tag can ligate to the plasmid because of the blunt end at the 5'-end and the *Xba* I site at the 3'-end. The separate amplicon and oligonucleotide can only ligate at one point but the plasmid should not re-circularize. Minipreps of colonies picked by blue-white screening yielded plasmid DNA with bands different from pGE20. It can be seen in figure 2.23 below that these plasmids are retained more by the agarose network than pGE20.

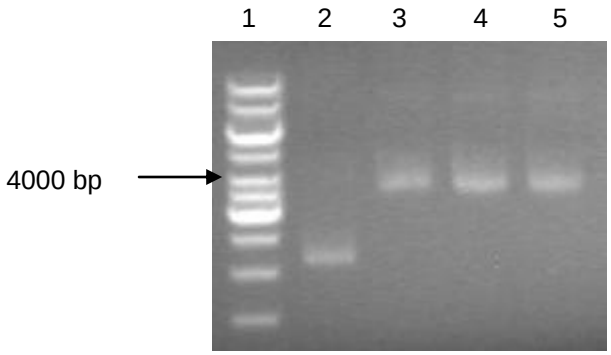


Figure 2.23: Agarose gel electrophoresis (1%) of plasmid isolated after ligation of MAC10-His-tag complex to *Xho* I and *Xba* I restriction enzyme digested pGE20. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pGE20, 3 to 5 pGEJS-His.

In figure 2.24 the confirmation PCR indicated that the MAC10 scFv gene was indeed present in the plasmid, but that the His-tag had not ligated to the PCR fragment. The first three lanes represent triplicates PCRs to confirm the presence of the scFv gene in the plasmid using Henuni F1/R1 as primers. As expected a band is observed at ~ 750 bp as the amplified gene region is 730 bp long. The next three lanes represent the triplicates of the PCR using Henuni F1/His-Xba to confirm that the annealed His-tag oligonucleotide and scFv gene had indeed ligated prior to ligation to the pGE20 plasmid. A very faint band was obtained at ~750 bp, and a much brighter band below it of ~ 500 bp. Thus ligation did not occur between the amplicon and the His-tag gene but between the amplicon and pGE20.

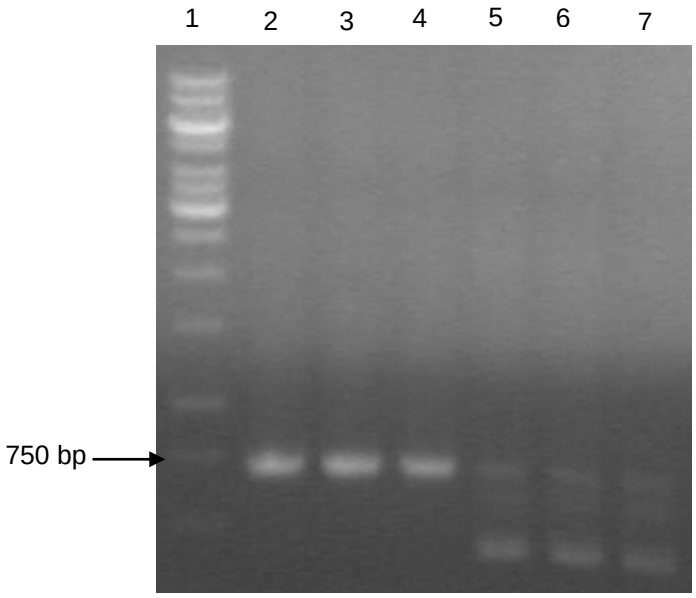


Figure 2.24: Agarose gel electrophoresis (1%) of PCR using Henuni F1 and His-Xba primers to confirm ligation of MAC10 and His-tag gene into *Xho* I and *Xba* I restriction enzyme digested pGE20. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 to 4 = amplicon(s) obtained by PCR of ligation product with Henuni primers, 5 to 7 = amplicon(s) obtained by PCR of ligation product with Henuni F1 and His-Xba primers.

Sequencing of the ligation reaction of the His-tag to the PCR amplicon indicated that *Eco* RI digestion of the PCR amplicon had occurred correctly (GAA TT instead of GAA TTC). Never the less sequencing also indicated that ligation of the His-tag had not occurred (figure 2.25).

$(T)_n$CTGATGGCGGCCGTGACGTTGGACGAGTCCGGGGGCGGCCTCCAGACGCCCGGAGGAGGGCT
CAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAGCAGTTACAGCATGGGTTGGGTGCGACAGGC
GCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTGATGATGATGGTAGTTTCACACACTACGGGTC
GGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACAACGGGCAGAGCACAGTGAGGCTGCAGCTG
AACAACCTCAGGGCTGAGGACACCGGCACCTACTACTGCGCCAAAAATTATTATATTATACGATTATC
GACGCATGGGGCCAGGGACCGAAGTCATCGTCTCCTCCGGTGGAGGCGGTTCAGGTGGAGGTGGC
TCTGGCGGAGGCGGATCGGCGCTGACTCAGCCGTCCTCGGTGTCAGCGAACCCGGGAGAAATCGT
CGAGATCACCTGCTCCGGGAGTGACAGCAGCAACAACTATGGCTGGTATCAGCAGAAGGCACCTGG
CAGTGCCCCTGTCACTGTGATCTATGACAACACCAACAGACCCTCGAACATCCCTTCACGATTCTCC
GGTTCCACATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCAAGCCGAGGACGAGGCTGCC
TATTTCTGTGCGAGTACAGACAGCATATTTGGGGCCGGGACAACCCTGACCGTCCTAGGCCAGCCC
AATGAATT$(A)_m$

Figure 2.25: Sequence of amplicon product obtained after ligation of His-tag annealing reaction to Nested PCR amplicon. Underlined are the first two and last two codons of the MAC10 scFv gene.

Restriction enzyme digestion of plasmid pAK400 with Sfi I

A simpler cloning method used the pAK400 plasmid instead of pGE20. The pAK100 and pAK400 plasmids are phage display and scFv expression plasmids respectively. Cloning into the pAK plasmid vectors is simpler because it uses only one restriction enzyme *Sfi* I. This is because of the unique sequence of the *Sfi* I restriction site. Between the two GGCC repeats there are 5 non specific nucleotides, thus allowing for directional cloning. In the pAK plasmids the non specific nucleotides of the two *Sfi* I sites are completely different ensuring directional cloning. Between the two *Sfi* restriction sites is a 2101 bp tet cassette which due to its size allows for much easier separation of double digested plasmid from single cut or uncut. The plasmids pAK100 and pAK400 were digested with *Sfi* I to remove the tetracycline (tet) resistance cassette to allow for the ligation of the MAC10 scFv gene (figure 2.26). As expected for both plasmid digests only two bands were obtained, suggesting complete digestion had occurred. The tet cassette band (2101 bp) was expected close to the 2000 bp fragment of the ladder. Digested pAK100 (4325 bp) was expected between the 4000 and 5000 bp fragments of the ladder. The digested pAK400 band (3836 bp) was expected near to be near the 4000 bp fragment.

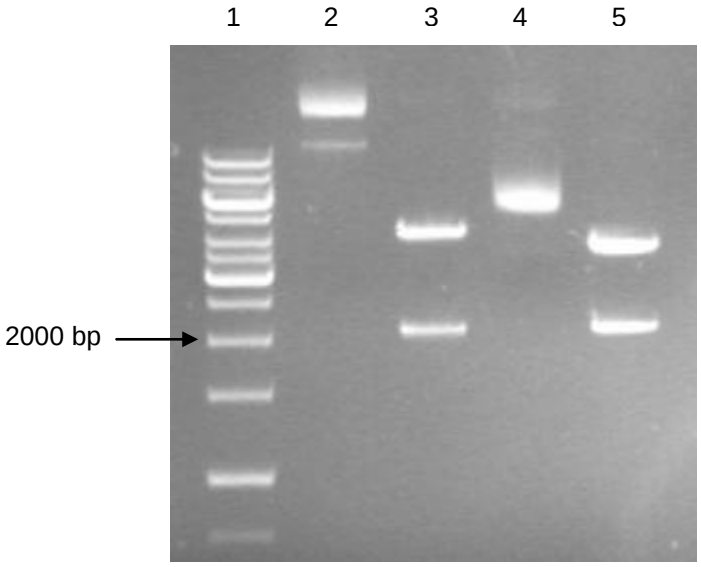


Figure 2.26: Agarose gel electrophoresis (1%) of *Sfi* I restriction enzyme digestion of pAK100 and pAK400. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pAK100, 3 = *Sfi* I cut pAK100, 4 = pAK400, 5 = *Sfi* I cut pAK400.

Nested Polymerase chain reaction with pAK primers

Nested PCR was performed to obtain an amplicon of the MAC10 scFv gene from the MB2 and pAKreverse primers. The PCR was done in duplicate and was expected to yield an amplicon of 754 bp in length. The amplicons obtained were consequently run on 1% agarose gel electrophoresis as seen in figure 2.27. In both of the duplicates a clear band was obtained from 10 µl of the 50 µl reaction close to the 750 bp fragment of the ladder.

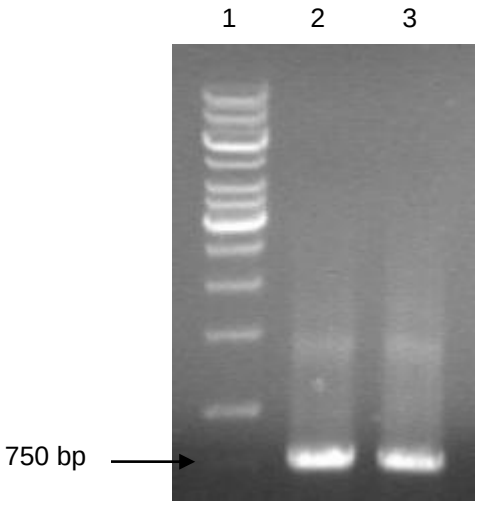


Figure 2.27: Agarose gel electrophoresis (1%) of Nested PCR amplicons from the MB2 and pAKreverse primers with MB primer amplicon as template. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 and 3 = amplicon obtained by nested PCR of MB primer amplicon and MB2 and pAKreverse primers.

Sequencing of the amplicon obtained from the primers MB2 and pAKreverse indicated that the *Sfi* I site had been added to the 5' end of the gene. Also the nested PCR strategy worked as expected as no non specific annealing of the primers had occurred. Only a few of the bases upstream of the MB2 *Sfi* I site were sequenced, but would be cut off by *Sfi* I in any case (figure 2.28).

AACTGC**GGCCCAGCCGGCC**CTGATGGCGGCCGTGACGTTGGACGAGTCCGGGGGCGGCCTCCAG
ACGCCCGGAGGAGGGCTCAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAGCAGTTACAGCATG
GGTTGGGTGCGACAGGCGCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTGATGATGATGGTAG
TTTCACACACTACGGGTCGGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACAACGGGCAGAGCA
CAGTGAGGCTGCAGCTGAACAACCTCAGGGCTGAGGACACCGGCACCTACTACTGCGCCAAAAAGA
TTATTATATTATACGATTATCGACGCATGGGGCCAGGGACCGAAGTCATCGTCTCCTCCGGTGGAGG
CGGTTCAGGTGGAGGTGGCTCTGGCGGAGGCGGATCGGCGCTGACTCAGCCGTCCTCGGTGTCAG
CGAACCCGGGAGAAATCGTCGAGATCACCTGCTCCGGGAGTGACAGCAGCAACAACTATGGCTGGT
ATCAGCAGAAGGCACCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACACCAACAGACCCTCGA
ACATCCCTTCACGATTCTCCGGTTCCACATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCA
AGCCGAGGACGAGGCTGCCTATTTCTGTGCGAGTACAGACAGCATATTTGGGGCCGGGACAACCCT
GACCGTCCTAGGCCAGCCCAATGC**GGCCGCGGGGGCC**C(A)$_m$

Figure 2.28: Sequence of Nested PCR amplicon from the MB2 and pAKreverse primers. Indicated in bold are the *Sfi* I restriction sites, m = uncertain number of adenine nucleotides.

Ligation of amplicon from pAK primers into plasmid pAK400 with T4 DNA ligase

The *Sfi* I digested PCR amplicon obtained from the MB2 and pAK reverse primers was ligated into the pAK400 plasmid with vector: insert ratios of 1.5:1 and 1:3. The number of colonies that grew for both ratios was about the same, a transformation efficiency of approximately 1 x 10^{-6} cfu/µg was obtained for the uncut pAK400 control. Plasmid was isolated from colonies picked and run on agarose gel electrophoresis as shown in figure 2.29. Digestion of these plasmids yielded two bands. The largest band was at the same position as the largest digested band of pAK400, near the 4000 bp fragment of the ladder. The second smaller and slightly fainter band was close to the 750 bp fragment of the ladder. Thus it was able to show successful cloning of the MAC10 scFv gene into pAK400.

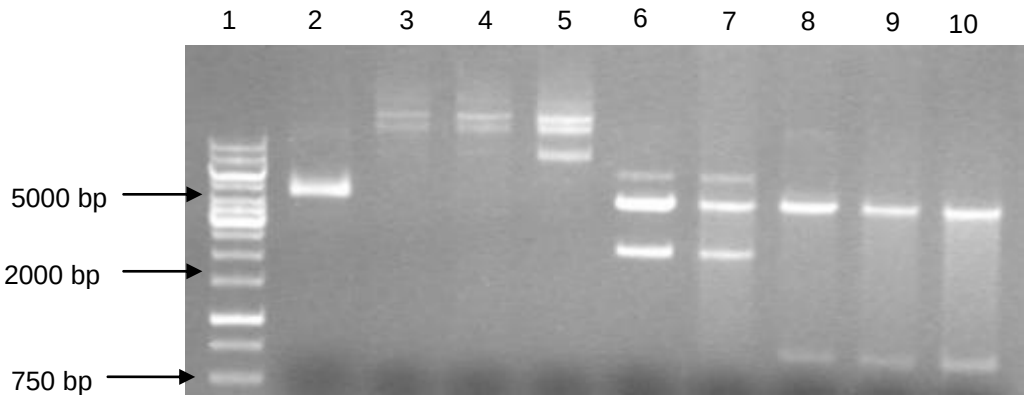


Figure 2.29: Agarose gel electrophoresis (1%) of pAK400 and pAKJS clones. As well as *Sfi* I restriction enzyme digestion of pAK400 and the pAKJS clones. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = pAK400, 3 to 5 = pAKJS, 6 and 7 = *Sfi* I cut pAK400, 8 to 10 = *Sfi* I cut pAKJS.

Sequencing of one of the clones indicated that the MAC10 scFv gene had been cloned into pAK400. The plasmid was identified by the T7g10 Shine-Dalgarno sequence and the His-tag (figure 2.30).

GAGAAGGAGATATACATATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGC**GGC**
CCAGCCGGCCCTGATGGCGGCCGTGACGTTGGACGAGTCCGGGGGCGGCCTCCAGACGCCCGGA
GGAGGGCTCAGCCTCGTCTGCAAGGCCTCCGGGTTCACCTTAAGCAGTTACAGCATGGGTTGGGTG
CGACAGGCGCCTGGCAAAGGGCTGGAATGGGTCGCTGGTATTGATGATGATGGTAGTTTCACACAC
TACGGGTCGGCGGTGAAGGGCCGTGCCACCATCTCGAGGGACAACGGGCAGAGCACAGTGAGGCT
GCAGCTGAACAACCTCAGGGCTGAGGACACCGGCACCTACTACTGCGCCAAAAAGATTATTATATTA
TACGATTATCGACGCATGGGGCCAGGGACCGAAGTCATCGTCTCCTCCGGTGGAGGCGGTTCAGGT
GGAGGTGGCTCTGGCGGAGGCGGATCGGCGCTGACTCAGCCGTCCTCGGTGTCAGCGAACCCGG
GAGAAATCGTCGAGATCACCTGCTCCGGGAGTGACAGCAGCAACAACTATGGCTGGTATCAGCAGA
AGGCACCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACACCAACAGACCCTCGAACATCCCTTC
ACGATTCTCCGGTTCCACATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCAAGCCGAGGA
CGAGGCTGCCTATTTCTGTGCGAGTACAGACAGCATATTTGGGGCCGGGACAACCCTGACCGTCCT
AGGCCAGCCCAATGC**GGCCGCGGGGGCC**GATCACCATCATCACCATCAT**TAG**AAA

Figure 2.30: Sequence of the MAC10 scFv gene obtained by semi-nested PCR using MB2 and pAKreverse primers with pMAC10 template, cloned into *Sfi* I restriction enzyme digested pAK400. Underlined is the T7g10 Shine-Dalgarno sequence, the first two codons of the *pelB* leader gene and the His-tag gene. In bold are the *Sfi* I restriction sites and the stop codon.

Discussion

Specific amplification of high GC content genes is difficult to achieve with a standard PCR. The MAC 10 scFv gene is an example due to the 5'-GGCGGCCG-3' sequence located at the 5'-end of the gene. The problem of how to specifically amplify a gene with a high GC content was encountered soon after the development of the PCR (Haqqi *et al*., 1988; Sarkar *et al*., 1990 and Dutton *et al*., 1993). The non specific amplification of high GC genes has been attributed to reduced denaturation at 94^{o}C, reduced access to the primer due to secondary structure and slower elongation due to secondary structure in the template (Dutton *et al*., 1993). Consequently because the primer cannot access its intended target it is likely to 'seek out' other similar sequences to which it can anneal. This then leads to the amplification of several bands which can then be seen on agarose gel electrophoresis, instead of a single defined band as should be the case. The primers, then because of their high GC content, as in this case can form many strong triple bond base pairs with the template, making annealing at non specific sites favorable.

Extending the length of the Henuni F primer so that it resembles MB2 would not resolve the issue as it would further increase the already high Tm and the likelihood of secondary structures. Some of the suggestions to increase the specificity of amplification include nested PCR (Haqqi *et al*., 1988), the addition of denaturant additives (Sarkar *et al*., 1990) or increasing the PCR cycle temperatures above standard (Dutton *et al*., 1993). Nested PCR is the best solution for the specific amplification of the MAC 10 gene as it requires the least amount of optimization. The MB primers and the cycling parameters used in the first round of nested PCR were previously designed and optimized (van Wyngaardt *et al*., 2004). Also the cycling parameters serve as a starting point from which to design the parameters for the second round. In contrast to the other methods nested PCR eliminates the non specific sequences outside the first set of primers thus restricting the second set of primers to the desired annealing positions.

Figure 2.12 A shows the sequence of the amplicon obtained by PCR of the MAC10 gene using the Henuni primers. Even at a very high annealing temperature of 70^{o}C there was an upstream sequence amplified along with the gene. This included not only the pelB leader gene from pHEN-1 but also a section of plasmid. The sequence of the upstream section beyond the pelB leader could not be corrected because it is beyond sequencing range. In addition the sequence of this section of the pHEN-1 plasmid was not published (Hoogenboom *et al*., 1991). Illustrated by figure 2.12 B is that it is known that the Henuni forward primer anneals upstream of where it should but probably close by. Agarose gel electrophoresis (figure 2.12 C) shows that the amplicon produced is very close to 750 bp in size. However when the amplicon obtained from a PCR using the same primers with an annealing temperature of 65^{o}C is ligated into restriction digested pGE20 the non specific annealing of the Henuni primers becomes clearer. As shown in figure 2.15 C both primers bind non specifically outside the regions they were designed to anneal to. The result is an amplicon of about 1200 bp instead of 750 bp, of which 820 bp is shown in figure 2.15 C. The sequences upstream and downstream of the 820 base pairs were not corrected since there is no literature reference and are shown in figures 2.15 D and E (Hoogenboom *et al*., 1991). For both pGEJS1.1 and 2.1 these sequences appear to have some similarity further confirming their pHEN-1 origin. As previously shown increasing the annealing temperature during PCR to 70^{o}C removes the non specificity of the reverse Henuni primer but the forward primer continues to annealing more favorably upstream of its intended position. The confirmation PCR as shown in figure 2.18 shows how favorable it is for each primer to bind to more than one location resulting in the formation of several amplicons.

In order to eliminate cycling parameters as the cause of non specific annealing of the Henuni primers, a PCR using the MB primers was run for 30 cycles. The original optimized reaction used the standard of 25 cycles. Illustrated in figure 2.13 A is the annealing positions of the MB primers (van Wyngaardt *et al*., 2004). The 30 cycles results as expected in a high concentration of amplicon product as shown in figure 2.13 B and further confirmed by Nanodrop spectrometry analysis. Sequencing determined that the MB primers annealed at the desired positions on pMAC10 plasmid despite the annealing temperature being much lower than the Tm of the primers and the 30 cycles of amplification. Therefore the increased number of cycles is not responsible for the non specific annealing of the Henuni primers which are similar in sequence to the MB primers but shorter and with much lower Tm.

Figure 2.14 shows that Nested PCR resolved the issue of non specific annealing of the Henuni primers. The illustration in figure 2.14 A is to show that using a PCR amplicon as template for the amplification forces the primers to bind only to the correct site. Agarose gel electrophoresis (figure 2.14 B) shows that a single band of the desired size was obtained. Again a high number of cycles (30) were used to eliminate this as the source of any errors in the amplicon product. Also to confirm that Nested PCR with a high number of cycles can be used for the error prone PCR aspect of this project. Sequencing of the Nested PCR amplicon (figure 2.14 C) confirms that the Nested PCR was successful. At the 5'-end the amplicon starts with the CTG ATG just downstream of the *Sfi* I in pMAC10. At the 3'-end the amplicon ends with CCC AAT upstream of the *Not* I site in pMAC10 followed by the *Eco* RI site from Henuni RI. At the 5' and 3' end Taq polymerase added extra thymidines and adenine nucleotides as expected.

Due to the repetitive sequence of the single stranded His-tag oligonucleotides it is possible for the primers to misalign during the annealing reaction. The misaligned double stranded oligonucleotides would be numerous and stable and therefore ligation to plasmid or PCR amplicon would not be possible. The solution would be to design the His-tag gene with two or more of the 5'-CAT-3' codons replaced with 5'-CAC-3'. Thereby the repetitive sequence is broken and the chance for misaligned double stranded oligonucleotides is reduced (Krebber *et al*., 1997 and Mohanty *et al*., 2004). Similar methods were used for the construction of plasmids with polyhistidine tags with six or ten histidines. Also once annealed the double stranded oligonucleotides coding for the His-tags were designed to have restriction sites at their end to allow for ligation to a plasmid. The difference was that the His-tag coding regions were only a small region of the entire oligonucleotide with the shortest of these being 75 bp in length. Secondary structure formation was not a concern as the single stranded oligonucleotide mixtures were only boiled and cooled on ice which would allow for annealing (Mohanty *et al*., 2004).

Ligation of the His-tag double stranded oligonucleotide to the nested PCR amplicon should have generated a fragment able to ligate to the *Xba* I site of pGE20. Without the His-tag the amplicon ends in the *Eco* RI site added by Henuni RI. Therefore any plasmid to which only the amplicon ligated would not circularize. Yet as figures 2.23 and 2.24 show this is not true. Not only is the amplicon cloned into pGE20 alone, but it is able to transform competent *Escherichia coli*. The confirmation PCR in figure 2.24 clearly shows that the MAC10 gene amplicon is present in the recombinant plasmid and the His-tag oligonucleotide is not. It is likely that at the 3'-end of the amplicon interaction with pGE20 was caused by the non dephosphorylation of the plasmid even though *Xba* I and *Eco* RI sites are not compatible. The cloning of the MAC10 gene without the His-tag was further assisted by the blunt end ligation of the 5' end of the amplicon to the blunted *Xho* I site of pGE20.

The cloning of the MAC10 gene into plasmid pGE20 was complicated by the presence of an *Xho* I restriction site inside the gene (underlined in figure 2.12 A). This is the cause of the complicated cloning strategy into plasmid pGE20. If this plasmid is to be considered further for subcloning of the MAC10 gene the *Xho* I restriction site should be eliminated in some way.

The sub cloning of the MAC10 scFv gene from pHEN-1 into pAK400 followed a similar method as with pGE20. Except that in this case it is only a semi nested PCR used because both the first and second rounds of PCR use the same forward primer, MB2. The reverse primer for the second round pAKreverse was designed to add a *Sfi* I restriction site to the 3'-end of the MAC10 gene to replace the Not I restriction site. Since *Sfi* I restriction sites have 5 random nucleotides, the same enzyme can be used to digest two essentially unique restriction sites. The digested pAK400 plasmid is easily identified as shown in figure 26, showing clearly the linear plasmid (3836 bp) and Tet cassette (2101 bp) (Krebber et al, 1997). Also what makes the pAK400 cloning method much more efficient is that this plasmid vector already contains its own His-tag. The PCR to amplify the MAC10 gene (figures 2.27 and 2.28) for cloning is only a semi nested PCR since the forward primer is the same for both rounds of amplification. However the high GC% of both primers is a clear indication that the nested PCR method remains necessary. Comparing the recombinant pAKJS plasmid to parental pAK400 by agarose gel (figure 2.29) shows that a change has occurred after ligation. The number of plasmid bands has increased to three (probably linear, circular and supercoiled) as well as the position of the plasmid. Further comparison of *Sfi* I digested pAKJS with digested pAK400 shows that a fragment of approximately 750 bp was cut out of the recombinant plasmid. Sequencing (figure 2.30) was able to identify both the amplicon inserted and the plasmid confirming successful in-frame cloning of the MAC10 scFv gene.

Conclusion

Due to a high percentage of guanine and cytosine nucleotides at the 5'and 3'ends of the MAC10 scFv gene PCR amplification is not straight forward. A nested PCR method has been investigated to eliminate the non specific annealing caused by the high GC percentage. The nested PCR method has been found to be a reliable and efficient method for amplification. The first round of amplification uses previously designed primers that amplify the scFv gene along with adjoining sequences from the plasmid pMAC10. The second round can then uses nested primers designed to produce an amplicon for subcloning. Two primer pairs were tested to investigate the applicability of the nested PCR method. Both produced specific amplicons which ligated to their respective plasmids. Since the 5' and 3' ends of all scFv genes from the Nkuku library are identical the nested PCR method is applicable to the entire library. Two plasmids were investigated for subcloning of the scFv gene, pGE20 and pAK400. Of the plasmid vectors investigated for subcloning pAK400 is the most suitable since it only requires a single step and one restriction enzyme to prepare the plasmid for cloning compared to three for pGE20.

Chapter 3 : Expression of anti-mycolic acid single chain variable fragment by TG-1 and HB2151 *Escherichia coli* strains from the pHEN-1 and pAK400 plasmids.

Introduction

During phage display scFv proteins are fused to the bacteriophage coat. This directly links the DNA sequence contained in the plasmid carried by the bacteriophage and the protein it codes for. This is the innate advantage of phage display which allows for a rapid screening and selection of binders from a group during the same experiment (McCafferty *et al*., 1990, Hoogenboom *et al*., 1991 and Krebber *et al*., 1997). However once selection of the desired scFv has been completed, the bacteriophage is no longer required. It may therefore subsequently become desirable to express the scFv as a 'soluble' protein separate from the bacteriophage (Hoogenboom *et al*., 1991, van Wyngaardt *et al*., 2004 and Krebber *et al*., 1997).

The pHEN-1 plasmid was constructed with an amber codon between the *c-myc* tag gene and gene 3(g3) of the M13 bacteriophage. The corresponding, gene 3 protein (g3p) is the minor coat protein located at one tip of the bacteriophage. Adsorption to the *Escherichia coli* host pilus is facilitated by the gene 3 protein which form knob like structures on the surface of the bacteriophage. Adsorption to the pilus is the initiation of infection of the *Escherichia coli* host. It is on the N-terminals of gene 3 proteins that proteins are presented during phage display. Phagemid vectors such as pHEN-1 can express the scFv gene as either a fusion to protein g3p or as soluble fragments. This is dependent on the *Escherichia coli* expression host being used. The type of strain used for phage display is called a suppressor (*sup*E) strain. Suppressor strains such as *Escherichia coli* TG-1 and XL-1 blue do not recognize the amber codon (TAG). Instead the amino acid glutamine is inserted in the place of the stop codon. The result is that the scFv protein is expressed as a fusion with the g3p of the bacteriophage envelope. A non suppressor strain on the other hand recognizes the amber stop codon. A scFv expressed by a non suppressor strain is produced as a soluble protein without fusion to the g3p (Hoogenboom *et al*., 1991 and Barderas *et al*., 2006).

The plasmid pAK400 is a derivative of pAK100 which has been enhanced for the soluble expression of scFv proteins. In order to do this the second Shine-Dalgarno sequence of pAK100 was replaced by the bacteriophage T7 gene 10 (T7g10) Shine-Dalgarno sequence, and the bacteriophage gene 3 was removed (Krebber et al., 1997). In prokaryotic translation, the Shine-Dalgarno sequence is part of the ribosome binding site that ensures interaction between rRNA and mRNA occurs in the correct reading frame. The T7g10 Shine-Dalgarno sequence is classified as strong, as it has been shown to result in the accumulation of transcription and translation products of foreign genes. It has been shown to be able to efficiently enhance foreign protein expression more than a hundred fold (Olins *et al*., 1989 and Makrides, 1996). In addition the bacteriophage coat protein gene 3 has been removed in plasmid pAK400 as it is no longer required.

In the pAK series of plasmids expression of the scFv proteins is placed under the control of the *lac* promoter. The *lac* promoter forms part of the *lac* operon of *Escherichia coli*. An operon is a cluster of genes and control elements arranged together on the *Escherichia coli* chromosome into a unit. The *lac* operon consists of the genes for three proteins, β-galactosidase (Z); galactoside permease (Y) and thiogalactoside transacetylase (A), the cell needs to metabolize lactose, as well as regulatory sequences (operator and promoter). When lactose is not present a repressor protein is produced that binds to the operator, thus preventing transcription. When lactose is present it is converted to the inducer molecule allolactose which binds the repressor protein. Once the allolactose has bound to the repressor protein, it can no longer bind to the operator and transcription is initiated (Voet et al., 2004).

Thus the *lac* operon promoter is termed an inducible promoter. The inducible character of the *lac* promoter makes it ideal the expression of foreign protein genes cloned into plasmids. The transcription of genes under the control of the *lac* promoter can be controlled by either including the repressor gene in the plasmid, or by adding glucose to the growth medium during the logarithmic growth phase of the transformed *Escherichia coli*. Without being repressed, and in the absence of an inducer, the *lac* promoter allows for constitutive expression of genes under its control. Glucose is the preferred catabolite of *Escherichia coli*, therefore its presence in the growth medium results in repression of transcription of genes involved in the metabolism of other sugars. This catabolite repression can be used to control the expression of genes under the control of the *lac* promoter and prevent constitutive expression. For the *lac* promoter the most widely used inducer is the synthetic compound isopropylthiogalactoside (IPTG) (Voet *et al*., 2004; De Bellis *et al*., 1990 and Krebber *et al*., 1997).

The suppression of stop codons (amber, ochre and opal) in certain strains of *Escherichia coli* is due to single base mutations in the anti-codon pairs of transfer RNAs (tRNA). As with other prokaryotes, *Escherichia coli*, normally does not possess tRNAs that recognizes stop codons. However so-called suppressor mutations in the genes for certain tRNA result in anti-codons for these three codons. In the nomenclature of suppressor strain the *sup* refers to suppressor. While the capital letter (E) at the end designates the gene that had been mutated. The wild type of the gene is differentiated from the mutant by a plus sign (*sup*E^+ indicates the wild type). In the case of *sup*E the gene that has been mutated is the *glnV* (*α, β*) in which the anti-codon change was from CUG to CUA (Eggertsson *et al*., 1988).

Polyhistidine tagged fusion proteins can be purified from a crude mixture of proteins due to the affinity of the histidine imidazole rings for chelated transition metal ions. Metal ions like nickel are chelated by an immobilized ligand such as nitrilotriacetic acid (NTA) allowing for immobilized metal ion affinity chromatography (IMAC). Thus the high affinity polyhistidine peptide interacts with the metal ion and is separated from the *Escherichia coli* host cell proteins (Arnau *et al*., 2005 and Waugh, 2005).

The mycolic acid clone 10 (MAC10) scFv was selected from the Nkuku library. The next step is expression as the soluble protein. The protein was expressed in *Escherichia coli* TG-1 and HB2151 host strains using the recombinant plasmid pMAC10. Plasmid pAKJS the product of sub-cloning was also used for expression in the *Escherichia coli* HB2151 host strain.

Materials and Methods

Flow diagram of expression from plasmids pMAC10 and pAKJS

The following is a summarized flow diagram of the strategies used in the expression of the MAC10 scFv from two different plasmids and strains of *Escherichia coli*.

Transform competent *Escherichia coli* TG-1 cells with pMAC10

↓

Periplasmic expression of MAC10 scFv

↓

Extraction from periplasm by osmotic shock

Transform competent *Escherichia coli* HB2151 cells with pMAC10

↓

Periplasmic expression of MAC10 scFv

↓

Extraction from periplasm by osmotic shock

Transform competent *Escherichia coli* HB2151 cells with pAKJS

↓

Periplasmic expression of MAC10 scFv

↓

Extraction using TES buffer but no osmotic shock

Transform competent *Escherichia coli* HB2151 cells with pAKJS

↓

Periplasmic expression of MAC10 scFv

↓

Extraction using TES buffer with osmotic shock

Preparation of chemically competent *Escherichia coli*

For the preparation of chemically ($CaCl_2$) competent *Escherichia coli* a previously described method was followed (Ausubel *et al*., 2000). Briefly, competent *E. coli* cells (TG1 and HB2151) were prepared from an overnight culture (16 h at 37°C) shaking at 200 rpm. The overnight culture was grown from a single colony picked off an LB agar plate (1% tryptone, 0.5% yeast extract, 1% NaCl and 1.4% (w/v) agar) into liquid LB (10 ml). Of this culture 1 ml was diluted 100 x and grown with shaking (200 rpm) until an OD_{550} of approximately 0.6 was reached. The 100 ml culture was split into two 50 ml tubes and centrifuged (11 900 rcf, 20 – 25 °C for 8 min). The cell pellets were re-suspended and combined in ice cold 100 mM $CaCl_2$ in a single tube. The cells were again pelleted by centrifugation (11 900 rcf, 20 – 25 °C for 8 min). The resultant pellet was re-suspended in 50 ml ice cold 100 mM $CaCl_2$ and left on ice overnight at 4°C. The cells were subsequently pelleted by centrifugation (11 900 rcf, 4°C for 8 min). The cell pellet was then re-suspended in 86% (v/v) ice cold 100 mM $CaCl_2$ and 14% (v/v) glycerol. This solution was then split into 0.2 ml aliquots in pre-cooled eppendorff tubes and stored at -70°C until needed.

Transformation of chemically competent *Escherichia coli*

To transform competent cells of *Escherichia coli* TG1 and HB2151 strains approximately 100 ng of plasmid DNA was added to 50 µl of competent cells. This was then incubated on ice for 30 min. Following this, the cells were heat shocked (42 °C for 2 min). Pre- warmed (37 °C) LB medium (1 ml) was added immediately to the cells which were then incubated in a water bath (37 °C) for 1 h. After incubation, 100 µl aliquots of transformed cells was plated on either LB-ampicillin (100 µg/ml ampicillin) agar or non expression (NE) (1.6% tryptone, 1% yeast extract, 0.5% NaCl, 1% glucose and 25 µg/ml chloramphenicol) plates, which were incubated overnight (37°C).

Overnight culture of transformed *Escherichia coli*

Cultures were obtained by selecting single *E. coli* colonies from LB-ampicillin (100 µg/ml) or NE (25 µg/ml chloramphenicol) agar plates and inoculating them into liquid LB-ampicillin (100 µg/ml) or NE (25 µg/ml chloramphenicol) medium. The cells were incubated at 37°C for 16 h with shaking at 200 rpm.

Expression of an anti-mycolic acid single chain variable fragment in TG-1 and HB2151 cells

The MAC10 scFv was expressed as previously described (van Wyngaardt et al, 2004). A 5 ml 2 x TY (1.6% tryptone, 1% yeast extract, 0.5% NaCl) with 2% glucose and 100 µg/ml ampicillin and overnight pre-culture was added to 45 ml of 2 x TY with 2% glucose and 100 µg/ml ampicillin and allowed to grow until an OD_{600} of approximately 0.9 was reached. Thereafter the cells were collected by centrifugation (6000 rcf for 10 min at 25°C) and resuspended. Induced samples were resuspended in 1/5 volume of 2 x TY with 100 µg/ml ampicillin. Expression was induced by the addition of 1 mM IPTG. Uninduced samples were resuspended in 1/5 volume of 2 x TY with 100 µg/ml ampicillin and 2% glucose. After overnight (16 h) incubation (37°C) the cell pellet was collected by centrifugation (6000 rcf for 10 min at 25°C). The supernatant was kept separate and concentrated.

Expression of an anti-mycolic acid single chain variable fragment in HB2151 cells transformed with pAKJS

To express the MAC10 scFv protein in HB2151 cells transformed with the pAKJS plasmid two different methods were used (Krebber *et al*, 1997 and van Wyngaardt *et al*, 2004). In the former method 20 ml of expression medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl and 25 µg/ml chloramphenicol) was inoculated with pre-culture (200 µl or 1 ml) of non expression (NE) medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl 1% glucose and 25 µg/ml chloramphenicol) and incubated at 30°C shaking at 200 rpm until an OD_{600} of 0.5 was reached. Expression was induced by the addition of 1 mM IPTG after 4 h in a shaking incubator (200 rpm at 30°C) the cell pellet was collected by centrifugation (6000 rcf for 10 min at 25°C).

Alternatively the latter method was adapted from that used to express from pMAC10 in TG-1 cells. A 5 ml overnight pre-culture of NE medium was added to 45 ml NE medium and kept in a shaking incubator (200 rpm at 30°C) until an OD_{600} of approximately 0.9 was reached. The cells were then collected by centrifugation (6000 rcf for 10 min at 25°C) and resuspended in 1/5 volume (10 ml) of expression medium. Induction was suspended after 4 h or 16 h and the cell pellet collected by centrifugation (6000 rcf for 10 min at 25°C).

Periplasmic extraction

The cell pellet was collected by centrifugation and was subsequently resuspended in 500 µl of ice cold TES buffer (30 mM Tris (pH 8), 5 mM EDTA, 20% Sucrose). To which 750 µl of ice cold 1/5 diluted TES buffer in $ddd H_2O$ was added. After incubating on ice (30 min) the cell suspension was centrifuged (11 000 rcf for 10 min at 4°C) and the supernatant collected. To that 1 mM phenylmethanesulfonyl fluoride (PMSF) was then added to the supernatant which was subsequently frozen at -20°C.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis is a qualitative method for the analysis of protein mixtures based on separation according to size (Wilson *et al*., 2005). To prepare the separating gel (12% acrylamide), 5.3 ml of $ddd H_2O$, 3.7 ml of 1.5 M 4x Tris-HCl / 14 mM SDS (pH 8.8), 6 ml of 30% acrylamide/0.8% N,N'-methylenebisacrylamide, 10 µl of TEMED and 200 µl of 10% ammonium persulfate (APS) were added together. The stacking gel (5% acrylamide) consisting of 3 ml of $ddd H_2O$, 1.3 ml of 0.5 M 4x Tris-HCl /35 mM SDS (pH 6.8), 0.9 ml of 30% acrylamide/0.8% N,N'-methylenebisacrylamide, 5 µl of TEMED and 80 µl of 10% APS was added over the polymerized separating gel. After polymerization of the stacking gel, cold 1x SDS electrophoresis buffer (800 ml of $ddd H_2O$ and 200 ml of 5x SDS electrophoresis buffer [125 mM Tris, 960 mM glycine, 17.3 mM SDS and $ddd H_2O$ up to 1 L]) was added to the apparatus. The gel was loaded with sample (30 µl of protein sample and 10 µl of loading dye (50 mM Tris-HCl; (pH 6.8), 12.5 mM EDTA, 2% SDS, 1% β-Mercaptoethanol, 10% Glycerol and 0.02 % Bromophenol Blue) and 10 µl of PageRuler™ unstained protein ladder (Fermentas). The gel was then run at 30 mA for 2 h. The gel was then stained overnight in Coomassie staining solution (3 mM Coomassie brilliant blue R-250, 45% methanol, 10% glacial acetic acid and $ddd H_2O$ up to 1 L), followed by 4 h of destaining in approximately 20 ml wash solution (20% methanol, 10% glacial acetic acid and $ddd H_2O$ up to a volume of 1 L) changing the solution every 2h.

Table 3.1: Composition of a 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis separating gel.

Component	Volume (ml)
$dddH_2O$	5.3
1.5 M 4x Tris-HCl/SDS (pH 8.8)	3.7
30% acrylamide/0.8% N,N'-methylenebisacrylamide	6
TEMED	10 µl
10% ammonium persulfate	200 µl

Table 3.2: Composition of a 5% sodium dodecyl sulfate polyacrylamide gel electrophoresis stacking gel.

Component	Volume (ml)
$dddH_2O$	3
0.5 M 4x Tris-HCl/SDS (pH 6.8)	1.3
30% acrylamide/0.8% N,N'-methylenebisacrylamide	900 µl
TEMED	5 µl
10% ammonium persulfate	80 µl

Results

Expression of anti-mycolic acid single chain variable fragment protein in *Escherichia coli* TG-1 cells

The MAC10 scFv protein was expressed from the pMAC10 plasmid in the *Escherichia coli* TG-1 strain (van Wyngaardt *et al*, 2004). This strain of *Escherichia coli* is a so called suppressor strain which does not recognize the amber codon between the scFv and gene 3 DNA sequence. Therefore it was expected that soluble scFv as well as scFv-g3 protein fusions would be expressed. SDS-PAGE analysis of the periplasmic extracts from untransformed TG-1, uninduced as well as IPTG induced cells showed the following. The extraction process removes a number of proteins from the periplasm as shown below by the extract from the untransformed cells. The uninduced lane has similar bands to untransformed, except for the prominent band between 25 and 30 kDa. This is thought to be the MAC10 scFv protein since all scFv's are approximately 25 kDa in size. Also this band is not present in the periplasmic extract from untransformed cells. Although these are crude extracts which have not been standardized in terms of concentration it seems that induction with IPTG results in a significant increase in the amount of all the periplasmic proteins (figure 1).

SDS-PAGE analysis was also done of the supernatants from the *Escherichia coli* TG-1 transformed with pMAC10. It was expected that some of the expressed scFv's would be exported to the supernatant (van Wyngaardt *et al*., 2004 and Depetris *et al.*, 2008). A number of proteins including the scFv seem to have been exported from the cells into the supernatant. The amount of scFv protein transported into the supernatant is less than expected. Similar to the periplasmic extracts IPTG induction resulted in an increase in the amount of all proteins. SDS-PAGE analysis indicates that the amount of scFv is about the same for the uninduced and induced samples (figure 2).

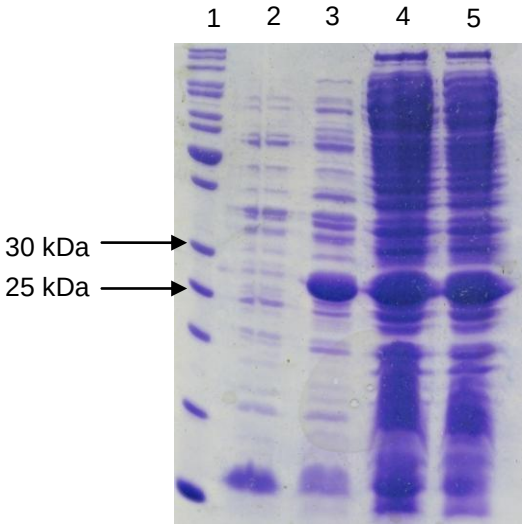


Figure 3.1: **Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of periplasmic extracts obtained from *Escherichia coli* TG-1 cells transformed with plasmid pMAC10.** 1 = PageRuler™ unstained protein ladder, 2 = untransformed *Escherichia coli* TG-1, 3 = uninduced, 4 and 5 IPTG induced.

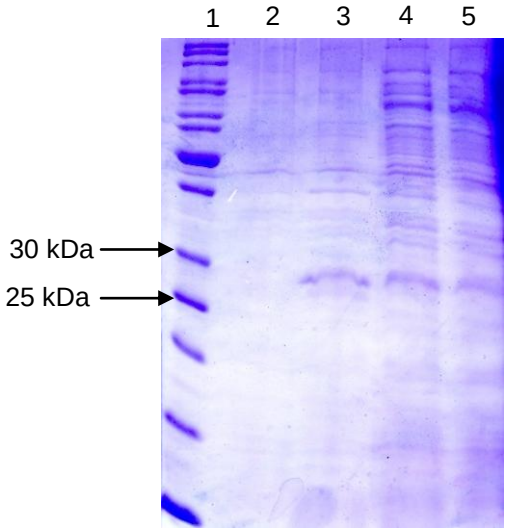


Figure 3.2: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of supernatants obtained from *Escherichia coli* TG-1 cells transformed with the plasmid pMAC10. 1 = PageRuler™ unstained protein ladder, 2 = untransformed *Escherichia coli* TG-1, 3 = uninduced, 4 and 5 IPTG induced.

Expression of anti-mycolic acid single chain variable fragment protein in *Escherichia coli* HB2151 cells from the plasmid pMAC10

The MAC10 scFv protein was expressed in *Escherichia coli* HB2151 transformed with the plasmid pMAC10. Non suppressor strains such as HB2151 recognize the amber stop codon. It is therefore expected that only the soluble form without fusion to the g3 protein will be expressed in HB2151 in contrast to TG-1. As with the TG-1 strain there is some expression of the scFv protein in the uninduced samples. In the induced samples the amount of scFv protein expressed increased relative to the uninduced samples. Although the increase in protein expressed seems not to be as drastic as with the TG-1 strain. Also quite clear is that in both the uninduced and the induced samples the amount of scFv protein is significantly greater than any of the periplasm associated *Escherichia coli* proteins. Induction with IPTG did not increase the amount of the periplasmic *Escherichia coli* proteins when expressing in HB2151 as it did in TG-1.

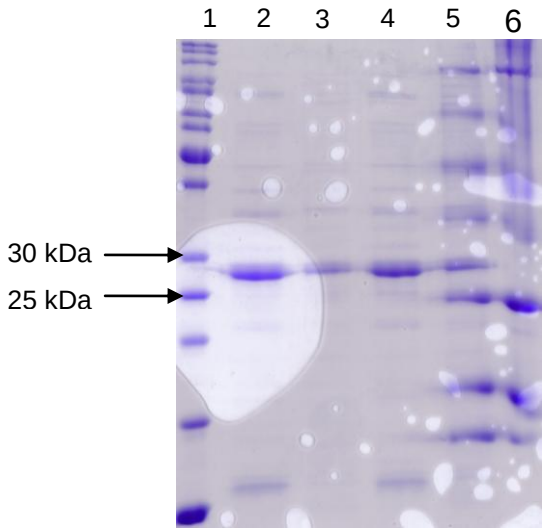


Figure 3.3: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of periplasmic extracts obtained from E*scherichia coli* HB2151 cells transformed with plasmid pMAC10. 1 = PageRuler™ unstained protein ladder, 2 and 4 = IPTG induced E*scherichia coli* HB2151, 3 and 5 = uninduced, 6 = untransformed E*scherichia coli* HB2151.

Expression of anti-mycolic acid single chain variable fragment from the pAKJS plasmid by HB2151 cells

After subcloning the MAC10 scFv gene into the pAK400 plasmid the protein was to be expressed in *Escherichia coli* HB2151 non suppressor strain. There is no difference in the protein composition of the periplasmic extracts from the untransformed cells, uninduced cells and the two groups of induced cells. The expressed scFv protein concentration is too low to be identified from the *Escherichia coli* proteins. Also increasing pre culture volume from 200 µl to 1 ml does not affect protein composition of the extract. It does however increase the amount of protein extracted and decrease the incubation time required to reach the OD_{600} of 0.5.

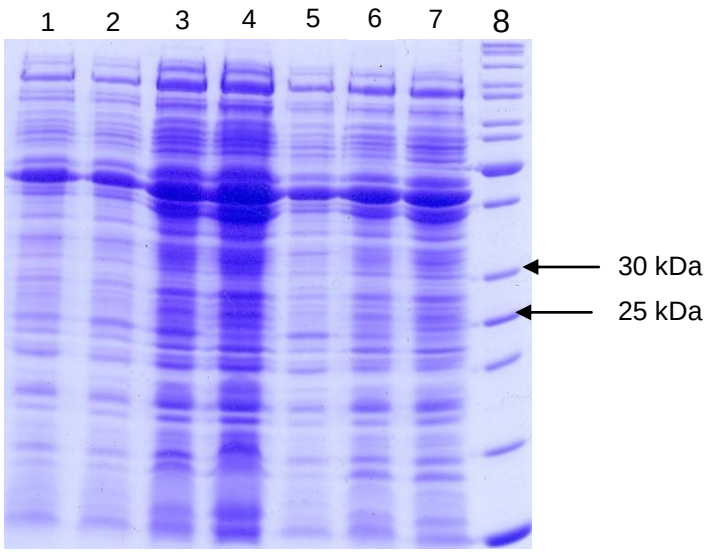


Figure 3.4: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of periplasmic extracts obtained from *Escherichia coli* HB2151 cells transformed with plasmid pAKJS. 1 and 2 = IPTG induced *Escherichia coli* HB2151 from 200 µl of preculture, 3 and 4 = IPTG induced *Escherichia coli* HB2151 from 1 ml preculture, 5 = uninduced, 6 and 7 = untransformed *Escherichia coli* HB2151, 8 = PageRuler™ unstained protein ladder.

In order to investigate an alternative method to express the MAC10 scFv, the TG-1 method was used instead of the method used for soluble expression from pAK plasmids. Using the same method for expression of the scFv from the pAKJS plasmid as from the pMAC10 plasmid does not produce sufficient amounts of protein. Although a larger volume of cells was used (50 ml compared to 20 ml) the amount of protein in the periplasmic extracts is less compared to that of the pAK method. There is still no detectable difference between the untransformed cell extracts and the transformed cell extracts. The profile of proteins is the same for the extracts made after 4 h and 16 h of induction.

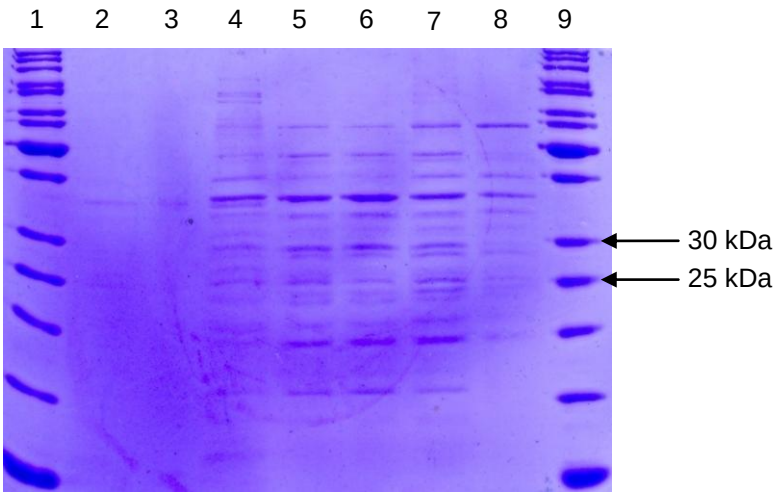


Figure 3.5: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of periplasmic extracts obtained from *Escherichia coli* HB2151 cells transformed with plasmid pAKJS plasmid. 1 and 9 = PageRuler™ unstained protein ladder, 4 = *Escherichia coli* HB2151 induced for 16 h with IPTG, 5 and 6 = *Escherichia coli* HB2151 induced for 4 h with IPTG, 7 = uninduced, 8 = untransformed *Escherichia coli* HB2151.

In order to determine whether the expressed scFv protein is exported to the supernatant (growth medium); HB2151 cells were transformed with the pAKJS plasmid. Periplasmic extracts were obtained and the supernatants were concentrated 10 times. There was no export of the protein to the supernatant, opposite of what was expected for at least the 16 h induction sample. The amount of protein is consistently low for all samples. However induction resulted in an increase in the expression of two *E. coli* proteins. SDS-PAGE analysis could not determine whether expression of MAC10 scFv occurred.

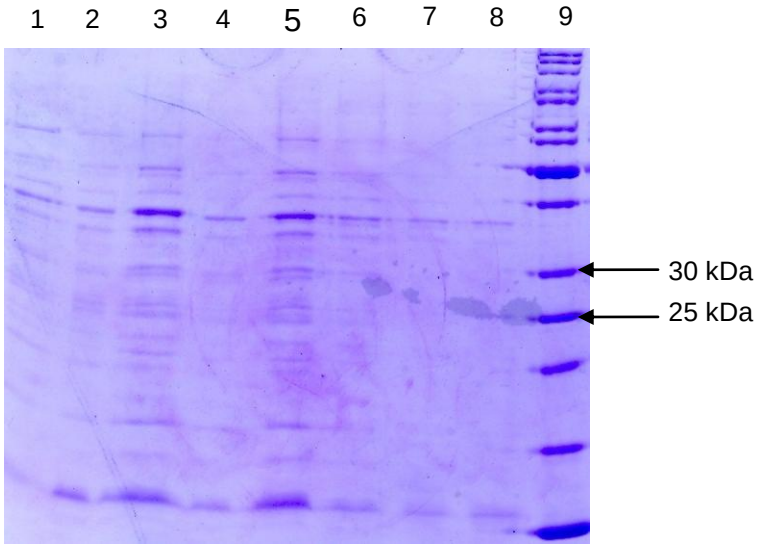


Figure 3.6: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%) of periplasmic extracts and supernatants obtained from *Escherichia coli* HB2151 cells transformed with plasmid pAKJS. 1 = periplasmic extract from *Escherichia coli* HB2151 16 h of IPTG induction, 2 = supernatant from_*Escherichia coli* HB2151 16 h of IPTG induction, 3 = periplasmic extract from *Escherichia coli* HB2151 4 h of IPTG induction, 4 = supernatant from *Escherichia coli* HB2151 4 h of IPTG induction, 5 = supernatant from uninduced *Escherichia coli* HB2151, 6 = periplasmic extract from uninduced *Escherichia coli* HB2151, 7 = supernatant from untransformed *Escherichia coli* HB2151, 8 = periplasmic extract from *Escherichia coli* HB2151, 9 = PageRuler™ unstained protein ladder.

Discussion

Expression of the MAC10 scFv protein was performed using plasmids pMAC10 and pAKJS. The expression strains used were *Escherichia coli* TG-1 and HB2151.

It is known from literature that long term IPTG induced periplasmic expression (more than 4h) would result in export of scFv proteins to the supernatant (Depetris *et al*., 2008, van Wyngaardt *et al*., 2004 and Hoogenboom *et al*., 1991). However as seen in figures 1 and 2 after 16h of expression the amount of scFv as well as the total protein amount is much lower in the supernatant than the bacterial periplasm. Previous periplasmic expression of scFv's from the Nkuku library used the supernatant exclusively as the source of scFv's for further analysis. Another plasmid vector system, also for the periplasmic expression of scFv's in *Escherichia coli* TG-1 did the same. Both expressions used the same method and media as was used to express MAC10. In the latter it was found that after overnight induction with IPTG the majority of scFv activity was in the supernatant and not the periplasm. In contrast with expression of only 4 h where the majority of activity was found in the periplasm (van Wyngaardt *et al*., 2004 and Depetris *et al*., 2008). Comparison between the SDS-PAGE analyses of the periplasm and supernatants of *Escherichia coli* TG-1 expressing MAC10 indicate less scFv protein present in the supernatant. The supernatant was concentrated 5 x prior to SDS-PAGE analysis therefore dilution is unlikely to be the reason for the low amount of protein. There is no evidence directly pointing towards degradation, but cannot be completely dismissed in this case. However in literature when scFv's were isolated from supernatants after 16 h of IPTG induction no mention was made of protein degradation being a problem that needed to be considered (van Wyngaardt *et al*., 2004 and Depetris *et al*., 2008).

Since *E. coli* TG-1 is a suppressor strain it was expected that most of the scFv's would be expressed as fusions to the gene 3 protein. Gene 3 protein has a molecular mass of 44 kDa and together with the 25 kDa of the scFv a band of approximately 70 kDa was expected (Wezenbeek *et al*., 1979). Due to induction the expression of proteins with molecular weights in that region increased drastically. It is therefore too difficult to identify single proteins on the SDS-PAGE gel.

The constitutive expression of the scFv protein is quite clear in figures 3.1 and 3.2 and to a lesser extent figure 3.3. What this means is that even without induction with IPTG the Lac Z promoter allows transcription to occur (De Bellis *et al*., 1990). It has been reported that IPTG induction is not required for expression from plasmid pHEN-1. Induction with IPTG would in fact lead to cell death due to the toxicity of the fusion protein (Hoogenboom *et al*., 1991). The constitutive expression of the scFv in *Escherichia coli* TG-1 allows for some options for expression. The scFv protein can be obtained from a periplasmic extract of an overnight uninduced culture. A periplasmic extract of 4 h induced sample can also be used to obtain scFv protein. Since only 4 h is required to obtain optimal levels of scFv in the periplasm. ScFv protein can also be obtained from the supernatant of an overnight (16 h) IPTG induced culture (Depetris *et al*., 2008). These expression options are of course available to all Nkuku library scFv's using recombinant pHEN-1 and *Escherichia coli* TG-1.

Repression of expression with glucose in the growth medium during log phase growth of was reported to control the LacZ promoter by catabolite repression. However these results indicate that scFv expression from pMAC10 is not repressed by glucose. In both the *Escherichia coli* TG-1 and HB2151 strains scFv protein was expressed by the uninduced samples. This despite there being glucose present in the medium during growth towards an OD_{600} of 0.9 and during the overnight period of 'induction'. This could be due to the glucose being consumed much quicker than expected or that in the case of this pMAC10 glucose has no effect on the LacZ promoter (Voet *et al*., 2004; De Bellis *et al*., 1990).

The constitutive expression of the MAC10 scFv is less in *Escherichia coli* HB2151 than in TG-1. This can be seen from the defined scFv protein bands obtained for uninduced and induced samples in figure 3.3. In comparison with the uninduced samples, the induced samples do not contain much more scFv protein. Thus IPTG induction does not appear to increase scFv expression in HB2151. Induction with IPTG did not result in an increase in the amount of periplasmic proteins in *Escherichia coli* HB2151. The amount of scFv protein appears to be much more than the *Escherichia coli* periplasmic proteins. The only exception is one of the uninduced samples where the all proteins appear to be present in equal amounts in the sample.

The SDS-PAGE analyses in figures 3.4 and 3.5 indicate that the expression of the MAC10 scFv using pAKJS in *Escherichia coli* HB2151 is not as simple as the previous system. There is clearly no constitutive expression of the scFv using pAKJS in *Escherichia coli* HB2151. There is also very little or no expression of the scFv during induction with IPTG. Two different methods of expression were used to express the scFv yet the result was the same. Figure 3.4 shows the SDS-PAGE analysis for the method developed for expression from recombinant pAK400 plasmids. The growth and induction temperatures were increased to 30°C from 24°C to speed up the growth rate of the culture. These temperatures were previously at 24°C to prevent misfolding of proteins and to maintain host cell integrity (Skerra *et al*., 1991). The higher temperatures would therefore not be the reason for non expression but rather the location of the proteins. Also different *Escherichia coli* host strains were used. The MAC10 scFv was expressed in *Escherichia coli* HB2151 while the hybridoma scFv's were expressed in *Escherichia coli* JM83 (Krebber *et al*., 1997).

Another difference is that periplasmic extraction without osmotic shock was used to obtain MAC10 instead of using a French press. Usually during periplasmic extraction the cell culture is grown to OD_{600} in one volume, induction is then done in one fifth of that volume. The French press method would have resulted in a mix of proteins from the cytoplasm as well as the periplasm. Which would mean correctly and incorrectly folded proteins would be mixed. As the SDS-PAGE (figure 3.4) shows other proteins were extracted from the periplasm therefore there is no reason why the MAC10 scFv would not have been. Also since the scFv was induced it was expected to be present in a significant amount in the periplasmic extract.

Figure 3.5 is the SDS-PAGE of the second method used in an attempt to obtain the MAC10 scFv. This time expression was done under the conditions required for periplasmic extraction. Induction was also done for 4 h and 16 h to determine if expression stopped after 4 h. The reasons for looking at 4 h is previous expression from recombinant pAK400 was induced for only that amount of time. As well as the fact that expressed proteins are exported from the periplasm into the supernatant after more than 4 h of expression (Krebber *et al*., 1997 and Depetris *et al*., 2008). Clearly the change in induction time made no difference as to whether the scFv was expressed or not. Although there are proteins present in the 25 to 30 kDa range there is no difference in expression levels between uninduced and induced. Interestingly less periplasmic proteins were obtained from these extractions than for those of figure 3.4.

The supernatants and periplasmic extracts of the 4 and 16 h induced samples were subsequently compared by SDS-PAGE analysis. According to the SDS-PAGE analysis shown in figure 6 there are some periplasmic *Escherichia coli* proteins present in the supernatants. It could however not be determined whether the MAC10 scFv had been expressed. Expression of the scFv was clearly not induced by the addition of IPTG. Induction also had no effect on the expression of the *Escherichia coli* proteins in the periplasmic extracts or the supernatants.

The SDS-PAGE analyses in figures 3.4 through 3. 6 clearly indicate that the MAC10 scFv is not being expressed at all in *Escherichia coli* HB2151 transformed with plasmid pAKJS. As stated earlier, expression of foreign genes cloned into pAK400 are under the transcriptional control of the LacZ promoter. Therefore the lack of expression of the MAC10 scFv must have resulted from repression of the *Lac*Z promoter. Prior to induction the *Escherichia coli* HB2151 transformed with plasmid pAKJS were grown in expression medium as well as 2 x TY-glucose. Expression of the MAC10 scFv did not occur in either case. The repression of the promoter can therefore only be attributed to the Lac repressor protein produced by the *Lac* I gene. The pAK plasmids all contain LacI genes to allow for *Escherichia coli* strain independent repression of the promoter (Krebber *et al*., 1997). This is in contrast to the pHEN-1 plasmid which contains no LacI gene (Hoogenboom *et al*., 1991). In addition the *Escherichia coli* HB2151 strain contains within its genome a functioning LacI gene (Hoogenboom *et al*., 1991). The LacI gene in *Escherichia coli* HB2151 is in fact a $LacI^{Q}$ mutant. This *Lac* I^{Q} mutation in the -35 hexamer of the promoter, results in a 100 copies of repressor protein being expressed per cell. The unmutated LacI promoter only produces around 10 to 20 copies of repressor protein per cell (Baneyx., 1999). It therefore means that when *Escherichia coli* HB2151 cells are transformed with pAKJS more repressor protein is produced than the IPTG can necessarily bind to. Subsequently repressor protein is still available in the cell to bind to the LacZ promoter and stop transcription of the scFv gene. The *Escherichia coli* JM83 strain on the other hand has no genomic Lac I gene therefore repressor protein is produced from the plasmid only (Fiedler *et al*., 2001).

Conclusion

The anti-mycolic acid clone 10 (MAC10) scFv protein has been expressed from the recombinant plasmid pMAC10 using the *Escherichia coli* TG-1 and HB2151 strains. Using *Escherichia coli* TG-1 strain transformed with pMAC10 it has been determined that 16 h of IPTG results in a large amount of protein in the bacterial periplasm. However, according to literature only 4 h are required to obtain sufficient expression of the scFv protein. The *Lac*Z promoter of the pMAC10 plasmid does not control expression of the scFv very tightly resulting in constitutive expression even without IPTG induction. The *Escherichia coli* HB2151 strain appears to express less of the scFv protein constitutively. The pAK plasmids were designed to contain their own *Lac* I genes to allow for strain independent suppression of scFv expression. Therefore a strain such as *Escherichia coli* JM83 will be required for effective expression using the pAKJS plasmid.

Chapter 4 : Random mutagenesis of an anti-mycolic acid single chain fragment variable

Introduction

The *Thermus aquaticus* (*Taq*) polymerase is naturally error prone due to the fact that it lacks 3' to 5' proof reading ability (Eckert *et al*., 1991). Error prone polymerase chain reaction (EP-PCR) is a modification of the standard reaction that seeks to exploit this characteristic. This reaction is a random mutagenesis procedure designed to introduce random point mutations into the sequence of an amplicon. The objective in designing the procedure was for a reaction that was capable of decreasing the fidelity of Taq without decreasing its amplification ability (Cadwell *et al*., 1992).

Error prone PCR is one of several *in vitro* mutagenesis methods that have been utilized to insert random mutations into gene sequences. The list of random mutagenesis methods includes the use of mutator strains of *Escherichia coli*, incorporation of nucleotide analogues, chemical mutagens, and the incorporation of randomized oligonucleotides. Chemical mutagens include hydroxylamine, sodium bisulphite, nitrous acid, formic acid and hydrazine are unable to generate all the base substitutions or produce highly mutated amplicons. Degenerate oligonucleotides on the other hand are not ideal since the mutations produced are not truly random. This is due to the fact that the space in which the mutation occurs is defined, each nucleotide is mutated with a given probability, and the base inserted at a position is selected (Fromant *et al*., 1995). No matter which type of *in vitro* mutagenesis method is used the design is to allow for the investigation of structurally and functionally important residues and positions. Random mutagenesis becomes especially useful when combined with a screening method when functionally and/or structurally important positions are not known (Cadwell *et al*., 1992).

In error prone PCR a mutagenic environment is established by using unequal concentrations of dNTPs. This is achieved by either a higher concentration of the pyrimidines compared to the purines or by having a higher concentration of a single base (Cadwell *et al*., 1992 and Fromant *et al*., 1995). In addition the concentration of $MgCl_2$ the co-factor of *Taq* is increased and $MnCl_2$ is added to the reaction as well. The increase in $MgCl_2$ is thought to stabilize non complimentary base pairs. Manganese ions on the other hand have a known mutagenic effect on a range of polymerases. This is by affecting nucleotide substrate specificity of the polymerase enzyme (Cadwell *et al*., 1992, Beckman *et al*., 1985, Tabor *et al*., 1989 and Cadwell *et al*., 1994).

The *Taq* polymerase enzyme does not have 3' to 5' exonuclease ability and therefore does not 'proofread' or check incorporated bases (Eckert *et al*., 1991). During EP-PCR the lack of proofreading ability by *Taq* is exploited by the misincorporation of nucleotides by the polymerase. Even under non mutagenic conditions the majority of changes in nucleotide sequence are due to errors made by the polymerase (Eckert *et al*., 1991). The efficiency with which the polymerase extends the primer chain beyond a mismatch is an important factor in determining PCR product yield. Whether a mismatch results in a mutated product is thus dependant on how efficiently the *Taq* polymerase can continue polymerization beyond it (Kwok *et al*., 1990).

There are 12 possible base pair mismatches. However mismatches containing the same nucleotide bases have been found to be symmetrical in terms of the effects on *Taq* polymerase efficiency during PCR. In other words a (primer: template) A: G mismatch had an equivalent effect to a G: A mismatch to chain elongation during PCR. There are thus only eight unique nucleotide mismatches possible (Kwok *et al*., 1990).

The mutation frequency from EP-PCR may be affected by the GC% of the target gene. There is some evidence that the mutation frequency is higher for AT base pairs than for GC base pairs. Therefore sequences with a higher number of GC base pairs could affect the average of mutations (Shafikhani *et al*., 1997).

During PCR under non mutagenic reaction conditions *Taq* was found to have a bias favouring certain types of substitutions. The AT to GC base pair transition was the most frequently observed substitution (Tindall *et al*., 1988 and Keohavong *et al*., 1989). Under EP-PCR conditions the bias favouring AT to GC base pair transition was also observed (Fromant *et al*., 1995, Vartanian *et al*., 1996, Lin-Goerke *et al*., 1997 and Vanhercke *et al*., 2004). Although initially EP-PCR was found to produce equal numbers of both transitions (Cadwell *et al*., 1992 and Cadwell *et al*., 1994). An ideal EP-PCR reaction is expected to produce twice as many transversions as transitions. This is because there are twice as many possible transversions as transitions (Wong et al., 2006, Cadwell *et al*., 1992 and Cadwell *et al*., 1994).

The error rate during a PCR varies according to the sequence of the amplified region as well as the *in vitro* conditions during which DNA synthesis occurs. The error rate as measured as number of mutations per nucleotide per cycle varies greatly between different observations. In fact the values observed by different investigators can differ by more than 10-fold (Eckert *et al*., 1991). The error rate is probably around one error per 9000 base pair substitutions per nucleotide synthesized or one mutation per 10 000 nucleotides per cycle (Tindall *et al*., 1988 and Eckert *et al*., 1991).

Literature differs on the appropriate error rate that should be attained in order to achieve the appropriate level of randomization and yet retain protein function. Some sources state that during a single EP-PCR reaction a mutation rate of 6 mutations per gene is sufficient. A higher rate would lead to an accumulation of genes with non functional protein products (Shafikhani *et al*., 1997). Others have shown that there is a disproportionally high number of functional as well as improved function mutants produced by high rates (Daugherty *et al*., 2000).

In this chapter error prone PCR was investigated to randomize the MAC10 scFv gene by directed evolution. This is to determine whether the method can be used to randomize the scFv gene sequence, but also to contribute to the identification of residues influencing antigen binding function.

Materials and methods

Flow diagram of Random mutagenesis by EP-PCR

The following is a summarized flow diagram of the methods used in the randomization by EP-PCR of the MAC10 scFv gene. At first the gene was amplified under EP-PCR conditions. Then the amplicon as well as plasmid pAK400 were digested with the restriction enzyme *Sfi* I. This was followed by ligation of the digested amplicon into the digested pAK400. Some of the recombinant plasmids obtained by ligation were used to transform electrocompetent *Escherichia coli* HB2151 by electroporation. A dilution series was made of the transformed cells and an aliquot of each concentration was plated on solid media and grown overnight.

EP-PCR of MAC10 → *Sfi* I restriction digestion of EP-PCR amplicon and plasmid pAK400

↓

Ligation of restriction digested EP-PCR amplicon into restriction digested plasmid pAK400

↓

Transformation by electroporation of *Escherichia coli* HB2151 ligation product

↓

Selection of colonies obtained by overnight incubation at 37°C on NE-agar plates

↓

Plasmid isolation from selected colonies → Sequencing of recombinant plasmids

Error prone polymerase chain reaction

The MAC10 scFv was randomly mutated using error prone PCR. Basic reaction set up was according to previously designed protocols (van Wyngaardt *et al.*, 2004). Adjustments were made to the basic reaction set up to establish a reaction environment that encourages mutagenic amplification Cadwell *et al*., 1992, Shafikhani *et al*., 1997 and Ausubel *et al*., 2002). The reactions consisted of PCR reaction buffer (20 mM Tris-HCl, 50 mM KCl, pH 8.5) (Jena Bioscience, Germany), 7 mM $MgCl_2$ (Jena Bioscience, Germany), dATP/dGTP (200 µM each) and dCTP/dTTP (1 mM each) (Jena Bioscience, Germany), 400 nM of each primer (Inqaba Biotech, South Africa), approximately 760 pg of template, 500 µM $MnCl_2$ (Sigma-Aldrich, USA), 5 U Taq DNA polymerase (5 U Taq, 400 µM Tris-HCl, 2 mM KCl, 2 µM EDTA, 20 µM DTT, 0.01% (v/v) Tween-20, 0.01% (v/v) Nonidet P-40, 1% (v/v) Glycerol, pH 8.0 (25°C)) (Jena Bioscience, Germany) in dddH_2O to a final volume of 50 µl. Thermal cycling was carried out as follows: one cycle of 1 min at 94°C after that 30 cycles of 94°C for 1 min and then 65°C for 1 min then 72°C for 2 min and a final extension cycle of 72°C for 5 min.

Table 4.1: Oligonucleotide primers used in error prone amplification of the mycolic acid clone 10 scFv gene from plasmid pAKJS as template.

Primer	Sequence	Restriction enzyme site	Tm (°C)[a]	Tm (°C)[b]
MB2	GTC CTC GCA ACT GC**G GCC CAG CCG GCC** CTG ATG GCG GCC GTA ACG	*Sfi* I	156	85
pAKreverse	TG**G GCC CCC GAG GCC** GCA TTG GGC TG	*Sfi* I	92	76

a) Tm = 4x (G+C) + 2x (A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

Polymerase chain reaction clean up and gel extraction

For PCR clean up, the specifications of the Macherey-Nagel NucleoSpin® Extract II kit were followed. Two volumes (200 µl) of buffer NT was added to one volume of sample (100 µl). This solution was then loaded onto a NucleoSpin® Extract II column which was centrifuged at 11 000 rcf for 1 min. The flow-through was discarded followed by a washing step with 700 µl buffer NT3 and centrifuged at 11 000 rcf for 1 min. The flow-through was again discarded. As a drying step the column was centrifuged for 2 min at 11 000 rcf to remove any traces of ethanol. In the elution step the bound DNA was eluted from the silica membrane by adding 30 µl of buffer NE and incubated at room temperature for 4 min, followed by centrifugation at 11 000 rcf for 1 min. A second elution step was then performed.

For extraction of DNA from an agarose gel using the NucleoSpin® Extract II kit, the desired band was excised from the gel using a clean scalpel blade. For each 100 mg of agarose gel 200 µl of NT buffer was added. The gel samples were incubated at 50°C in a water bath until the agarose had completely dissolved in the buffer, usually within 20 min, with infrequent vortexing. The sample was then loaded onto a NucleoSpin® Extract II column and centrifuged at 11 000 rcf for 1 min. The flow-through was discarded. During the washing step 700 µl of buffer NT3 was added and the column centrifuged for 1 min at 11 000 rcf. The eluate was discarded, after which the column was centrifuged again at 11 000 rcf for 1 min to remove residual ethanol. After a 2 min room temperature incubation, DNA bound to the silica membrane of the column was eluted with 60 µl of buffer NE and centrifugation at 11 000 rcf for 1 min.

Restriction enzyme digestion of plasmid pAK400 and PCR amplicon with *Sfi* I

To digest plasmid pAK400 or the PCR amplicon, 500 ng of the DNA was combined with 1 x buffer G (1 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 5 mM NaCl and 10 µg/ml BSA) (Fermentas), 10 U of enzyme (500 µM Tris-HCl (pH 7.4 at 25°C), 15 mM NaCl, 500 µM $MgCl_2$, 50 µM DTT, 0.0075% Triton X-100, 10 µg/ml BSA and 2.5% glycerol) (Fermentas) and $dddH_2O$ up to a volume of 30 µl. The reaction was then incubated at 50°C in a water bath for 3h.

Table 4.2: Restriction digest reaction of pAK400 with *Sfi* I.

Component	Amount
Template DNA	500 ng
Buffer G	3 µl
Sfi I restriction enzyme	10 Units
$dddH_2O$	Reaction volume up to 30 µl

Ligation of amplicon from pAK primers into pAK400 with T4 DNA ligase

Two types of ligation reactions were set up, the first according to literature (Krebber *et al*., 1997) and the second according to the manufacturers specifications. In the first method, plasmid pAK400 (100 ng) and insert (13 ng) were combined in 1x ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP) (Promega, USA) with 3 U of enzyme (500 µM Tris-HCl (pH 7.4), 2.5 mM KCl, 50 µM DTT, 5 µM EDTA and 2.5% glycerol) (Promega, USA) and $dddH_2O$ up to a volume of 20 µl at a ratio of 1:1.5. The amount of insert was calculated as follows:

Ng of insert = (ng of vector x kilobase size of insert)/(kilobase size of vector) x molar ratio of insert/vector

Thus: (100 ng x 0.76 kbp)/(3.836 kbp)x 1/1.5 = 13.2 ng of insert

Table 4.3: Ligation of pAK primer set amplicon into pAK400 at a molar ratio of 1: 1.5

Component	Amount
Vector	100 ng
Insert	13 ng
Ligase buffer	2 µl
T4 DNA ligase	3 U
$dddH_2O$	Reaction volume up to 20 µl

In the second method pAK400 (100 ng) and insert (58 ng) were combined in 1x ligase buffer (30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP) (Promega, USA) with 3 U of enzyme (500 µM Tris-HCl (pH 7.4), 2.5 mM KCl, 50 µM DTT, 5 µM EDTA and 2.5% glycerol) (Promega, USA) and $dddH_2O$ up to a volume of 20 µl at a ratio of 1:3. In both cases ligation of the *Sfi* I restriction digested PCR amplicon and plasmid pAK400 was left overnight at 4°C. the amount of insert was calculated using the same formula as before but with the vector: insert ratio of 1:3.

Table 4.4: Ligation of pAK primer set amplicon into pAK400 at a molar ratio of 3: 1

Component	Amount
Vector	100 ng
Insert	58 ng
Ligase buffer	2 µl
T4 DNA ligase	3 U
$dddH_2O$	Reaction volume up to 30 µl

Preparation of electrocompetent *Escherichia coli* HB2151

Electrocompetent *Escherichia coli* HB2151 were obtained by growing a 10 ml culture of LB medium in a shaking incubator at 200 rpm and 37°C. The culture was split into 2 x 5 ml aliquots and inoculated into 500 ml of prewarmed LB medium. The cultures were then grown in a shaking incubator at 200 rpm and 37°C till an OD_{600} of between 0.5 and 0.6 was reached. The cultures were transferred to prechilled centrifuge bottles and kept on ice for 20 min. After which the cells were pelleted by centrifugation at 2000 rcf, 4°C for 20 min. The pellets were kept on ice till a washing step in which the pellet was resuspended in 10 ml ice cold $dddH_2O$. To this suspension 240 ml more ice cold $dddH_2O$ was added and mixed. The cells were then pelleted at 2000 rcf, 4°C for 20 min and retained while the supernatant was discarded. This cell pellet was then resuspended in residual supernatant. The entire washing step was subsequently repeated. Then the cell pellet was resuspended in 10 ml of 10 % glycerol and incubated on ice for 60 min. The cells were again pelleted as before by centrifugation. Followed by resuspension in 500 µl of 10% glycerol and then stores at -70°C.

Transformation of *Escherichia coli* HB2151 by electroporation

For transformation 90 µl aliquots of electrocompetent cells and 10 µl of plasmid DNA were mixed and inserted into a pre-chilled electroporation cuvette. Electroporation was performed at 2000 V for 30 mS. To the cells in the cuvette 100 µl of prewarmed LB-glucose was added. The mixture was then added to 800 µl of LB-glucose and grown at 37°C for 1 h. A 1:10 and a 1:50 dilution were made of the mixture with LB-glucose. Aliquots of 100 µl of all three were plated on NE-agar plates and grown overnight at 37°C. From the colonies that subsequently grew 6 were picked for plasmid isolation and subsequent sequencing.

Nucleotide sequencing

Plasmid and PCR amplicon DNA needed to be sequenced to determine the order of nucleotides in and around the MAC10 gene. Both template types were sequenced using the BigDye 3.1 chemistry. Template (approximately 100 ng), primers (3.2 pmol), 5 x dilution buffer and BigDye 3.1 Terminator Ready Reaction Mix were combined in $dddH_2O$ to a final volume of 15 µl. The sequencing reaction was subjected to an initial denaturing cycle of 96°C for 1 min followed by 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s and elongation at 60°C for 4 min. The product of the sequencing reaction was precipitated by addition of cold 100 % ethanol (64 µl) to the sequencing reaction and 270 mM sodium acetate (pH 4.8) made up to a final volume of 22 µl with $dddH_2O$. This mixture was centrifuged at 11 000 rcf at 4°C for 30 min. The supernatant was discarded and the pellet washed twice with 60 µl of 70% ethanol and centrifugation at 11 000 rcf at 4°C for 15 min. The following primer was used for sequencing of the pAKJS clones.

Table 4.5: Oligonucleotide primer used in the sequencing of inserts obtained by error prone amplification of the mycolic acid clone 10 scFv gene.

Primer	Sequence	Tm (°C)[a]	Tm (°C)[b]
pAKseq F	GTG GAA TTG TGA GCG	46	48

a) Tm = 4x (G+C) + 2x(A+T)

b) Tm = 69.3 + 0.41(GC%) – 650/primer length

Sequence analysis

Sequencing results were obtained in BioEdit format as a chromatogram and corresponding sequence. A ClustalW multiple alignment was performed in BioEdit to compare the mutant sequences with that of the parental reference sequence. The multiple alignment along with the sequencing chromatograms were used to correct the sequences, eliminating gaps and sequencing artifacts. Corrected sequences were aligned with the parental reference sequence in the GeneDoc program to mark base substitution mutations. Complementarity determining regions (CDRs) were determined using the Kabat system as well as previously published possible locations in scFv's of the Nkuku phage display library.

Results

Error prone polymerase chain reaction

An error prone polymerase chain reaction was performed using a method based on previously described protocols (Cadwell *et al*., 1992, Shafikhani *et al*., 1997 and Ausubel *et al*., 2002). The purpose of the reaction was to cause base substitution mutations randomly spread over the length of the gene being amplified. A template free control PCR was performed along with the EP-PCR triplicate. As figure 4.1 indicates, the control reaction yielded no amplification as expected only primer dimmers. The EP-PCR triplicates all yielded a single band at the expected 750 bp mark.

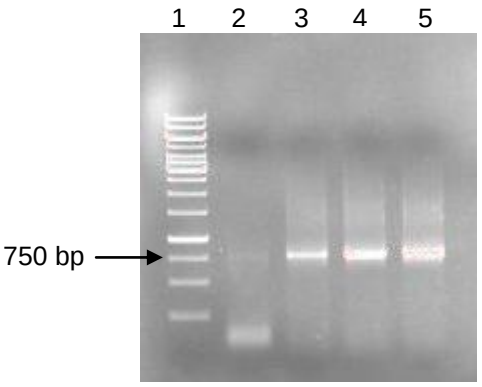


Figure 4.1: Amplicon of the mycolic acid clone 10 scFv gene obtained by error prone polymerase chain reaction with pAKreverse and MB2 primers from pAKJS as template. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = template free negative control, 3 to 5 = EP-PCR amplicons of the MAC10 scFv gene.

Restriction enzyme digestion of plasmid pAK400 with Sfi I

The plasmid pAK400 was chosen as cloning vector for the EP-PCR amplicon products. This is because of the ease with which the plasmid is restriction digested. Digestion of plasmid pAK400 requires only a single restriction enzyme *Sfi* I. Between the two GGCC repeats of the *Sfi* I restriction site there are 5 non specific nucleotides. The non specific nucleotides of the two *Sfi* I sites are completely different in plasmid pAK400. Directional cloning is therefore possible since an insert would be required to have two distinct non complementary ends. Between the two *Sfi* I restriction sites is a 2101 bp tetracycline resistance (tet) cassette which due to its size allows for easy separation of double digested plasmid from single cut or uncut. The plasmid pAK400 was digested with *Sfi* I to remove the tet cassette to allow for the ligation of the MAC10 scFv gene. The tet cassette band (2101 bp) was expected close to the 2000 bp fragment of the ladder. In figure 4.2, the digested pAK400 lanes had two bands one was close to 6000 bp and the other close to 4000 bp. The latter is expected to be the double digested (3836 bp) pAK400. The other band is probably single digested pAK400.

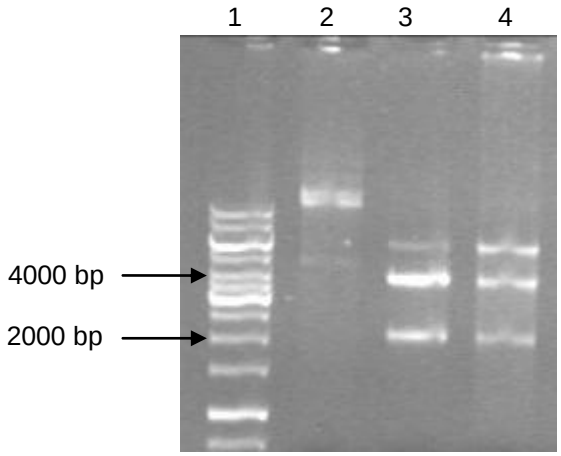


Figure 4.2: Agarose gel electrophoresis (1%) of *Sfi* I restriction enzyme digestion of plasmid pAK400. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 = undigested plasmid pAK400, 3 and 4 = *Sfi* I restriction digested plasmid pAK400.

Ligation of error prone polymerase chain reaction amplicons into pAK400

The amplicons obtained by EP-PCR of the MAC10 scFv gene were ligated into plasmid pAK400 and then used to transform electrocompetent *Escherichia coli* HB2151. Six colonies were randomly selected. As indicated in figure 4.3, the plasmid DNA purified from 5 of the 6 colonies showed similar banding patterns. The agarose gel electrophoresis of the plasmids suggests that the insert of the sixth plasmid is different from the others.

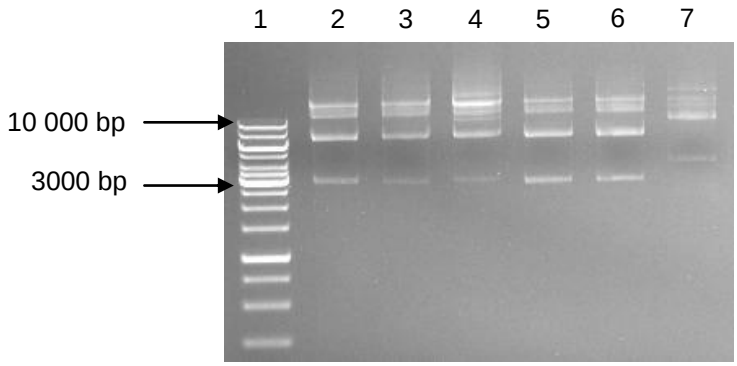


Figure 4.3: Agarose gel electrophoresis of recombinant plasmids obtained by ligation of error prone polymerase chain reaction amplicons to *Sfi* I restriction digested plasmid pAK400. 1 = O'GeneRuler™ 1 kb DNA molecular mass ladder, 2 to 7 = pAKJS clones obtained from overnight cultures of *Escherichia coli* HB2151 in NE medium.

Sequence analysis of error prone polymerase chain reaction amplicons

The amplicons obtained by EP-PCR of the MAC10 scFv gene were sequenced from the recombinant plasmids. The sequence of the insert from pAKJS6 showed no identity to the parental sequence and BLAST analysis indicated no *Gallus gallus* immunoglobulin gene was present. The sequence of pAKJS6 was therefore omitted from further analysis. The number of transition and transversion mutations as well as the total for each mutant was calculated. The total number of mutations were as follows, pAKJS1 = 8; pAKJS2 = 11; pAKJS3 = 14; pAKJS4 = 11; pAKJS5 = 7 (table 4.6). The mutation rate of the EP-PCR reaction was determined by dividing the total number of mutations from 5 mutants by the total number of base pairs sequenced from 5 mutants multiplied by 1000. The formula used was (Total # of mutations from all mutants/Total # of base pairs sequenced) x 1000 = mutation rate. Thus 51 mutations/5893 base pairs x 1000 = 8.6 mutations/1000 base pairs. The expected mutation rate was 11 mutations/1000 base pairs. Tables 4.7 to 4.11 list the type of base pair, codon and amino acid change that occurred in each of the pAKJS mutants. The base pairs are numbered according to their positions in the chromatograms obtained from the sequencing. For all the mutants except pAKJS5 the majority of base pair substitutions were transversions. The changes in codons were classified as missense if they resulted in an amino acid change, silent if no change in amino acid or nonsense if the mutation resulted in a stop codon. All mutant sequences contained at least one silent mutation and three out of the five mutants contained one nonsense mutation. Missense mutations resulted in changes in side group structure and polarity at the residue level.

Table 4.6: Frequency of different transition and transversion mutations in the JS single chain fragment variable mutants

Mutation	pAKJS1	pAKJS2	pAKJS3	pAKJS4	pAKJS5
Purine to purine	1	2	2	2	3
Pyrimidine to pyrimidine	1	2	2	3	3
Purine to pyrimidine	3	5	8	2	1
Pyrimidine to purine	3	2	2	4	0

Table 4.7: Amino acid substitutions in the JS1 single chain fragment variable gene

Position	Type of point mutation	Parental codon	Mutant codon	Type of mutation	Parental amino acid	Mutant amino acid
259	Transversion	TGG	AGG	Missense	Tryptophan	Arginine
322	Transition	GCC	ACC	Missense	Alanine	Threonine
467	Transversion	GTC	TTC	Missense	Valine	Phenylalanine
533	Transition	CCG	TCG	Missense	Proline	Serine
565	Transversion	ATC	ATA	Silent	Isoleucine	Isoleucine
736	Transversion	ACT	TCT	Missense	Threonine	Serine
747	Transversion	CAA	CAT	Missense	Glutamine	Histidine
769	Transversion	TTC	ATC	Missense	Phenylalanine	Isoleucine

Table 4.8: Amino acid substitutions in the JS2 single chain fragment variable gene

Position	Type of point mutation	Parental codon	Mutant codon	Type of mutation	Parental amino acid	Mutant amino acid
127	Transversion	TTG	TAG	Nonsense	Leucine	Stop
132	Transition	GAG	AAG	Missense	Glutamic acid	Lysine
253	Transversion	GAA	GTA	Missense	Glutamic acid	Valine
343	Transversion	CAG	CTG	Missense	Glutamine	Leucine
379	Transition	GCT	GTT	Missense	Alanine	Valine
477	Transversion	GGA	GGT	Silent	Glycine	Glycine
563	Transversion	GTC	GAC	Missense	Valine	Aspartic acid
598	Transversion	AAC	TAC	Missense	Asparagine	Tyrosine
730	Transition	ATC	GTC	Missense	Isoleucine	Valine
749	Transversion	GAG	GTG	Missense	Glutamic acid	Valine
766	Transition	TTC	CTC	Missense	Phenylalanine	Leucine

Table 4.9: Amino acid substitutions in the JS3 single chain fragment variable gene

Position	Type of point mutation	Parental codon	Mutant codon	Type of mutation	Parental amino acid	Mutant amino acid
137	Transition	CCC	CCT	Silent	Proline	Proline
143	Transversion	GGA	GGC	Silent	Glycine	Glycine
201	Transition	TGG	TAG	Nonsense	Tryptophan	Stop
232	Transversion	GAA	GAT	Missense	Glutamic acid	Aspartic acid
360	Transversion	GAG	GTG	Missense	Glutamic acid	Valine
413	Transversion	CGA	CGT	Silent	Arginine	Arginine
564	Transversion	GAC	TAC	Missense	Aspartic acid	Tyrosine
621	Transversion	ACT	CCT	Missense	Tyrosine	Proline
636	Transition	AAC	GAC	Missense	Asparagine	Aspartic acid
678	Transversion	TCC	GCC	Missense	Serine	Alanine
684	Transition	ACA	ATA	Missense	Threonine	Isoleucine
755	Transversion	AGT	AGA	Missense	Serine	Arginine
758	Transversion	ACA	ACT	Silent	Threonine	Threonine
763	Transversion	AGC	ATC	Missense	Serine	Arginine

Table 4.10: Amino acid substitutions in the JS4 single chain fragment variable gene

Position	Type of point mutation	Parental codon	Mutant codon	Type of mutation	Parental amino acid	Mutant amino acid
146	Transition	GGC	GGT	Silent	Glycine	Glycine
180	Transversion	TGC	GGC	Missense	Cysteine	Glycine
214	Transversion	AGC	ATC	Missense	Serine	Isoleucine
243	Transition	AAA	GAA	Missense	Lysine	Glutamic acid
294	Transversion	TAC	AAC	Missense	Tyrosine	Asparagine
336	Transversion	AAC	TAC	Missense	Asparagine	Tyrosine
389	Transition	ACC	ACT	Silent	Threonine	Threonine
547	Transition	AAC	GAC	Missense	Asparagine	Aspartic acid
585	Transversion	AGT	AGA	Missense	Serine	Arginine
612	Transversion	TAT	TAG	Nonsense	Tyrosine	Stop
692	Transition	TTC	TCC	Missense	Phenylalanine	Serine

Table 4.11: Amino acid substitutions in the JS5 single chain fragment variable gene

Position	Type of point mutation	Parental codon	Mutant codon	Type of mutation	Parental amino acid	Mutant amino acid
218	Transition	ATG	GTG	Missense	Methionine	Valine
244	Transition	GGC	GGT	Silent	Glycine	Glycine
384	Transversion	GAG	GTG	Missense	Glutamic acid	Valine
391	Transition	ACC	ACT	Silent	Threonine	Threonine
521	Transition	GCG	GCA	Silent	Alanine	Alanine
539	Transition	TCG	TCA	Silent	Serine	Serine
781	Transition	GCG	GTG	Missense	Alanine	Valine

The alignment of the mutant sequences with the unmutated parental reference sequence is indicated in figure 4.4. Mutations are not limited to specific regions of the sequence but are spread along the length of the scFv gene. Mutagenesis therefore occurred in a random way as expected. Figure 4.5 A) to D) indicates segments from the sequencing chromatograms of the mutants as they were obtained from the sequencing facility at the University of Pretoria after samples were run on an ABI 3130 sequencer.

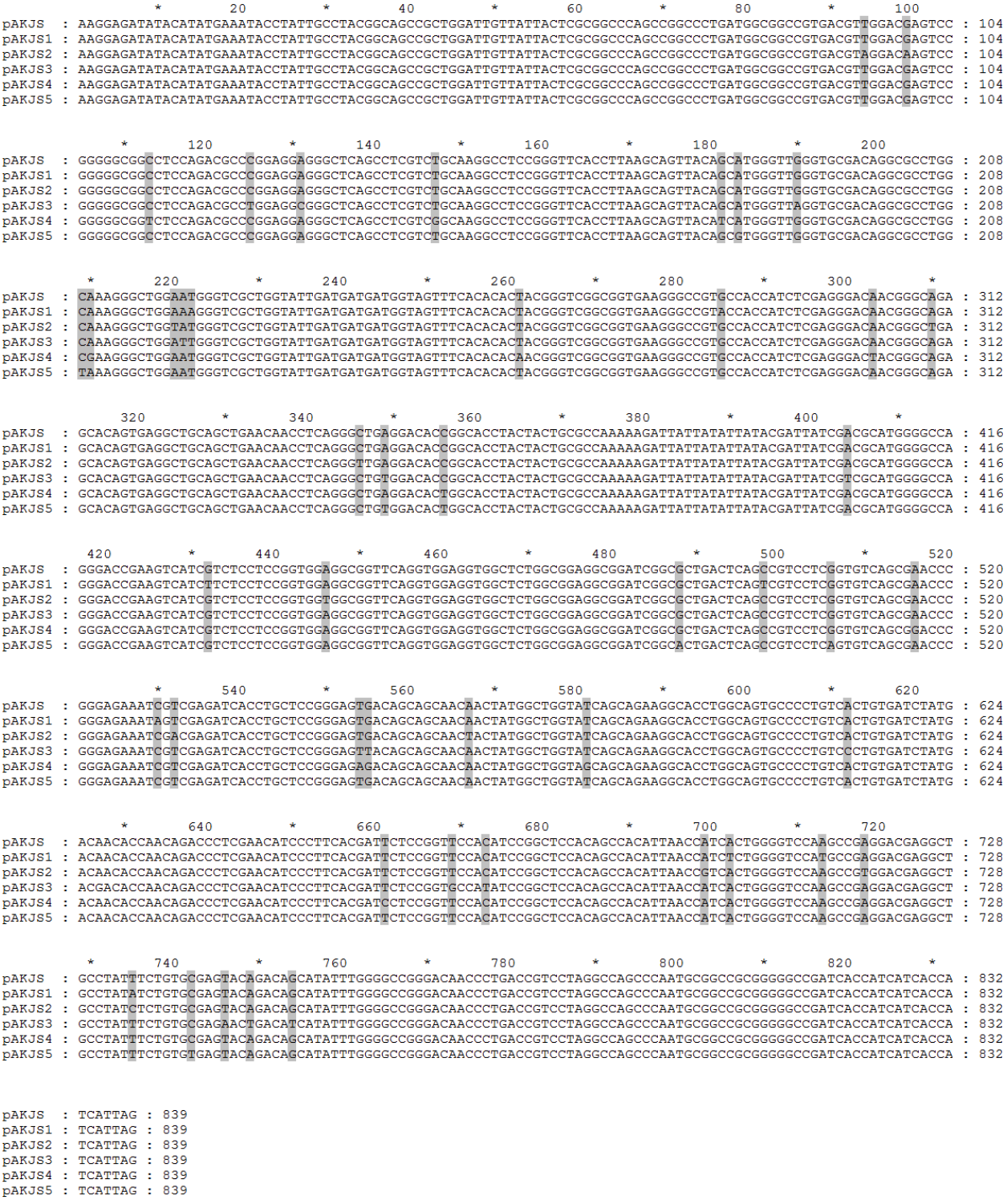
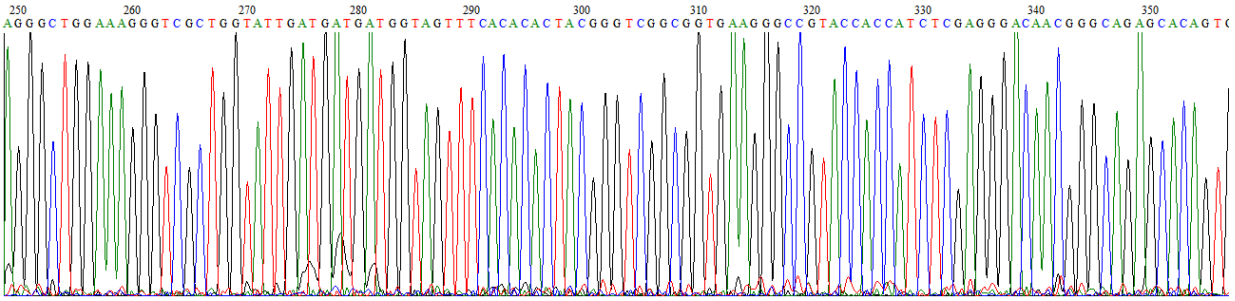
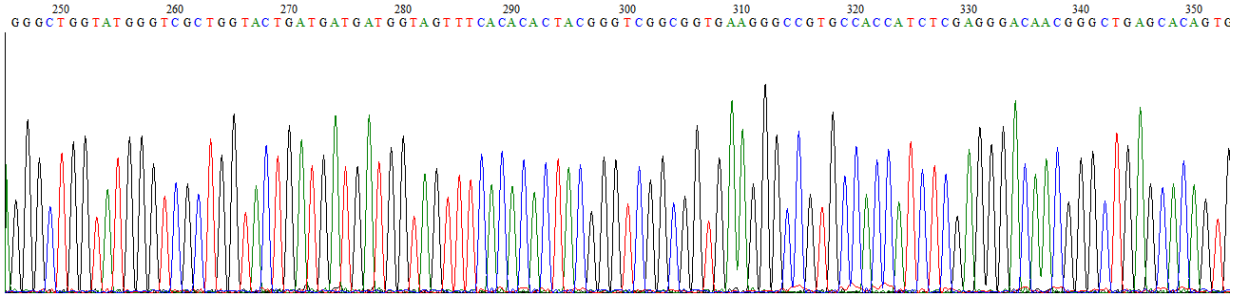


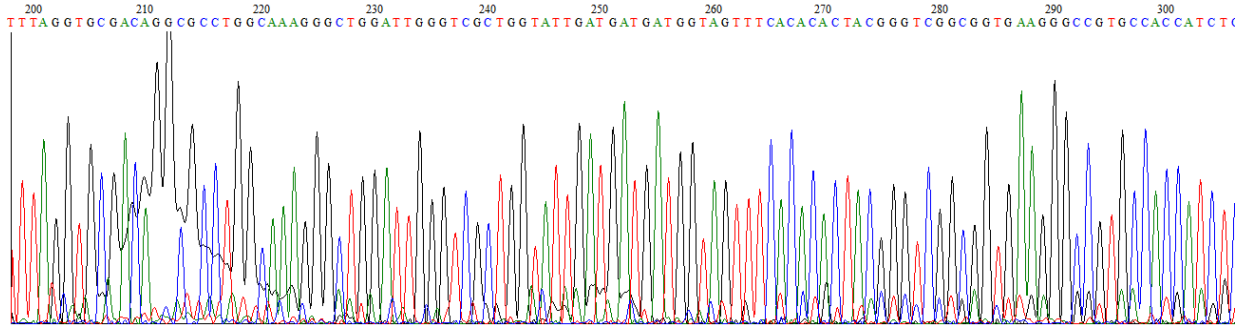
Figure 4.4: Alignment of DNA sequences from unmutated and mutated single chain fragment variable genes ligated into pAK400. Sequences were aligned using the GeneDoc Multiple sequence alignment editor & shading utility program. The location of mutations is indicated by gray shading of all sequences at the particular aligned position.



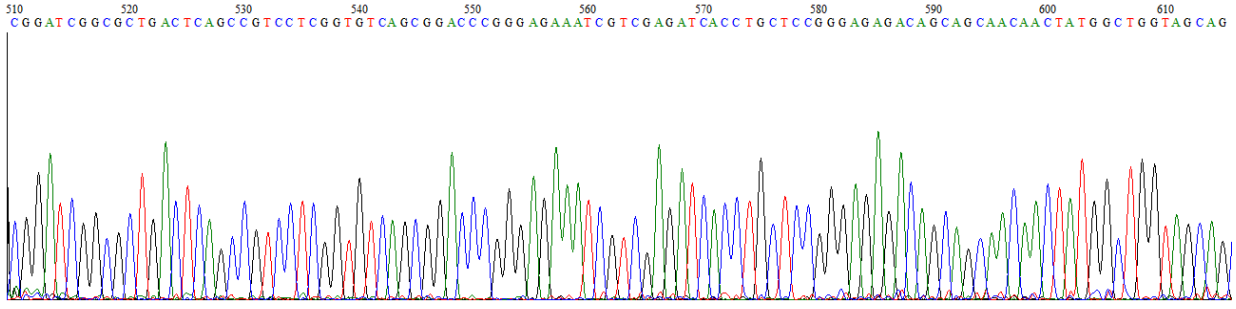
A)



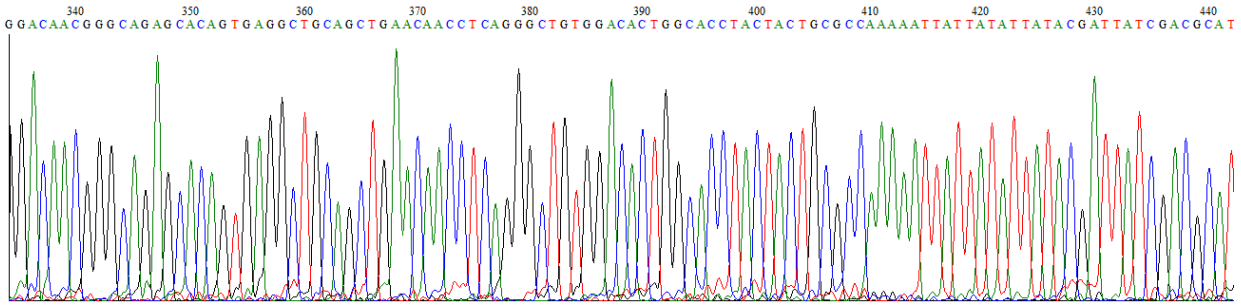
B)



C)



D)



E)

Figure 4.5: Sequencing chromatograms of single chain fragment variable mutants obtained by error prone polymerase chain reaction of the MAC10 gene. A) pAKJS1, B) pAKJS2, C) pAKJS3, D) pAKJS4, E) pAKJS5. Sequencing chromatograms were obtained as ABI files in the BioEdit sequencing alignment editor.

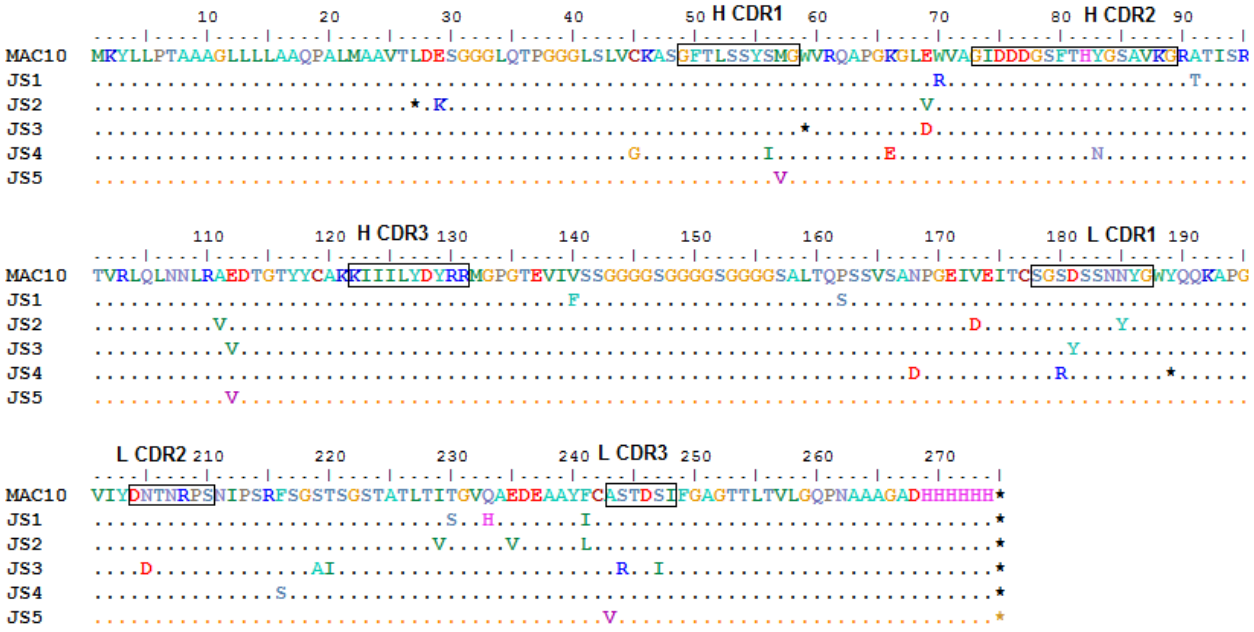


Figure 4.6: Alignment of JS protein mutant sequences to parental MAC10. Complementarity determining regions (CDRs) indicated by black squares. H = heavy chain, L = light chain, asterisk = stop codon positions. Sequences include pelB leader and polyhistidine tag. Protein sequences were aligned in the BioEdit sequencing alignment editor. Dots indicate identity to the parental reference amino acid sequence.

Discussion

A gene coding for an anti-mycolic acid single chain variable fragment was subjected to random mutagenesis by error prone polymerase chain reaction. Five randomly selected mutants were obtained and sequenced.

The changes in codon sequence that occurred due to EP-PCR in the scFv gene (figure 4.4) were classified as missense, silent or nonsense (tables 4.7 to 4.11). It is at this point unknown what the impacts are of the amino acid substitutions (figure 4.6) that occurred in each of the sequences. Therefore no matter how drastic the change in side group structure or polarity was, the effects on function were not determined. The amino acid changes that occurred in the CDR regions will however be discussed later. Also amino acid sequence changes that will probably have deleterious effects will be discussed. Substitution mutations are thus only classified as missense mutations. Of the five mutants three contained stop codons resulting in nonsense mutants. It had been expected that the majority of mutant sequences would contain stop codons. What was not expected was that the number was limited to one nonsense mutation per scFv mutant sequence.

The mutation rate of 8.6 mutations per 1000 base pairs that was obtained is in close agreement with the desired level of 11 mutations per 1000 base pairs (Ausubel *et al*., 2002). However at this point it must be stressed that the five mutants selected do not reflect a statistical sampling of the entire population of mutated amplicons generated by error prone PCR. The calculated mutation rate is therefore expected to deviate from the true rate. Another factor which may influence the mutation rate is the concentration of template used for the reaction. In order for the mutation rate to be of the desired value the template has to be diluted to 5 pM (760 pg) (Ausubel *et al*., 2002 and Shafikhani *et al*., 1997). This is a very low concentration and difficult to obtain and measure. The amount of template present in the reaction may therefore be slightly higher than desired leading to the mutation rate being lowered.

Dilution of the template to 10 pM causes 1024 fold amplification during PCR which is the equivalent to 10 effective cycles of DNA synthesis (Shafikhani *et al*., 1997). Therefore a template starting concentration half of that, 5 pM would result in approximately 20 cycles of DNA synthesis. According to literature 20 cycles of DNA synthesis, or doublings, would result in approximately 11 base substitution mutations in an 800 bp amplicon (Ausubel *et al*., 2002).

The ratio of total transversions to transitions obtained was 1:0.7 (tables 4.7 to 4.11) which is in agreement with the literature ratio of 1:0.75 (Cadwell *et al*., 1992). Reaction conditions and the sequence being amplified have an influence on this ratio (Cadwell *et al*., 1992, Fromant *et al*., 1995). In addition it was shown that Mn^{2+} ions may enhance the proportion of transversion mutations (Vartanian *et al*., 1996). The lower purine concentration relative to pyrimidine concentration was intended to force all the transversions to be significantly more favourable over all the transitions (Cadwell *et al*., 1992). However not all the transversion causing mismatches were equally likely to occur due to their different effects on *Taq* polymerases chain elongation efficiency. The G: G, G:A, A:A and C:C mismatches significantly decrease the enzymes extension efficiency (Kwok *et al*., 1990 and Eckert *et al*., 1991). Thus it is likely that the ratio of transitions to transversions will remain almost 0.5:1 but not reach this ideal. This is unless further modifications that can overcome the negative effects of these mismatches on the *Taq* polymerase are made. These would include an increased polymerase concentration and a longer extension times (Cadwell *et al*., 1992).

A ratio of 1:0.5 was previously calculated for the total number of mutations at A-T base pairs versus total mutations at G-C base pairs (Shafikhani *et al*., 1997). The ratio of total mutations at A-T base pairs versus total mutations at G-C base pairs was calculated as 1:0.6 for the pAKJS mutants. The high GC% of the MAC10 scFv gene (59.64%) resulted in only a slight increase in mutations at G-C base pairs.

Amongst the transversion mutations the majority were T to A and A to T substitutions. This trend was previously observed (Cadwell *et al*., 1994, Fromant *et al*., 1995, Casson *et al*., 1995, Vartanian *et al*., 1996, Shafikhani *et al*., 1997, Drummond *et al*., 2005 and Vanhercke *et al*., 2005). The frequency of these substitutions has been attributed to a bias in the *Taq* enzyme favouring mutations at A-T over G-C base pairs (Shafikhani *et al*., 1997). This biased trend is probably due to the extension efficiency of *Taq* after T:T mismatches (Vartanian *et al*., 1996 and Kwok *et al*., 1990). The reverse mismatch A:A could not have contributed greatly to these transversions due to the reasonable decrease it causes in *Taq* polymerases chain extension efficiency (Kwok *et al*., 1990). Changes in dNTP concentration during EP-PCR could decrease the dominance of T to A and A to T substitutions amongst transversions (Cadwell *et al*., 1994, Fromant *et al*., 1995 and Vartanian *et al*., 1996). No G-C to C-G substitutions or the reverse was observed in any of the pAKJS mutants. This can be explained by the observed difficulty in chain extension after C-C and G-G mismatches (Kwok *et al*., 1990, Eckert *et al*., 1991 and Fromant *et al*., 1995). The effect of the G-G mismatch on chain extension efficiency was however measured at equimolar dNTP concentration and not at low purine concentrations (Kwok *et al*., 1990).

Of the transitions that occurred in the pAKJS mutants most were from G-C to A-T base pairs. In total twice as many G-C to A-T transitions occurred as the reverse. These are the result of T:G and A:C mismatches which are believed to have little effect on strand extension efficiency (Kwok *et al*., 1990). An ideal EP-PCR should have both transitions occurring at a ratio of 1 (Wong et al., 2006). A ratio of 1 for both transitions has already previously been observed using this method (Cadwell *et al*., 1992). An EP-PCR scFv library generated previously showed a similar dominance of G-C to A-T base pair substitutions. An identically generated library however showed a dominance of the opposite transition (Drummond *et al*., 2005). The deviation from the ideal could simply be due to the high GC% of the MAC10 scFv gene (Shafikhani *et al*., 1997 and Fromant *et al*., 1995). Transitions make up a smaller number of the total mutations and in the majority of individual mutants. Thus the effect of the high GC% may be more noticeable amongst the transition mutations.

The prevalence of a certain transition could be determined by reaction conditions and sequence context (Cadwell *et al*., 1994). Previously the reverse transition A-T to G-C was observed to be dominant in EP-PCR (Fromant *et al*., 1995, Vartanian *et al*., 1996, Lin-Goerke *et al*., 1997 and Vanhercke *et al*., 2005). Such was the prevalence of this transition that it was concluded that *Taq* polymerase has a bias favouring A-T to G-C transitions (Tindall *et al*., 1988, Keohavong *et al*., 1989 and Vanhercke *et al*., 2005). Although the findings all pointed towards this bias, all reactions differed from this EP-PCR and each other. Specifically in terms of template, dNTP concentration, primer concentration and PCR profile. Also this trend was originally observed under non EP-PCR conditions (Tindall *et al*., 1988 and Keohavong *et al*., 1989). However the result of EP-PCR of a V_H gene showed a much smaller bias favouring A-T to G-C transitions (Casson *et al*., 1995). The ratio was also shown to be sensitive to dGTP and dTTP concentrations (Cadwell *et al*., 1992, Cadwell *et al*., 1994 and Fromant *et al*., 1995). And as stated in the previous paragraph two libraries, generated using the same template sequence and reaction conditions had different results in terms of which transition was in the majority (Drummond *et al*., 2005). Lastly the prevalence of G-C to A-T transitions may be due to the small sample size. For pAKJS4 there were almost equal numbers of both transitions, but G-C to A-T transitions were slightly in the majority. Thus the trend in total number of transition could be due to the trend in the majority of colonies that were picked.

The low number of C to T mutations suggests that deamination of cytosine to uracil was not very prevalent during this reaction. Cytosine deamination could occur during high temperatures such as the denaturation step of PCR (Eckert *et al*., 1991). As expected mutations at the first and second nucleotide of a codon was more likely to result in missense mutations. Substitution of the last nucleotide of a codon was more likely to result in a silent mutation. This is due to the degeneracy of the genetic code (Voet *et al*., 2004). The efficiency with which *Taq* polymerase extends beyond a mismatch is also inhibited if there is an internal primer: template close by downstream to the terminal (Kwok *et al*., 1990). This probably explains the absence of mutations next to one another in the sequence.

Even though there were biases favouring certain nucleotide mutations, significant changes occurred at the amino acid level in each of the pAKJS mutants (figure 4.6). These mutations are significant at the amino acid level (primary structure) due to the differences in side chain structure and/or polarity between the parental and the mutants. The complementarity determining regions of the parental MAC10 sequence (figure 4.6) were determined using the Kabat system as well as previous literature on the Nkuku library and chicken derived scFv's (Kabat *et al*., 1977, Kabat *et al*., 1991, van Wyngaardt *et al*., 2004 and Andris-Widhopf *et al*., 2000). Only JS4 and 5 had mutations in their (heavy chain) H CDR1. In JS4 a Serine was replaced by an Isoleucine and in JS5 a Methionine by a Valine. Although both resulted in noticeable changes in side chain structure only the Serine to Isoleucine conversion was in terms of polarity as well. One mutant, JS4, contained a change in H CDR2 from Tyrosine to Asparagine. This appears to be more a change in side chain structure than polarity. Of the five JS mutants none contained changes in H CDR3. Three of the mutants, JS2, 3 and 4 contained changes in the sequence of L CDR1. An Asparagine to Tyrosine, an Aspartic acid to Tyrosine and a Serine to Arginine change were obtained for the mutants respectively. The latter two changes represent changes in charge as well as side chain structure. One of the mutants, JS3 contained an Asparagine to Aspartic acid change in the L CDR2 sequence. This represents more of a change in side chain charge rather than structure since the two amino acid side chains are structurally very similar. Two of the mutants, JS3 and 5 contained changes in the L CDR 3 sequence. The former however contains two sequence changes, a Serine to Arginine and a Serine to Isoleucine. Both changes represent changes in side chain structure and charge. The latter, JS5 contains only an Alanine to Valine change that is only a side chain structural change.

Subsequently it is now clear that most of the mutations occurred outside the CDRs and that the substitution amino acids generally have similar properties to the parental. In the pAKJS mutants the substitute amino acid may be related to the original in terms of structure or charge. This tendency to retain amino acid properties has been referred to as 'structural fidelity'. It is believed to be an evolutionary adaptation to protect against mutations and subsequent changes in protein structure and function (Goldberg *et al*., 1966, Wong *et al*., 2006 and Zhu *et al*., 2003). It has been stated that the ultimate goal of a mutagenesis method should be to replace any amino acid with the 19 others in a statistical manner without impacting expression (Wong *et al*., 2006). Although this may seem ideal in terms of mutagenesis it may not be for the retention of function. If there is a tendency to replace amino acids with others that are related to the original in terms of charge or structure, there is a greater chance of affecting function without losing it entirely. Therefore this statement is not entirely true when considering that often random mutagenesis is used to obtain improved function mutants and not just mutants in general. If it is diversification that is desired perhaps multiple rounds of EP-PCR with different mutation rates should be considered (Shafikhani *et al*., 1997). Much higher mutation rates have also been shown to be beneficial in the discovery of higher affinity scFv's from random libraries (Daugherty *et al*., 2000 and Drummond *et al*., 2005). It is also quite likely that the effects of amino acid changes on the activity of a protein can be additive, synergistic or dependant on one another in some way (Shafikhani *et al*., 1997). Thus the relation of a replacement amino acid in a mutant protein to the one in the original may not be a limiting factor if improved functional proteins are desired.

The location of many mutations outside of the predicted CDR loop regions is a consequence of the randomness of EP-PCR and was expected. Although the residues of the CDR binding loops are responsible for the majority of antigen interaction, non contact residues may also affect binding. Experimental evidence indicates that mutation of second sphere (non contact) residues could also affect interaction with antigen (Dubreuil *et al*., 2005 and Zahnd *et al*., 2004). The positive effects that mutations have on affinity may be attributed to changes in interaction between structural residues. This in turn causes subtle changes in stability and folding of the protein, resulting in improved antigen interaction (Zahnd *et al*., 2004).

Even if an EP-PCR experiment does not lead to an improved binding antibody it can identify important residues. This information can then be used for the targeted mutation of these residues and subsequently obtain improved function mutants (Saviranta *et al*., 1998).

Despite the fact that no functional assay was performed to determine the effects of the amino acid substitutions on binding for the scFv some estimation can be made. The nonsense mutations will result in the expression of an incomplete truncated protein. This will most likely have an effect on the binding ability of the scFv protein. Another example is the loss of a Cysteine residue in pAKJS4. The result of which will be the inability of the protein to be correctly folded. Consequently the scFv protein will be unable to bind to its antigen. The prevalence of mutations that have a clear negative effect on function such as the nonsense mutation and the loss of a cysteine residue is an indicator that some selection should be applied to the EP-PCR amplicon pool.

Conclusion

An error prone polymerase chain reaction method was developed to investigate its viability as a method of randomizing the MAC scFv gene. From the 5 randomly selected mutant colonies selected it was determined that the reaction has an approximate mutation rate of 8.6 mutations per 1000 bp. This mutation rate compares with published values of 11 mutations per 1000 bp. Transversion mutations appeared most frequently in the total number of mutations, occurring mostly at A-T base pairs. Transition substitution mutations occurred mostly at G-C base pairs. Patterns of mutations (biases) during EP-PCR and PCR in general were previously attributed to the *Taq* polymerase enzyme and reaction conditions. In order to overcome these biases changes in nucleotide concentration should be investigated, most likely dGTP. A longer extension time during amplification should also be investigated as a way of overcoming the detrimental effects certain base pair mismatches have on the polymerase chain extension efficiency. The majority of the selected mutants contained mutations that would result in the expressed protein being non functional. It is therefore clear that a method of selection such as a display technology should be applied to the amplicon pool in order to eliminate loss of function mutants. Although a random mutagenesis procedure may not result in improved function mutants it can be used as a method to identify important residues and positions within the coding sequence. In order to better understand the trends of substitution mutations that occurred due to random mutagenesis further experiments should include a larger number of mutated sequences.

Concluding discussion

Mycolic acids are large, hydrophobic yet antigenic lipids from the cell walls of *Mycobacterium tuberculosis*. From the Nkuku phage display library 11 scFv's were obtained that bind to mycolic acids. One of these was selected to undergo random mutagenesis to obtain more information on the positions and identities of amino acid residues involved in function. Also to investigate random mutagenesis as a method of generating a 'library' of randomized sequences from which improved function variants could be isolated in the future.

In chapter 2 a nested PCR method was investigated as a means to overcome non specific annealing of PCR primers during amplification. The nested PCR method uses primers and a reaction set up developed previously. Since the 5'and 3' ends of all scFv genes in the Nkuku library are identical the nested PCR approach is applicable to the entire library. Nested PCR is also the preferred method of amplification during EP-PCR according to literature (Cadwell *et al*., 1994). If plasmid pGE20 is still considered as a possible subcloning plasmid for expression the *Xho* I restriction site inside the MAC10 gene should be eliminated in some way, such as site directed mutagenesis. With a single base pair substitution a silent mutation can eliminate the restriction site. The codon in question would still code for the correct amino acid. Once this has been accomplished the MAC10 gene can then be cloned into pGE20 between the *Xho* I and *Spe* I restriction sites. It should then be investigated whether the B10 tag present in pGE20 should be replaced by a His-tag or not. If a His-tag is to replace the B10 tag the oligonucleotide coding for the His-tag should not consist only of repetitions of the CAT codon. Every second or third codon should be a CAC codon. This non repetitive sequence of the oligonucleotide could decrease the potential for incorrect annealing of the single stranded oligonucleotides.

The gene for the MAC10 scFv has been successfully cloned into the plasmid pAK400 resulting in the recombinant plasmid pAKJS. Fast and simplified cloning of the desired PCR amplicon as well as the presence of all the needed tag sequences make this the preferred plasmid of the two that were investigated. The plasmid pAK400 is derivative of the plasmid pAK100 (Krebber *et al*., 1997). These plasmids share their simplified cloning strategy. This is important for future investigation of EP-PCR libraries generated as pAK100 is a phage display plasmid.

In chapter 3 the expression of the scFv from two plasmids and two strains of *Escherichia coli* were investigated. It was clearly evident that that long periods of induction are not required when using plasmid pMAC10 and *Escherichia coli* TG-1. However due to the constitutive expression in the absence of IPTG induction this particular system may prove undesirable for expression of functional scFv's. The glucose suppression which is supposed to prevent constitutive expression by plasmid pHEN-1 derived recombinant plasmids appears not to be reliable. This observation is clear from the constitutive expression of the MAC10 scFv by both *Escherichia coli* TG-1 and HB2151. In addition to having a simple cloning procedure the plasmid pAK400 has a much more reliable mechanism of suppressing expression prior to induction compared to pHEN-1. However since the plasmid pAK400 produces its own Lac repressor protein expression would require an *Escherichia coli* strain that does not produce repressor protein. The *Escherichia coli* JM83 strain would be the best to investigate since it has already been shown to work well with plasmid pAK400 recombinants for expression (Krebber *et al*., 1997).

The reaction set up for the PCR amplification of Nkuku library scFv genes proved easily adaptable to EP-PCR conditions. Based on the recommendations of several publications the previously designed reaction and amplification profile were adapted to a mutagenic environment (Cadwell *et al*., 1992; Shafikhani *et al*., 1996 and Ausubel *et al*., 2002). Future investigation should start by looking at the dependence of the mutation rate on the starting amount of template. The initial EP-PCR method called for a starting template concentration of 200 pM which was then amplified approximately 1000 times after 30 cycles (Cadwell *et al*., 1992 and Cadwell *et al*., 1994). Later protocols called for a lower template concentration to obtain adequate mutation rates (Shafikhani *et al*., 1996 and Ausubel *et al*., 2002). The pAKJS mutants obtained from 5 pM of starting template which is half the concentration specified by the former of the mentioned protocols (Shafikhani *et al*., 1996).

A solution to obtaining the concentrations required for the reaction would be a master mix set up. In a master mix a larger volume of the reaction is assembled which is aliquoted to obtain multiple reactions of the required volume. However this does not guarantee that the exact amount of a reagent such as the template is present in the reaction. Although it was not a problem in the current reaction it was previously advised to use a nested PCR approach rather than amplify from a plasmid template (Cadwell *et al*., 1994). This is a further reason to investigate the template dependence of the mutation rate as the amount of PCR amplicon template would be even less than for plasmid.

Mutation rate and diversification are important factors to consider in mutation experiments which aim towards improved function. As previously stated there is evidence that higher mutation rates might actually lead to an increase in improved function mutants present in a library (Daugherty *et al*., 2000 and Drummond *et al*., 2005). Others have stressed that a mutation rate of less than 10 mutations per 1000 base pairs is the maximum allowed before all mutants become non functional. Also that higher mutation rates should rather be achieved by repeated rounds of amplification. Whether the mutations are achieved by repeated rounds of EP-PCR or a single high error rate amplification reaction, selection should be applied to the resulting library. Selection by phage display should be a starting point, since the pAK100 phage display plasmid is already available and shares a cloning strategy with pAK400.

The biases towards mutations at certain base pairs can most likely be overcome by adjusting the concentrations of dATP and dGTP in the reaction. Previous recommendations have been to double the concentration of dGTP (Cadwell *et al*., 1994). However the *Taq* polymerase has been shown to be sensitive to the concentrations of all nucleotides (Cadwell *et al*., 1992, Cadwell *et al*., 1994 and Fromant *et al*., 1995). Thus adjusting the concentrations of both of the purine nucleotides either separately or together may lead to the balance in diversification required. Whether a particular mutation is retained in a complete amplicon product is also determined by how efficiently *Taq* polymerase is able to elongate beyond the causative mismatch (Kwok *et al*., 1990 and Eckert *et al*., 1991). Thus a longer extension time may result in an increase in mutations that result from efficiency decreasing mismatches. The number of mutants sampled after EP-PCR experiment is important in order to determine the trends within the mutations obtained. Most recently it was argued that the number of unique mutants within an error prone PCR generated library could range in the 10's of 1000's. Thus in order to obtain the desired mutations or mutant a sample larger than five sequences should be investigated (Daugherty *et al*., 2000 and Drummond *et al*., 2005).

Appendix

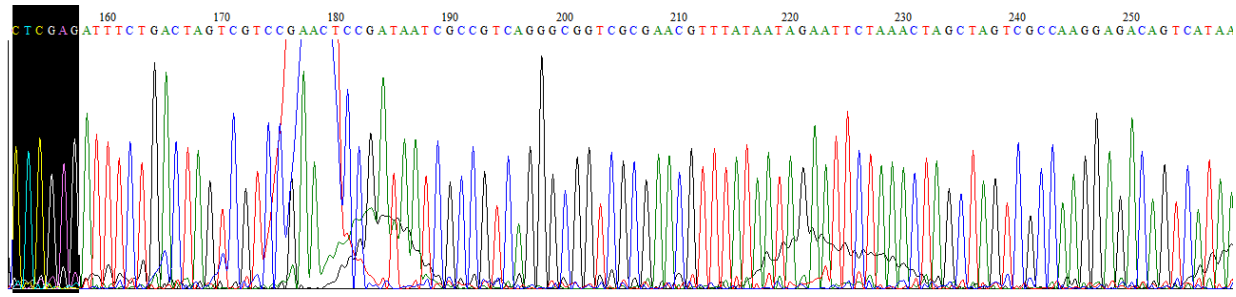


Figure A1: Partial sequencing chromatogram of plasmid pGE20 cloning site obtained with M13R as forward primer. Highlighted in black is the *Xho* I restriction site downstream is the *Spe* I restriction site.

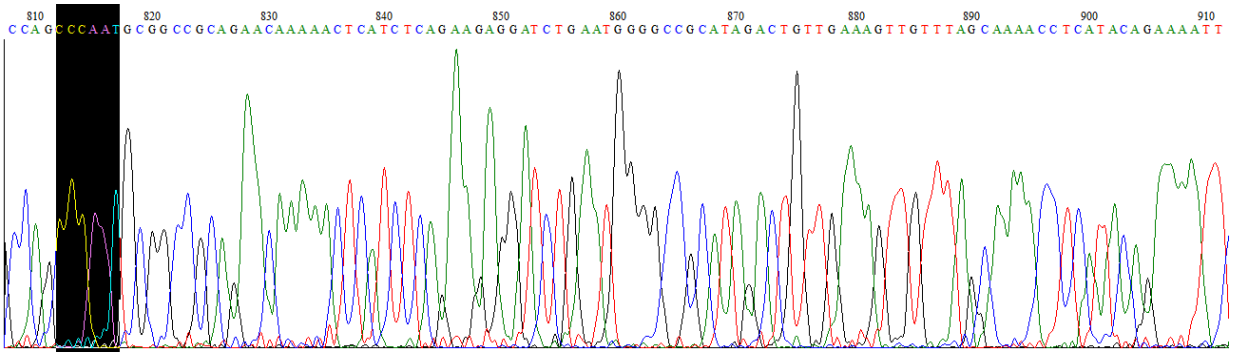
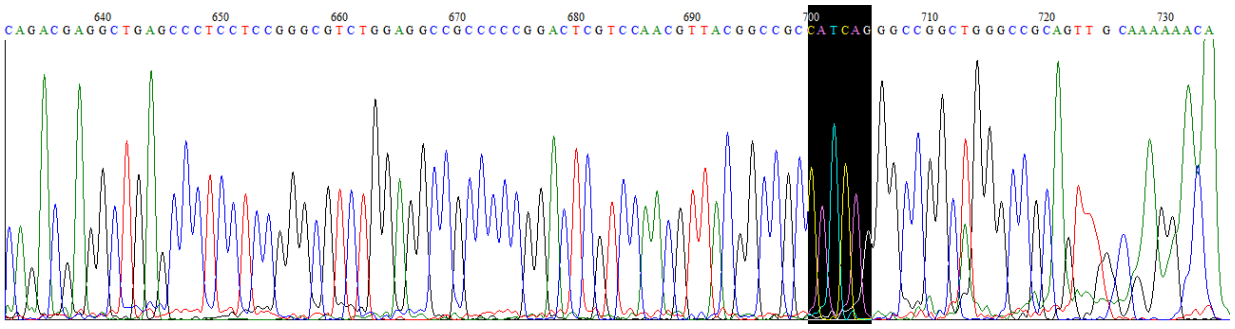
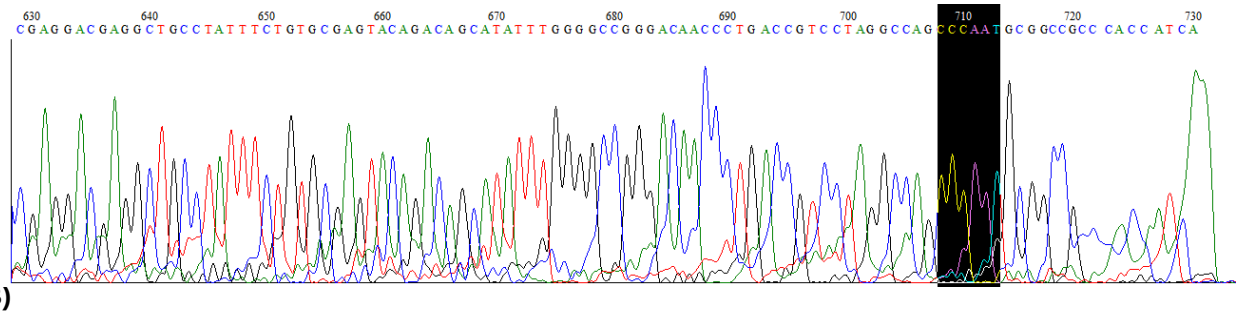


Figure A2: Partial sequencing chromatogram of pscFJS1.1 obtained with M13R as forward primer. Highlighted in black are the last two codons of the MAC10 scFv gene, downstream of this is the *Not* I restriction site and *c-myc* tag sequence.



A)



B)

Figure A3: Partial sequencing chromatograms of MB primer set PCR amplicon obtained with MB primers as reverse and forward primers. A) Partial sequencing chromatogram obtained with MB1 as reverse primer. Highlighted in black is the reverse compliment of the first two codons of the MAC10 scFv gene, downstream is the *Sfi* I restriction site. B) Partial sequencing chromatogram obtained with MB2 as forward primer. Highlighted in black are the last two codons of the MAC10 scFv gene, downstream is the *Not* I restriction site.

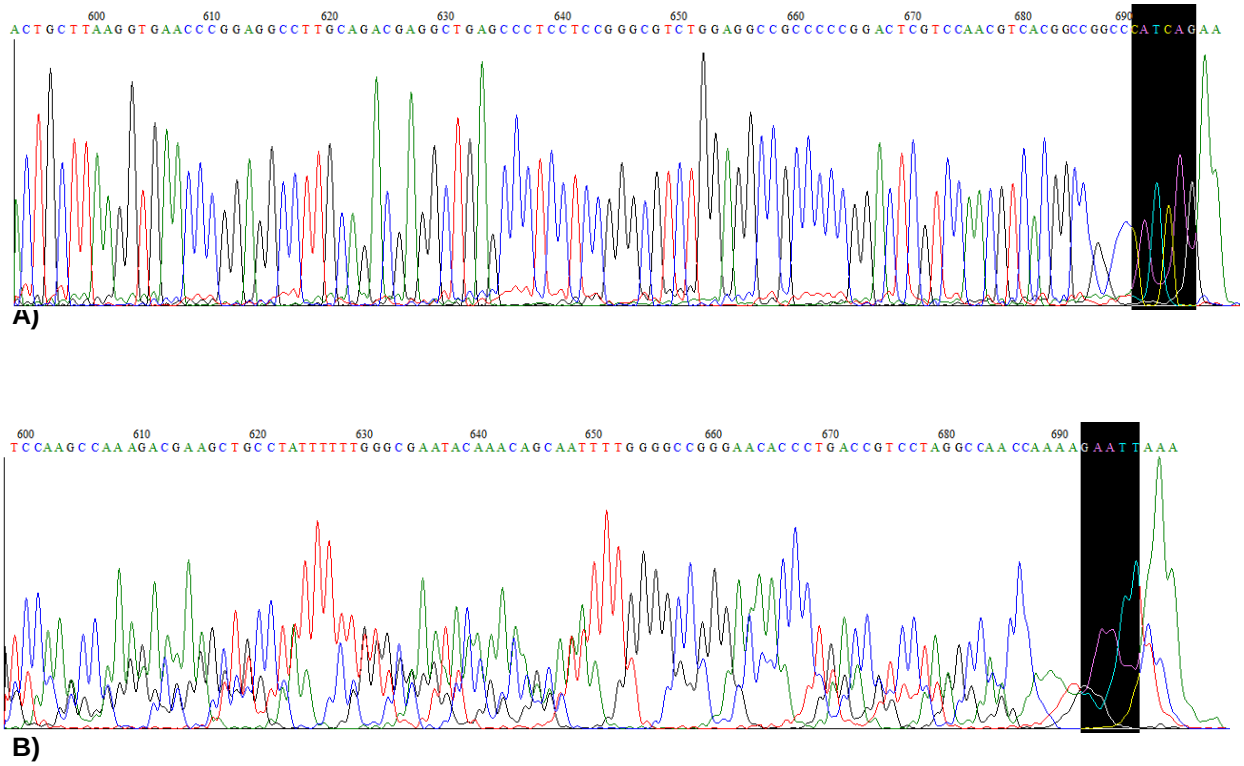
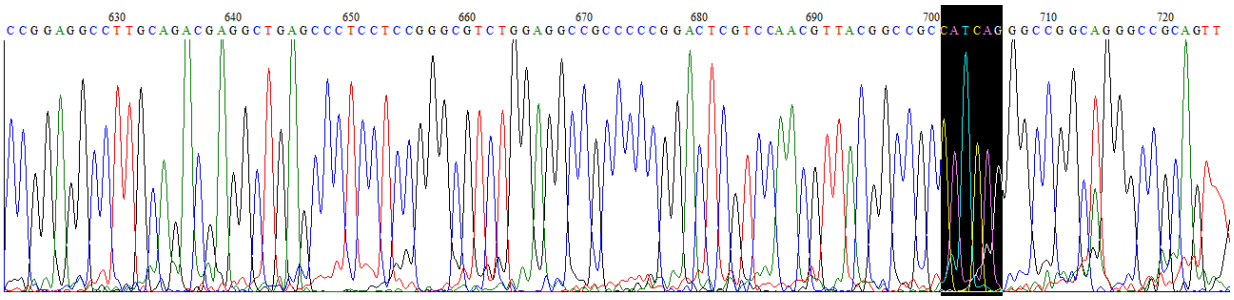
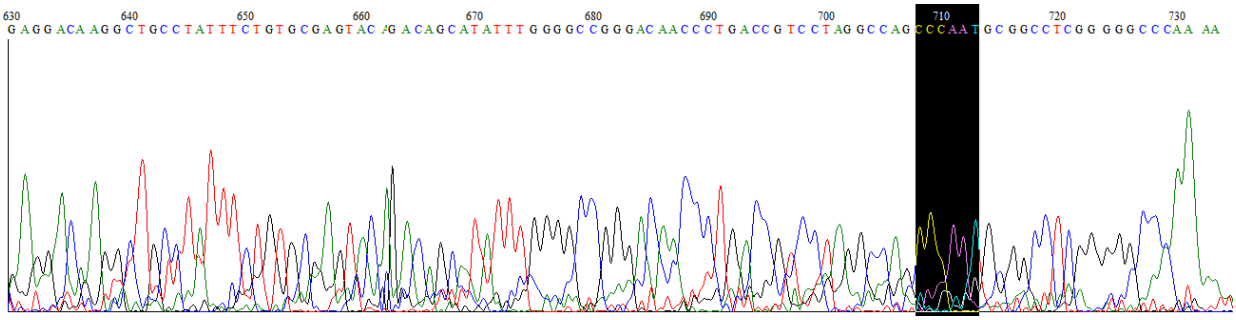


Figure A4: Partial sequencing chromatograms of Henuni primer set PCR amplicon obtained with Henuni primers as reverse and forward primers. A) Partial sequencing chromatogram obtained with Henuni R1 as reverse primer. Highlighted in black is the reverse compliment of the first two codons of the MAC10 scFv gene, downstream are adenine nucleotides added by amplification. B) Partial sequencing chromatogram obtained with Henuni F1 as forward primer. Highlighted in black are last two codons of MAC10 scFv gene, downstream are adenine nucleotides added by amplification.

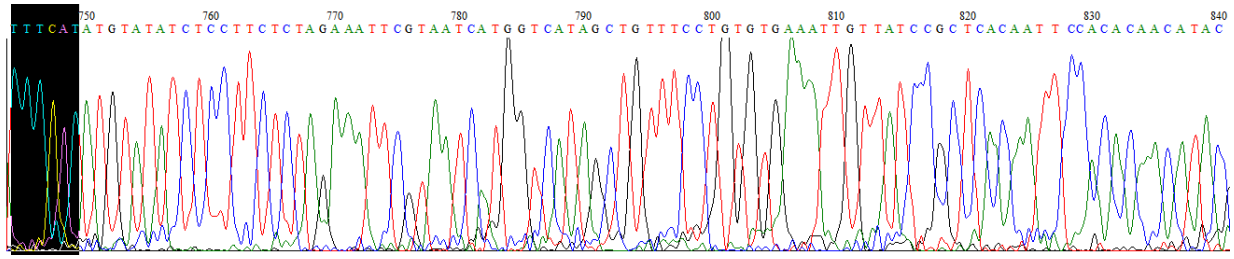


A)

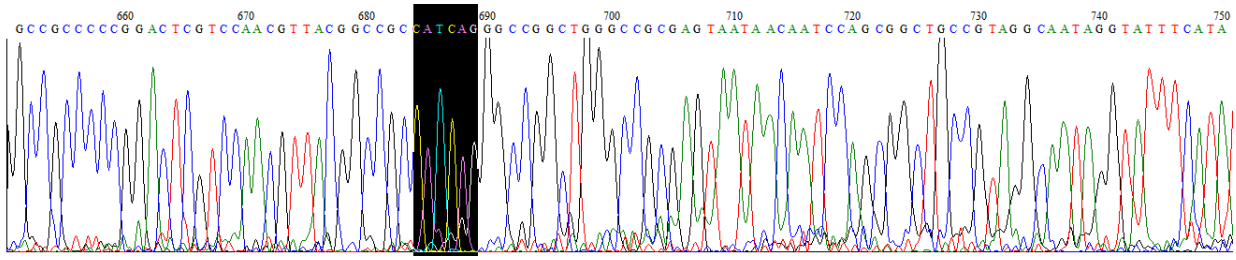


B)

Figure A5: Partial sequencing chromatograms of pAK primer set PCR amplicon obtained with pAK primers as reverse and forward. A) Partial sequencing chromatogram obtained with pAKreverse as reverse primer. Highlighted in black is the reverse compliment of the first two codons of the MAC10 scFv gene, downstream is the sequence of the MB2 primer containing a *Sfi* I restriction site. B) Partial sequencing chromatogram obtained with MB2 as forward primer. Highlighted in black are last two codons of MAC10 scFv gene, downstream is the pAKreverse primer sequence containing a *Sfi* I restriction site.



A)



B)

Figure A6: Partial sequencing chromatograms of recombinant plasmid pAKJS obtained with pAKreverse as reverse sequencing primer. A) Partial sequencing chromatogram obtained with pAKreverse as sequencing primer. Highlighted in black are the first two codons of the *pelB* leader sequence. B) Partial sequencing chromatogram obtained with pAKreverse as sequencing primer. Highlighted in black are the first two codons of the MAC10 scFv gene.

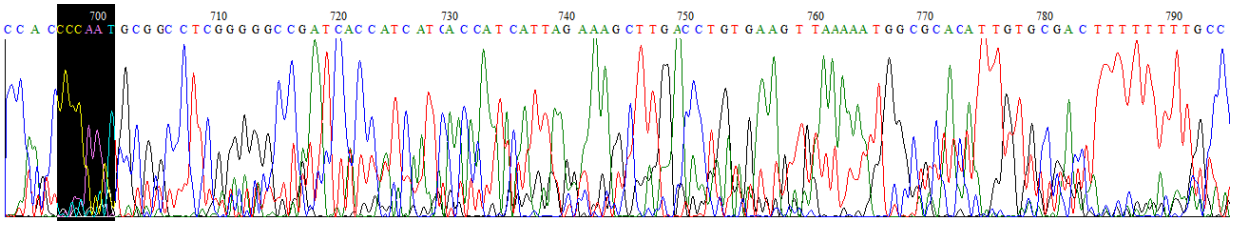


Figure A7: Partial sequencing chromatogram of recombinant plasmid pAKJS obtained with MB2 as forward sequencing primer. Indicated in black are the last two codons of the MAC10 scFv gene, downstream of this is the *Sfi* I restriction site where ligation occurred as well as the polyhistidine tag of pAK400.

References

Adams, GP; Shaller, CC; Dadachova, E; Simmons, HH; Horak, EM; Tesfaye, A; Klein-Szanto, AJP; Marks, JD; Brechbiel, MW and Weiner, LM; *A single treatment of yttrium-90-labeled CHX-A"–C6.5 diabody inhibits the growth of established human tumor xenografts in immunodeficient mice*; Cancer research; 2004; 64; 6200-6206

Andris-Widhopf, J; Rader, C; Steinberger, P; Fuller, R and Barbas III, CF; *Methods for the generation of chicken monoclonal antibody fragments by phage display*; Journal of immunological methods; 242; 2000; p 159-181

Arnau, J; Lauritzen, C; Petersen, GE and Pederson, J; *Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins*; Protein Expression and Purification; 48; 2006; p 1-13

Arntz, Y; Seelig, JD; Lang, HP; Zhang, J; Hunziker, P; Ramseyer, JP; Meyer, E; Hegner, M and Gerber, C; *Label-free protein assay based on a nanomechanical cantilever array*; Nanotechnology; 2003; 14; p 86-90

Ausubel, FM; Brent, R; Kingston, RE; Moore, DD; Seidman, JG; Smith, JA and Struhl, K; *Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology: 1. Introduction of plasmid DNA into cells: basic protocol 1: transformation using $CaCl_2$*; 5; 1; 2002; p 1.29

Ausubel, FM; Brent, R; Kingston, RE; Moore, DD; Seidman, JG; Smith, JA and Struhl, K; *Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology: 8. Mutagenesis of cloned DNA: 8.3 Random mutagenesis by PCR basic protocol: Mutation of DNA sequences*;5; 1; 2002; p 8.7-8.9

Backmann, N; Zahnd, C; Huber, F; Bietsch, A; Plueckthun, A; Lang, HP; Guentherodt, HJ; Hegner, M and Gerber, C; *A label-free immunosensor array using single-chain antibody fragments*; Proceedings of the national academy of sciences; 102; 41; p 14587-14592

Baneyx, F; *Recombinant protein expression in Escherichia coli*; Current opinion in Biotechnology; 1999; 10; p 411-421

Barderas, R; Shochat, S; Martinez-Torrecuadrada, J; Altschuh, D; Meloen, R and Casal, J I; *A fast mutagenesis procedure to recover soluble and functional scFv's containing amber stop codons from synthetic and semisynthetic antibody libraries*; Journal of immunological methods; 312; 2006; p 182–189

Beckman, RA; Mildvan, AS and Loeb, LA; *On the Fidelity of DNA Replication: Manganese Mutagenesis in Vitro*; Biochemistry; 24; 1985; p 5810-5817

Bershtein, S and Tawfik, DS; *Advances in laboratory evolution of enzymes*; Current opinion in chemical biology; 2008; 12; p 151-158

Bej, AK; Mahbubani, MH and Atlas; *Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications*; Critical reviews in biochemistry and molecular biology; 26; 3; 1991; p 301-334

Beukes, M; Lemmer, Y; Deysel, M; Al Dulayymi, JR; Baird, MS; Koza, G; Iglesias, MM; Rowles, RR; Theunissen, C; Grooten, J; Toschi, G; Roberts, VV; Pilcher, L; Van Wyngaardt, S; Mathebula, N; Balogun, M; Stoltz , AC; Verschoor, JA; *Structure–function relationships of the antigenicity of mycolic acids in tuberculosis patients*; Chemistry and physics of lipids; 2010; 163; p 800-808

Bird, RE; Hardman, KD; Jacobson, JW; Johnson, S; Kaufmann, BM; Lee, SM; Lee, T; Pope, SH; Riordan, GS and Whitlow; *Single-chain antigen-binding proteins*; Science, 1988; 242; 4877; p 423-426

Borrebaeck, CAK; *Antibodies in diagnostics – from immunoassays to protein chips*; Immunology today; 2000; 21; 8; p 379-382

Brannigan, JA and Wilkinson, AJ; *Protein engineering 20 years on*; Nature; 2002; 3; p 964-970

Cadwell, RC and Joyce, GF; *Randomization of genes by PCR mutagenesis*; Genome research; 2; 1992; p 28-33

Cadwell, RC and Joyce, GF; *Mutagenic PCR*; Genome research; 3; 1994; p s136-s140

Casson, L P and Manser, T; *Evaluation of loss and change of specificity resulting from random mutagenesis of an antibody V_H region*; The journal of immunology; 155; 1995; p 5647-5654

Chasteen, L; Ayriss, J; Pavlik, P and Bradbury, ARM; *Eliminating helper phage from phage display*; Nucleic acids research; 2006; 34; 21

Conroy, PJ; Hearty, S; Leonard, P and O'Kennedy, RJ; *Antibody production, design and use for biosensor-based applications*; Seminars in cell and developmental biology; 2009; 20; p 10-26

Daugherty, P S; Chen, G; Iverson, B L and Georgiou, G; *Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies*; Proceedings of the national academy of sciences; 97; 5; 2000; p 2029-2034

De Bellis, D and Schwartz, I; *Regulated expression of foreign genes fused to lac: control by glucose levels in growth medium*; Nucleic acids research; 18, 5; 1990; p 1311

Depetris, M; Casalis, P; Kratje, R; Etcheverrigaray, M and Oggero, M; *A scFv antibody fragment as a therapeutic candidate to neutralize a broad diversity of human IFN-alpha subtypes*, Journal of Immunological methods; 334; 2008; p 104-113

de Marco, A; *Biotechnological applications of recombinant single-domain antibody fragments*; Microbial cell factories; 2011; 10; 44; p 1-14

Dillon, PP; Manning, BM; Daly, SJ; Killard, AJ and O'Kennedy, R; *Production of a recombinant anti-morphine-3-glucuronide single-chain variable fragment (scFv) antibody for the development of a "real-time" biosensor-based immunoassay*; Journal of immunological methods; 2003; 276; p 151-161

Drummond, D A; Iverson, B L; Georgiou, G and Arnold, F; *Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins*; Journal of molecular biology; 350; 2005; p806-816

Dubreuil, O; Bossus, M; Graille, M; Bilous, M; Savatier, A; Jolivet, M; Menez, A; Stura, E and Ducancel, F; *Fine tuning of the specificity of an anti-progesterone antibody by first and second sphere residue engineering*; The journal of biological chemistry; 280; 26; 2005; p 24880-24887

Dunne, L; Daly, S; Baxter, A; Haughey, S and O'Kennedy, R; Surface plasmon resonance-based immunoassay for the detection of aflatoxin B1 using single-chain antibody fragments; Spectroscopy letters; 2005; 38; p 229-245

Dutton, CM; Paynton, C and Sommer, SS; *General method for amplifying regions of very high G + C content*; Nucleic Acids Research; 21, 12, 1993; p 2953–2954

Eckert, K A and Kunkel, T A; *DNA polymerase fidelity and the polymerase chain reaction*; Genome research; 1; 1991; p 17-24

Eggertsson, G and Soell, D; *Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli*; Microbiological reviews; 52, 3; 1988; p 354 – 374

Encarnacou, JM; Rosa, L; Rodrigues, R; Pedro, L; da Silva, FA; Goncales, J and Ferreira, GNM; *Piezoelectric biosensors for biorecognition analysis: Application to the kinetic study of HIV-1 Vif protein binding to recombinant antibodies*; Journal of biotechnology; 2007; 132; p 142-148

Fromant, M; Blanquet, S and Plateau, P; *Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction*; Analytical biochemistry; 224; 1995; p 347-353

Fujiwara, N; Pan, J; Enomoto, K; Terano, Y; Honda, T and Yano, I*; Production and partial characterization of anti-cord factor (trehalose-6,6P-dimycolate) IgG antibody in rabbits recognizing mycolic acid subclasses of Mycobacterium tuberculosis or Mycobacterium avium*; Federation of European microbiological societies: Immunology and medical microbiology; 1999; 24; p 141-149

Gavilondo, JV and Larrick, JW; *Antibody Engineering at the Millennium*; Biotechniques; 2000; 29; p 128-145

Giovane, C; Travé, G; Briones, A; Lutz, Y; Wasylyk, B and Weiss, E; *Targeting of the N-terminal domain of the human papillomavirus type 16 E6 oncoprotein with monomeric ScFv's blocks the E6-mediated degradation of cellular p53*; Journal of molecular recognition; 12; 1999; p 141 – 152

Goldberg, A L and Wittes, R E; *Genetic code: aspects of organization;* Science; 153; 1966; p 420-424

Greenman, J; Jones, E; Wright, MD and Barclay, AN; *The use of intracellular single-chain antibody fragments to inhibit specifically the expression of cell surface molecules*; Journal of immunological methods; 1996; 194; p 169-180

Grennan, K; Strachan, G; Porter, AJ; Killard, AJ and Smyth, MR; Atrazine analysis using an amperometric immunosensor based on single-chain antibody fragments and regeneration-free multi-calibrant measurement; Analytica chimica acta; 2003; 500; p 287-298

Hanes, J and Pluekthun, A; *In vitro selection and evolution of functional proteins by using ribosome display*; Proceedings of the national academy of sciences; 1997; 94; p 4937-4942

Haqqi, T M; Sarkar, G; David, CS and Sommer, SS; *Specific amplification with PCR of a refractory segment of genomic DNA*; Nucleic Acids Research; 16, 24; 1988; p 11844

Holliger, P and Hudson, PJ; *Engineered antibody fragments and the rise of single domains*; Nature biotechnology; 2005; 23; 9; p 1126-1136

Horacek, J; Garrett, SD; Skladal and Morgan, RA; *Characterization of the interactions between immobilized parathion and the corresponding recombinant scFv antibody using a piezoelectric biosensor*; Food and agricultural immunology; 1998; 10; p 363-374

Horton, J and Rattcliffe, NA; M; Roitt, I (Ed); Brostoff, J (Ed) and Male, D (Ed); *Evolution of immunity in Immunology*; 2001; 6; p 222

Hoogenboom, H R; Griffiths, A D; Johnson, K S; Chiswell, D J; Hudson, P and Winter; *Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains*; Nucleic Acids Research; 19, 15; 1991 p 4133 – 4137

Hu, WG; Thompson, HG; Alvi, AZ; Nagata, LP; Suresh, MR and Fulton, RE; *Development of immunofiltration assay by light addressable potentiometric sensor with genetically biotinylated recombinant antibody for rapid identification of Venezuelan equine encephalitis virus*; Journal of immunological methods; 2004; 289; p 27-35

Hudson, PJ and Souriau, C; *Engineered antibodies*; Nature medicine; 2003; 9; 1; p 129-134

Huse, WD; Sastry, L; Iverson, SA; Kang, AS; Alting-Mees, M; Burton, DR; Benkovic, SJ and Lerner, RA; *Generation of a large combinatorial library of immunoglobulin repertoire in phage lambda*; Science; 246; 1989; p 1275-1281

Huston, JS; Levinson, D; Mudgett-hunter, M; Tai, MS; Novotny, J; Margolies, MN; Ridge, RJ; Bruccoleri, RE; Haber, E; Crea, R and Oppermann, H; *Protein engineering of antibody binding sites: Recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli*; Proceedings of the national academy of sciences; 1988; 85; p 5879-5883

Iwahashi, M; Milenic, D; Padlan, EA; Bei, R; Schlom, J; Kashmiri, SVS; *CDR substitutions of a humanized monoclonal antibody (CC49): contributions of individual CDRs to antigen binding and immunogenicity*; Molecular immunology; 1999; 36; p 1079-1091

Iyer, G; Michalat, X; Chang, YP; Pinaud, FF; Matyas, SE; Payne, G; and Weiss, S; High affinity scFv-hapten pair as a tool for quantum dot labelling and tracking of single proteins in live cells; Nano letters; 2008; 8; 12; p 4618-4623

Kabat, EA; Wu, TT and Bilofsky, H; *Unusual distributions of amino acids in complementarity determining (Hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining Sites*; The journal of biological chemistry; 252; 19; 1977; p 6609-6616

Kabat, E A and Wu, TT; *Identical V region amino acid sequences and segments of sequences in antibodies of different specificities: relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites*; The journal of immunology; 147; 5; 1991; p 1709-1719

Kang, AS; Barbas, C; Janda, KD; Benkovic, SJ and Lerner, RA; *Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces*; Proceedings of the national academy of sciences; 1991; 88; p 4363-4366

Kanno, S; Yanagida, Y; Haruyama, T; Kobatake, E and Aizawa, M; *Assembling of engineered IgG-binding protein on gold surface for highly oriented antibody immobilization*; Journal of biotechnology; 2000; 76; p 207-214

Karyakin, AA; Presnova, GV; Rubtsova, MY and Egorov, AM; *Oriented Immobilization of antibodies onto the gold surfaces via their native thiol groups*; Analytical chemistry; 2000; 72; p 3805-3811

Keohavong, P and Thilly, W G; *Fidelity of DNA polymerases in DNA amplification; Biochemistry*; 86; 1989; p 9253-9257

Knight, DM; Wagner, C; Jordan, R; McAleer, MF; DeRita, R; Fass, DN; Coller, BS; Weisman, HF and Ghrayeb, J; *The immunogenicity of the 7e3 murine monoclonal Fab antibody fragment variable region is dramatically reduced in humans by substitution of human for murine constant regions*; Molecular Immunology; 1995; 32; 16; p 1271-1281

Kousparou, CA; Epenetos, AA and Deonarain, MP; *Antibody-guided enzyme therapy of cancer producing cyanide results in necrosis of targeted cells*; International journal of cancer; 2002; 99; p 138-148

Krauss, J; Arndt, MAE; Vu, BK; Newton, DL and Rybak, SM; *Targeting malignant B-cell lymphoma with a humanized anti-CD22 scFv-angiogenin immunoenzyme*; British journal of haematology; 2005; 128; p 602-609

Krebber, A; Bornhauser, S; Burmester, J; Honegger, A; Willuda, J; Bosshard, H R and Plueckthun, A; *Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system*; Journal of Immunological Methods; 201; 1997; p 35 – 55

Krinner, EM; Hepp, J; Hoffmann, P; Bruckmaier, S; Petersen, L; Petsch, S; Parr, L; Schuster, I; Mangold, S; Lorenczewski, G; Lutterbuese, P; Buziol, S; Hochheim, I; Volkland, J; Mølhøj, M; Sriskandarajah, M; Strasser, M; Itin, C; Wolf, A; Basu, A; Yang, K; Filpula, D; Sørensen, P ; Kufer, P; Baeuerle, P and Raum, T; *A highly stable polyethylene glycol-conjugated human single-chain antibody neutralizing granulocyte-macrophage colony stimulating factor at low nanomolar concentration*; Protein engineering design and selection; 2006; 19; 10; p 461-470

Kuegler, M; Stein, C; Schwenkert, M; Saul, D; Vockentanz, L; Huber, T; Wetzel, SK, Scholz, O; Plueckthun, A; Honegger, A and Fey, GH; *Stabilization and humanization of a single-chain Fv antibody fragment specific for human lymphocyte antigen CD19 by designed point mutations and CDR-grafting onto a human framework*; Protein engineering design and selection; 2009; 22; 3; p 135-147

Kubetzko, S; Balic, E; Walbei, R; Zangmeister-Wittke, U and Pluekthun, A; *PEGylation and multimerization of the anti-p185^{HER-2} single chain Fv fragment 4D5*; The journal of biological chemistry; 2006; 281; 46; p 35186-35201

Kwok, S; Kellogg, DE; McKinney, N; Spasic, D; Goda, L; Levenson, C and Sninsky, JJ; *Effects of primer - template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies*; Nucleic acids research; 18; 4; 1990; p 999-1005

Labrijn, AF; Aalberse, RC and Schuurman, J; *When binding is enough: nonactivating antibody formats; Current opinion in immunology*; 2008; 20; p 479-485

Labrou, NE; *Random mutagenesis methods for in vitro directed enzyme evolution*; Current protein and peptide science; 2010; 11; p 91-100

Lei, S; Lin, H; Wang, S; Callaway, J and Wilcox, G; Characterization *of the Erwinia carotovora pelB gene and its product pectate lyase*; Journal of bacteriology; 169; 9; 1987; p 4379-4383

Lesk, AM; *Introduction to protein science: Architecture, function and genomics*; 2004; p 267-273

Lin-Goerke, J L; Robbins, D J and Burczak; *PCR-based random mutagenesis using manganese and reduced dNTP concentration*; Biotechniques; 23; 3; 1997; p 409-412

Saviranta, P; Pajunen, M; Jauria, P; Karp, M; Pettersson, K; Mantsala, P and Lovgren, T; *Engineering the steroid-specificity of an anti-17b-estradiol Fab by random mutagenesis and competitive phage panning*; 11; 2; 1998; p 143-152

Lori, JA; Morrin, A; Killard, AJ; Smyth, MR; *Development and characterization of nickel-NTA-polyaniline modified electrodes*; Electroanalysis; 2006; 18; 1; p 77-81

Love, TE; Redmond, C and Mayers, CN; *Real time detection of anthrax spores using highly specific anti-EA1 recombinant antibodies produced by competitive panning*; Journal of immunological methods; 2008; 334; p 1-10

Makrides, SC; *Strategies for achieving high level expression of genes in Escherichia coli*; Microbiology reviews, 60; 3; 1996; p 512-538

McCafferty, J; Griffiths, AD; Winter, G and Chiswell, DJ; *Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains*, Nature, 348, 6301, 1990, p 552-554

Mersich, C and Jungbauer, A; *Generic method for quantification of FLAG-tagged fusion proteins by a real time biosensor*; Journal of biochemical and biophysical methods; 2007; 70; p 555-563

Miyazaki, C; Iba, Y; Yamada, Y; Takahashi, H; Sawada, J and Kurosawa, Y; *Changes in the specificity of antibodies by site-specific mutagenesis followed by random mutagenesis*; Protein engineering; 1999; 12; 5; p 407-415

Mohanty, A K; and Wiener, M C; *Membrane protein expression and production: effects of polyhistidine tag length and position*; Protein Expression and Purification; 33; 2004; p 311 – 325

Nanduri, V; Bhunia, AK; Tu, SI; Paoli, GC and Brewster, *SPR biosensor for the detection of L. Monocytogenes using phage-displayed antibody*; Biosensors and bioelectronics; 2007; 23; p 248-252

Neylon, C; Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution; Nucleic acids research; 2004; 32; 4; 1448-1459

O'Kennedy, R; Townsend, S; Donohoe, GG; Leonard, P; Hearty, S and Byrne, B; *Speedy, small, sensitive, and specific—reality or myth for future analytical methods*; Analytical letters; 2010; 43; p 1630-1648

Olins, PO and Rangwala, SH; *A novel sequence element derived from bacteriophage T7 mRNA act as an enhancer of translation of the lacZ gene in Escherichia coli*, The journal of biological chemistry; 269; 29; 1989; p 16973-16976

Orfanoudakis, G; Karim, B; Bourel, D and Weiss, E; *Bacterially expressed FABS of monoclonal antibodies neutralizing tumor necrosis factor alpha In vitro retain full binding and biological activity*; Molecular Immunology; 30, 16; 1993; p 1519 - 1528

Osborne, BA; Kindt, TJ and Goldsby, RA; *Kuby immunology*; 6; 2007; figure 5.25b

Pan, J: Fujiwara, N; Oka, S; Maekura, R; Ogura, T and Yano, L; *Anti-cord factor (trehalose 6, 6'-dimycolate) IgG antibody in tuberculosis patients recognizes mycolic acid subclasses*; Microbiology and Immunology 1999; 43; p 863–869

Park, JW; Hong, K and Kirpotin, DB; Anti-HER2 *Immunoliposomes: Enhanced Efficacy Attributable to Targeted Delivery; Clinical cancer research*; 2002; 8; p 1172-1181

Peluso, P; Wilson, DS; Do, D; Tran, H; Venkatasubbaiah, M; Quincy, D; Heidecker, B; Poindexter, K; Tolani, N; Phelan, M; Witte, K; Jung, LS; Wagner, P and Nock, S; *Optimizing antibody immobilization strategies for the construction of protein microarrays*; Analytical biochemistry; 2003; 312; p 113-124

Pirevincezi, RT; Reichert, WM and Hellinga, *Genetic engineering of a single-chain antibody fragment for surface immobilization in an optical biosensor*; Biosensors and bioelectronics; 1998; 13; 3-4; p 305-312

Presta, LG; *Engineering of therapeutic antibodies to minimize immunogenicity and optimize function*; Advanced drug delivery reviews; 2006; 58; p 640-656

Rippmann, JF; Pfizemaier, K; Mattes, R; Rettig, WJ and Moosmayer, D; *Fusion of the tissue factor extracellular domain to a tumour stroma specific single-chain fragment variable antibody results in an antigen-specific coagulation-promoting molecule*; Biochemical journal; 2000; 349; p 805-812

Sarkar, G; Kapelner, S and Sommer, SS; *Formamide can dramatically improve the specificity of PCR*; Nucleic Acids Research; 18, 24, 1990, p 7465

Schneider, WP; Glaser, SM; Kondas, JA and Hakimi, J; *The anti-idiotypic response by cynomolgus monkeys to humanized anti-tac is primarily directed to complementarity-determining regions H1, H2, and L3*; The journal of immunology; 1993; 150; 7; p 3086-3090

Schroff, RW; Foon, KA; Beatty, SM; *Human anti-murine immunoglobulin responses in patients receiving monoclonal antibody therapy*; Cancer research; 1985; 45; p 879-885

Scibek, JJ; Evergren, E; Zahn, S; Canziani, GA; Van Ryk, D and Chaiken, IM; Biosensor analysis of dynamics of interleukin 5 receptor subunit β_c interaction with IL5:IL5Rα complexes; Analytical biochemistry; 2002; 307; p 258-265

Sheibani, N; *Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins*; Preparative Biochemistry and Biotechnology; 29: 1; 1999; p 77 – 90

Shafikhani, S; Siegel, R A; Ferrari, E; and Schellenberger, V; *Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization*; Biotechniques; 23; 2; 1997; p 304-310

Short, JM; Fernandez, JM; Sorge, JA and Huse, WD; *λ ZAP: a bacteriophage λ expression vector with in vivo excision properties*; Nucleic acids research; 16; 1988; p 7583-7600

Siegel, DL; *Recombinant monoclonal antibody technology*; Transfusion and clinical biology; 2002; 9; p15-22

Skerra, A; Pfitzinger, I and Pluekthun; *The functional expression of antibody Fv fragments in Escherichia coli: improved vectors and a generally applicable purification technique*; Bio/technology; 9; 1991; p 273-278

Smith, KA; Nelson, PN; Warren, P; Astley, SJ; Murray, PG and Greenman; *Demystified...recombinant antibodies*; Journal of clinical pathology; 2004; 57; 912-917

Tabor, S and Richardson, C C; *Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I*; Proceedings of the national academy of sciences; 86; 1989; p 4076-4080

Tindall, K R and Kunkel, T A; *Fidelity of DNA Synthesis by the Thermus aquaticus DNA Polymerase*; Biochemistry; 27; 16; 1988; p 6013-6020

Turner, M; Roitt, I (Ed); Brostoff, J (Ed) and Male, D (Ed); *Antibodies* in *Immunology*; 2001; 6; p 65-83

Van Beusechem, VW; Grill, J; Mastenbroek, J; Wickham, TJ; Roelvink, PW; Haisma, HJ; Lamfers, LM; Dirven, CMF; Pinedo, HM and Gerritsen, WR; *Efficient and selective gene transfer into primary human brain tumors by using single-chain antibody-targeted adenoviral vectors with native tropism abolished*; Journal of virology; 2002; 76; 6; p 2753-2762

Vanhercke, T; Ampe, C; Tirry, L and Denolf, P; *Reducing mutational bias in random protein libraries; Analytical biochemistry*; 339; 2004; p 9-14

van Wezenbeek, P and Schoenmakers, JGG; *Nucleotide sequence of the genes III, VI and I of bacteriophage M13*; Nucleic acids research; 6, 8; 1979; p 2799-2818

Van Wyngaardt, W; Malatji, T; Mashau, C; Fehrsen, J; Jordaan, F; Miltiadou, D and du Plessis, D; *A large semi-synthetic single-chain Fv phage display library based on chicken immunoglobulin genes*; BMC Biotechnology; 4, 6; 2004

Vartanian, J P; Henry, M and Wain-Hobson, S; Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions; Nucleic acids research; 24; 14; 1996; 2627-2631

Voet, D and Voet, JG; *Biochemistry*; 3; 2004; p 94, 109, 116, 1218 and 1290

Waugh, D S; *Making the most of affinity tags*; Trends in Biotechnology; 23, 6; 2005; p 316 – 320

Wingren, C; Steinhauer, C; Ingvarsson, J; Persson, E; Larsson, K and Borrebaeck; *Microarrays based on affinity-tagged single-chain Fv antibodies: Sensitive detection of analyte in complex proteomes*; Proteomics; 2005; 5; p 1281-1291

Wong, T S; Zhurina, D and Schwaneberg, U; *The diversity challenge in directed protein evolution*; Combinatorial chemistry and high throughput screening; 9; 2006; p 271-288

Zagursky, RJ and Berman, ML; *Cloning vectors that yield high levels of single-stranded DNA for rapid DNA sequencing*; Gene; 1984; 27; p 183-191

Zahnd, C; Spinelli, S; Luginbuehl, B; Amstutz, P; Cambillau, C and Pluekthuen, A; *Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity*; The journal of biological chemistry; 279; 18; 2004; p 18870-18877

Zhu, C T; Zeng, X B and Huang, W D; *Codon usage decreases the error minimization within the genetic code*; Journal of molecular evolution; 57; 2003; p 533-537

Zinder, ND and Horiuchi, K; *Multiregulatory element of filamentous bacteriophages*; Microbiological reviews; 1985; 49; 2; p 101-106