ISOLATION OF ANTIGENIC PEPTIDES OF *COWDRIA RUMINANTIUM* AND THEIR ENCODING GENES USING A GENOME-DERIVED PHAGE DISPLAY LIBRARY.

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

RHODES UNIVERSITY

by

JEANNI FEHRSEN

March 2003

ABSTRACT

The development of new and effective vaccines and immunodiagnostic reagents requires the characterisation of antigenically relevant proteins and their interactions with the products of the immune system. Phage display technology was investigated as a means of elucidating some of the antigenic properties of the rickettsial parasite, *Cowdria ruminantium* (*Cowdria*). Randomly fragmented gene-derived libraries have been useful in elucidating viral and other epitopes, but only limited work has been done with entire genomes. A phage display library expressing a repertoire of *Cowdria* peptides was constructed. It was sufficiently large to represent the organism's genome, but lacked phages displaying peptides coded for by genes containing a *Pvu* II restriction enzyme site, including the one coding for the major antigenic protein 1 (MAP1). This was considered advantageous since MAP1 is immunodominant and has already been well characterised. Affinity selection with antibodies against *Cowdria* proteins other than MAP1 allowed several antibody-reactive peptides to be isolated. These selected sequences were placed in the context of the genome by screening a lambda bacteriophage library and by comparison with *Cowdria* DNA sequences. Apart from showing that antigenic mimics were present in the phage display library, six open reading frames encoding putative *Cowdria* proteins were identified. All had similarities to, or motifs in common with, membrane proteins and are thus likely to be exposed to the host's humoral immune system. Some of the proteins identified were larger than the antigens used to elicit the antibodies used for selection, probably as a result of the presence of cross-reactive epitopes.

Despite limitations experienced when extending a fragmented-gene approach for epitope location to genomes, it was possible to identify an antigenic region on MAP1 by comparison with selected mimics. In addition, binding peptide sequences were identified with two monoclonal antibodies that had been raised against non-*Cowdria* antigens. An epitope on the VP7 protein of bluetongue virus was identified and peptides were found that reacted with a monoclonal antibody directed against

malignant catarrhal fever virus. Thus, apart from being able to identify several potentially important *Cowdria* epitopes and genes, the fragmented-genome library holds promise as a universal reagent for identifying useful mimics.

TABLE OF CONTENTS

List of Abbreviations.......... viii
List of Figures.......... xi
List of Tables.......... xiii
Acknowledgements.......... xiv
Communications resulting from this work.......... xv

CHAPTER 1: LITERATURE REVIEW.

1.1. INTRODUCTION.......... 1
- 1.1.1. Antigens and antibodies.......... 2
- 1.1.2. Types of epitopes.......... 5
- 1.1.3. Antibody specificity.......... 6
- 1.1.4. Epitope mapping.......... 7

1.2. IDENTIFYING PROTEINS AND THEIR ENCODING GENES.......... 9
- 1.2.1. Protein sequencing.......... 9
- 1.2.2. Expression libraries.......... 10
- 1.2.3. Phage display.......... 11
  - 1.2.3.1. Filamentous bacteriophage.......... 11
  - 1.2.3.2. The origin of phage display.......... 14
  - 1.2.3.3. Affinity selection.......... 15
  - 1.2.3.4. Random peptide libraries.......... 19
  - 1.2.3.5. Other phage display systems.......... 20
  - 1.2.3.6. Epitope mapping with random display libraries.......... 24
  - 1.2.3.7. Fragmented-gene phage display.......... 26
  - 1.2.3.8. Fragmented-genome and cDNA phage display.......... 30
  - 1.2.3.9. Phage display: a tool for vaccine development.......... 33
  - 1.2.3.10. Antibodies expressed on phage.......... 35
- 1.2.4. Genomics.......... 36

1.3. *Cowdria ruminantium* 37
- 1.3.1. The disease, heartwater.......... 37
- 1.3.2. Classification.......... 38
- 1.3.3. Symptoms of heartwater.......... 39

1.3.4. The life cycle of *Cowdria*........ 39
1.3.5. *In vitro* cultivation of *Cowdria*........ 40
1.3.6. Control of heartwater........ 40
1.3.7. New approaches to heartwater control........ 41
1.3.8. Antigenic characterisation of *Cowdria*........ 44
1.3.9. Molecular characterisation of *Cowdria*........ 45
1.3.10. Serodiagnosis........ 48
1.3.11. DNA-based diagnosis........ 49
1.3.10. Immunity to heartwater........ 50

CHAPTER 2: CONSTRUCTION AND ANALYSIS OF A LAMBDA ZAP II LIBRARY.
2.1. INTRODUCTION........ 54
2.2. MATERIALS AND METHODS........ 56
2.2.1. *In vitro* cultivation of *Cowdria*........ 56
2.2.2. Purification of *Cowdria* organisms........ 57
2.2.3. Isolation of *Cowdria* DNA........ 57
2.2.4. Determining the bovine DNA content in the *Cowdria* DNA preparation........ 58
2.2.5. Bacteriophage λ library construction........ 59
2.2.6. Antibodies used for screening the λ ZAP II library........ 60
2.2.7. λ library screening with antibodies........ 61
2.2.8. *Map*1 primers and PCR........ 62
2.2.9. *Map*1 probe synthesis and hybridisation........ 62
2.2.10. Excision of pBluescript phagemids from the λ ZAP II genome........ 63
2.2.11. Protein expression from phagemid clones........ 63
2.2.12. Sequencing of the *map*1 gene........ 64
2.3. RESULTS........ 64
2.3.1. Genomic library construction........ 64
2.3.2. Screening the library with antibodies........ 67
2.3.3. Isolation and characterization of the *map*1 gene........ 67
2.4. DISCUSSION........ 70

CHAPTER 3: CONSTRUCTION OF A FRAGMENTED-GENOME PHAGE DISPLAY LIBRARY.

3.1. INTRODUCTION... 73
3.2. MATERIALS AND METHODS.. 75
3.2.1.Preparation of the *Cowdria* DNA fragments.. 75
3.2.2. Phage display vector preparation.. 76
3.2.3. Ligation and electroporation.. 77
3.2.4. PEG precipitation of phages... 77
3.2.5. Titration of phage solutions.. 78
3.2.6. PCR of fUSE2 bacterial colonies... 78
3.2.7. Sequencing of fUSE2 clones... 78
3.3. RESULTS.. 79
3.3.1. Preparation of the *Cowdria* DNA fragments.. 79
3.3.2. The construction of the *Cowdria* fragmented-genome phage display library..... 80
3.4. DISCUSSION... 82

CHAPTER 4: IDENTIFICATION OF *COWDRIA* EPITOPES AND PROTEINS.

4.1. INTRODUCTION... 84
4.2. MATERIALS AND METHODS.. 86
4.2.1. Antisera.. 86
4.2.2. Affinity purification of antibodies... 86
4.2.3. Affinity selection of phage displayed peptides (panning).................................. 87
4.2.4. Antigenic reactivity of the selected fusion peptides... 89
4.2.5. Screening the λZAP II *Cowdria* library... 90
4.2.5.1. Radiolabelled deoxyoligonucleotide probes...................................... 90
4.2.5.2. DIG-labelled deoxyoligonucleotide probes....................................... 91
4.2.6. Sequencing and analysing the λZAP II clones.. 91
4.2.7. *Cowdria* genome sequence database search.. 92
4.2.8. Bacterial expression of potential ORFs.. 93
4.3. RESULTS.. 94
4.3.1. Affinity selection with MAP1 antibodies... 94
4.3.1.1. Antigenicity of the selected fusion peptides...................................... 96
4.3.1.2. Screening the λZAP II *Cowdria* genomic library.............................. 98

4.3.1.3. *Cowdria* genome sequence database search..................................... 99
4.3.2. Affinity selection with antibodies against variable immunodominant Gardel isolate proteins..................................... 101
4.3.2.1. BLAST database analysis of peptide sequences..................................... 106
4.3.2.2. Antigenicity of the selected fusion peptides..................................... 106
4.3.2.3. Screening the λZAP II *Cowdria* genomic library..................................... 108
4.3.2.4. Characterisation of the potential ORFs..................................... 113
4.3.2.5. *Cowdria* genome sequence database search..................................... 115
4.3.3. Affinity selection with antibodies against potentially protective proteins of the Welgevonden isolate..................................... 118
4.3.3.1. *Cowdria* genome sequence database search..................................... 121
4.3.4. Summary of ORFs identified..................................... 122
4.4. DISCUSSION..................................... 126

CHAPTER 5: ANTIGENIC MIMICS IN THE FRAGMENTED-GENOME LIBRARY.
5.1. INTRODUCTION..................................... 132
5.2. MATERIALS AND METHODS..................................... 133
5.2.1. Antibodies..................................... 133
5.2.2. Affinity selection..................................... 134
5.2.3. Antigenic reactivity..................................... 134
5.2.4. Sequence analysis..................................... 135
5.3. RESULTS..................................... 135
5.3.1. Affinity selection with monoclonal antibodies against VP7 of BTV..................................... 135
5.3.2. Affinity selection with a monoclonal antibody against MCFV..................................... 136
5.3.3. Potential *Cowdria* antigenic mimics..................................... 138
5.4. DISCUSSION..................................... 143

CONCLUDING DISCUSSION..................................... 146
APPENDICES..................................... 155
REFERENCES..................................... 164

List of Abbreviations

AT	Adenine and thymine
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine serum albumin
BTV	Bluetongue virus
CDR	Complementarity determining regions
CFE	Continuous flow electrophoresis
CFU	Colony forming unit
cos	Cohesive ends
Cowdria	*Cowdria ruminantium*
CTAB	Hexadecyltrimethyl ammonium bromide
DIG	Digoxigenin
dH_2O	Distilled water
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
FPLC	Fast performance liquid chromotography
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immuno deficiency virus
HSP	Heat shock proteins
IFA	Indirect fluorescent antibody
INF-γ	Interferon gamma
IL-2	Interleukin-2
IPTG	Isopropyl-β-D-thiogalactopyranoside
h	Hour
kb	Kilo bases
λ	Lambda
LB	Luria-Bertani
mAb	Monoclonal antibody
MAP1	Major antigenic protein 1
MAP2	Major antigenic protein 2

MCFV malignant catarrhal fever virus
MHC Major histocompatibility complex
NaN_3 Sodium azide
OMP Outer membrane protein
ORF Open reading frame
PAGE Polyacrylamide gel electrophoresis
PBMC Peripheral blood mononuclear cells
PBS Phosphate-buffered saline
PBS1%MP PBS containing 1% milk powder
PBS2%MP PBS containing 2% milk powder
PBST PBS containing 0.05% Tween 20
PCR Polymerase chain reaction
PFU Plaque forming unit
PM/Scl Polymyositis / scleroderma
ProtScale Profile produced by amino acid scales
PSI-BLAST Position specific iterated BLAST
PSORT Prediction of protein localisation sites in cells
RAPD Random amplified polymorphic DNA
rBoIFNγ Recombinant bovine interferon gamma
RF Replicative form
RITE Reverse identification of transcriptional effectors
rRNA Ribosomal RNA
RSCU Relative synonymous codon usage
s Seconds
SDS Sodium dodecyl sulfate
SDS-PAGE SDS-polyacrylamide gel electrophoresis
SMART Simple Modular Architecture Research Tool
ss Single stranded
TEMED N,N,N’,N’-Tetramethyl-ethylenediamine
TBS Tris buffered saline
TBS-T.05 TBS containing 0.05% Tween 20
TBS-T 0.5 TBS containing 0.5% Tween 20
TNF Tumour necrosis factor

U	Units
VP7	Viral capsid protein 7
X-gal	5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside

LIST OF FIGURES.

		Page
Figure 1.1	The Ff bacteriophage.	12
Figure 1.2	Life cycle of filamentous bacteriophages.	13
Figure 1.3	Schematic representation of affinity selection (panning) of bacteriophages.	17
Figure 1.4	Schematic representation of the pJuFo system.	23
Figure 2.1	Schematic representation of the construction of the *Cowdria* λZAP II library.	55
Figure 2.2	Determination of bovine DNA content.	66
Figure 2.3	Western blots with goat antisera.	68
Figure 2.4	Agarose gel electrophoresis showing PCR of the Welgevonden isolate *map*1 gene from genomic DNA and λ clones and a western blot of the MAP1 expressing clone.	69
Figure 3.1	Schematic representation of the construction of the *Cowdria* fragmented-genome phage display library.	74
Figure 3.2	Agarose gel electrophoresis showing stages of *Cowdria* DNA fragment preparation.	81
Figure 4.1	Western blots and panning results obtained with MAP1 goat antiserum.	95
Figure 4.2	Sequence and immunoreactivity of fusion phage displayed peptides selected by affinity-purified MAP1 antibodies.	97
Figure 4.3	Specificity of hybridisation of the type 1-2 and type 4-5 oligonucleotide radiolabelled probes and the deduced amino acid sequence of the λ clones identified with the probes.	100
Figure 4.4	Schematic representation of the positions of peptides on the λ clones or the *Cowdria* genome regions.	101
Figure 4.5	Western blots and panning results obtained with antisera directed against Gardel isolate proteins.	102
Figure 4.6	Peptides selected by the rabbit antibodies directed against the 23 to 29 kDa range of Gardel isolate proteins.	105
Figure 4.7	Western blots of recombinant phage proteins reacting with rabbit antiserum against 23 to 29 kDa Gardel isolate proteins.	107
Figure 4.8	Hybridisation specificity and screening with the F12 DIG labelled oligonucleotide probe.	110
Figure 4.9	Deduced amino acid sequence of the *Cowdria* λ clones containing the F12, H1, E3 and L24 DNA sequences.	112
Figure 4.10	Coomassie blue stained PAGE gel of F12orf cloned into the vector pGEX2-T.	114

Figure 4.11	Deduced amino acid sequences of the *Cowdria* genome regions matching peptides M37, N49, D25, G4, I8, J15 and K18.	116
Figure 4.12	Western blots obtained with antisera directed against Welgevonden isolate proteins.	119
Figure 4.13	Deduced amino acid sequences of the *Cowdria* genome regions matching peptides O4, P8, Q23 and R36.	123
Figure 5.1	Phages selected with monoclonal antibody 15-A directed against MCFV.	139
Figure 5.2	Alignment of peptides with potential *Cowdria* ORFs.	141
Figure 5.3	Alignment of peptides with F12orf and M37orf.	142

LIST OF TABLES.

		Page
Table 4.1	Oligonucleotide probes derived from the selected phage displayed sequences.	99
Table 4.2	Peptides selected with antibodies against the 23 to 29 kDa range of Gardel isolate proteins.	104
Table 4.3	Comparison of peptide 47 with the corresponding region of the known NADH dehydrogenase proteins.	106
Table 4.4	Hybridisation profiles of oligonucleotide probes derived from the phage displayed sequences selected by rabbit antibodies against the 23 to 29 kDa range of Gardel isolate proteins.	109
Table 4.5	The results of pannings with antibodies directed against the 23 to 24 kDa and 18 to 22 kDa Welgevonden isolate proteins.	121
Table 4.6	Summary of genes identified.	124
Table 4.7	The GC content and codon usage of DNA sequences encoding peptides selected by phage display. Epitope prediction indicates whether the peptide would have been predicted as an antigenic region.	126
Table 5.1	Peptides selected with antibodies against VP7 of BTV.	136
Table 5.2	Results of pannings with monoclonal antibody 15-A directed against MCFV.	138

Acknowledgements.

I am sincerely grateful to all my colleagues, family and friends who have helped, advised and encouraged me during the last few years. Specifically, I want to thank my supervisor at Onderstepoort Veterinary Institute, Dr. D. H. du Plessis for his never-ending advice, ideas and encouragement. He introduced me to phage display and the world of antibodies. Without him this thesis would not have been written. Also, I thank my supervisor at Rhodes University, Prof. D. A. Hendry for all his kind words and advice on the project.

I appreciate all the time and energy my colleagues in the Immunology Division, Onderstepoort Veterinary Institute have devoted to me. Mirinda van Kleef advised me on the intricacies of *Cowdria*, supplied me with useful antibodies and 'suffered' with me. Frances Jordaan taught me how to work with filamentous phages.

I would also like to thank Dr. E. Zweygarth and A. Josemans for providing the *Cowdria* cell cultures, Drs. N. C. Collins, E. P. de Villiers and B. A. Allsopp for giving me access to the *Cowdria* genomic sequence data, C. Potgieter for his help with cloning and Dr. C. Bremer for collaborating with the MCFV work.

Lastly, I am indebted to Onderstepoort Veterinary Institute for the financial support and allowing me to further my education.

COMMUNICATIONS RESULTING FROM THIS WORK:

PUBLICATIONS:

Brayton, K. A., **Fehrsen, J.**, de Villiers, E. P., van Kleef, M. and Allsopp, B. A., 1997. Construction and initial analysis of a representative λZAPII expression library of the intracellular rickettsia *Cowdria ruminantium*: cloning of *map1* and three other *Cowdria* genes. *Veterinary Parasitology*. 72, 185-99.

Fehrsen, J. and du Plessis, D. H., 1999. Cross-reactive epitope mimics in a fragmented-genome library derived from the rickettsia, *Cowdria ruminantium*. *Immunotechnology.* 4, 175-84.

TALKS:

Du Plessis, D.H. and **Fehrsen, J**., Identification by phage display of peptide epitopes that cross-react with antibodies to the major antigenic protein of the rickettsia *Cowdria ruminantium*. “The Biomedical and biotechnological use and further development of Phage Display". 1997 Meeting of EU phage club, Smolenice, Slovakia.

Fehrsen, J. and du Plessis, D. H., Identification of genes encoding antigenic proteins of *Cowdria ruminantium* by fragmented-genome phage display. BIOY2K Combined Millennium Meeting (SASM, SASBMB), Grahamstown, January 2000.

POSTERS:

Fehrsen, J., Brayton, K. A. and Allsopp, B A., Searching for *Cowdria ruminantium* antigen genes. The 2nd International conference on Tick-borne Pathogens at the Host-Vector Interface, Aug 1995, Kruger Park.

Fehrsen, J. and du Plessis, D. H., Construction of a filamentous phage library displaying peptides derived from *Cowdria ruminantium.* The 4th Biennial Meeting of the Society for Tropical Veterinary Medicine, May 1997, Montpellier, France.

Du Plessis, D. H., **Fehrsen, J.**, van Kleef, M., Gunter, N. and Macmillan, H., Identification of protective and diagnostic epitopes of *Cowdria ruminantium*. Vaccines and Immunotherapeutics conference, Lorne, Australia, March 2000.

CHAPTER 1

LITERATURE REVIEW.

1.1. INTRODUCTION.

Specific diagnostic tests are required to help with the control of livestock diseases. Once diagnosed, vaccines can be administered thereby reducing stock losses and preventing reduced productivity due to disease. A common method of diagnosis is to detect antibodies to the organisms in the sera of suspect animals. For an effective antibody-based test, an antigen specific to the organism is usually required. Whole organisms are often used but the constituent proteins can contain cross-reactive epitopes that react with antibodies against closely related organisms, resulting in a false diagnosis. Recombinant antigens have advantages for specific serological tests but detailed information about the antigens and their encoding genes is usually required.

Veterinary diseases can be controlled by vaccination. Many veterinary vaccines are preparations of live organisms of a mild or attenuated strain that are administered to the animals. In some cases, the source of the organisms is blood from infected animals, which needs to be kept frozen until administered. There is of course a risk that the blood can contain other pathogens. Cross-protection between strains or isolates is not always achieved. Identifying and cloning the genes encoding the antigens that induce a protective immune response in the animal can help to address some of these problems. For instance, a cocktail of antigens or peptides from different isolates could be included in a vaccine to provide comprehensive immunity against a range of strains present in the field. It is therefore imperative to have an understanding of the molecular interactions between the antigens of the disease-causing organism and the components of the host's immune system before

serodiagnostic tests and vaccines can be developed. Identifying the relevant antigens and isolating the encoding genes are important steps towards these goals. If the nucleotide sequence of the organism's genome is at hand, potential protein coding regions can be identified. Many virus genomes have been sequenced and in some cases the interactions with the immune system have been well characterised. With more complex organisms and where the genome sequence is not yet available, a more empirical approach is required. This study describes an approach aimed at identifying genes encoding antigenic proteins of the intracellular rickettsial parasite *Cowdria ruminantium* (*Cowdria*), the causative agent of heartwater in ruminants. When this study was commenced, very limited sequence information was available. This necessitated the use of the interactions between antibodies against *Cowdria* and its antigens to identify the relevant proteins.

1.1.1. Antigens and antibodies.

When foreign material enters the body, components of the immune system recognise and dispose of it. The natural, or innate, branch of the immune system responds non-specifically to foreign entities by processes such as phagocytosis by macrophages, cell lysis by natural killer cells and by antimicrobial factors such as complement, acute phase proteins, lysozyme and interferons. This process does not improve with repeated exposure. Acquired or adaptive immunity on the other hand, depends on specific interactions and repeated exposures result in enhanced responses to specific foreign entities (Roitt, 1997). This interaction involves molecular recognition between the foreign entity and components of the immune system such as antibodies (secreted proteins) and receptors on cells (lymphocytes) of the immune system. Two types of lymphocytes exist. B cells develop in the bone marrow and differentiate into antibody producing plasma cells while T cells differentiate in the thymus. T cells recognise foreign entities only when they are presented as processed peptides in association with the major histocompatibility complex (MHC) molecules

(Roitt, 1997). Their functions include helping B cells to make antibody, killing virus-infected cells, regulating the level of the immune response and stimulating microbiocidal and cytotoxic activity of other cells of the immune system. Antibodies produced by B cells recognise the foreign entities free in solution or on the surface of cells. When an antibody expressed on the surface of a B cell reacts with a foreign molecule, the cell is signalled to proliferate and produce more of the same antibody. Upon a second exposure, the cells will expand and differentiate. The immunoglobulin genes undergo affinity maturation due to somatic mutations, enabling a more efficient secondary response. The acquired immune system can discriminate between millions of discrete molecular structures. This is achieved by the diversity of lymphocytic receptors derived from multiple variable germ line genes encoding the immunoglobulin and T cell receptors. Rearrangements, deletions, additions and somatic mutations all contribute to the diversity of the immune repertoire (Bona, 1996). In humans and mice (and many other animals) diversity is generated by rearrangement of gene segments. But there are alternative ways to generate diversity. For example, chickens and rabbits use gene conversion to alter the single heavy-chain locus with copies of variable-segment pseudogenes, while sharks have multiple rearranged variable-light regions in the germline genome and diversity is generated by simply activating expression of different copies. Since the immune system can discriminate between such a great diversity of molecular structures, there is a belief that immunological recognition is the most specific differentiation mechanism that exists in biology. If this were true we would all be dead, since without cross-reactions (see 1.1.3) we would not be able to 'handle' all the different molecular structures of all the strains of organisms and viruses that we can possibly encounter (van Regenmortel, 1998).

The foreign entities, which can be cells, organelles or proteins are called antigens if they are able to elicit an immune response and be recognised by the products of the immune response (Atassi,

1984). Most of our understanding of the process of immunological recognition is concerned with the interaction between proteins and antibodies. Antigenicity is restricted to certain parts of the protein known as the antigenic determinants, or epitopes (Jerne,1960). Thus epitopes are discrete patches on the surface of the molecules that bind specifically to the binding site, or "paratope", of antibodies. Proteins are folded molecules with a tertiary structure and it is thought that almost the entire accessible surface of the protein is a continuum of overlapping epitopes which can be recognised by antibodies (Berzofsky, 1985).

An epitope is formed by an array of atoms arranged on the surface of a molecule and which are recognised by the antibody. The 'antibody footprint', made by the complementary paratope was estimated to be 800$Å^2$ (van Regenmortel, 1998). The paratope mediates the recognition of the antigenic determinant by the antibody. It consists of six loops of hypervariable sequences at the ends of two variable domains of the antibody molecule. The six loops are formed by the three complementarity determining regions (CDRs) of the variable light chain and the three on the variable heavy chain (Capra and Kehoe, 1975). Two identical paratopes are present on each immunoglobulin molecule. Antibodies represent a unique functional design capable of recognising virtually any molecular structure. Classically the paratope was thought to be a pocket or groove into which the epitope fits, but X-ray crystallography data has shown that there can be mutual structural adaption involved in the antibody-antigen interaction (van Regenmortel, 1998). Within the paratope there are critical binding residues which contribute both to the binding energy, and/or its overall structure. The structural epitope is thought to be a fifteen to twenty amino acid region which is in contact with the paratope, but only atoms from three to five amino acid residues actually contribute to the binding energy of the interaction (van Regenmortel, 1995). Thus the relation between the epitope and the paratope has been described as having both a spatial and an activity component and accordingly belongs in a time-space continuum that cannot be separated. It is important to be aware

that epitopes and paratopes are termed as such due to their mutual complementarity and not by any intrinsic property of each entity independently (van Regenmortel, 1998).

1.1.2. Types of epitopes.

Epitopes can be classified as either discontinuous or continuous (Atassi and Smith, 1978). A discontinuous epitope consists of a cluster of residues that are not contiguous in sequence, but are in close proximity at the protein surface due to the folding of the peptide chain. Most antibodies recognise discontinuous epitopes. Continuous epitopes consist of amino acid residues that are in direct peptide linkage (Atassi, 1975; Atassi *et al.*, 1976).

Antigenic determinants have also been classified according to their binding properties. Heteroclitic epitopes show heterospecificity or antigenic cross-reactions, whereby the antibody binds more strongly to other epitopes or antigens than to the one against which it was raised (Mäkelä, 1965; Underwood, 1985). It is probably futile to search for the 'true' epitope since cross-reactive, rather than absolute fit is the rule underlying epitope-paratope interactions (van Regenmortel, 1998). This is due to the way antibody production is initiated. Clonal selection of B cells leads to antibody production, a process which can be triggered by a relatively low affinity binding of the antigen to the B cell receptor. The resulting antibodies are polyspecific and can have stronger binding to related antigens. Mimotopes on the other hand, are short peptides which contain critical binding residues that chemically mimic the antigenic determinants on the folded protein. Some mimotopes have no resemblance at all to the native protein's sequence (Geysen *et al.*, 1986). Mimotopes have since then been given an even broader meaning in that the mimic is only a true mimotope when it is able to elicit antibodies that react with the original antigen (van Regenmortel, 1999).

1.1.3. Antibody specificity.

Landsteiner (1962) concluded that the serological cross-reactivity he observed between cells from different animal species was a series of gradual transitions. He showed with a series of cross-absorption studies that an antigen was able to elicit not only one specific antibody, but a range of antibodies having various degrees of cross-reaction with related antigens. The cross-reactions of antibodies raised against a particular antigen have been exploited to show the relatedness in viruses, probably due to shared antigenic determinants. Shared cross-reactivity is when an antibody reacts with the same epitope on two different proteins, whereas true cross-reactivity is when an antibody recognises a different, structurally related epitope on a heterologous antigen. The homologous reaction is usually stronger than the heterologous reaction (van Regenmortel, 1998).

Antibodies are generally called specific if they can discriminate between two or more antigens at a required level. It may thus be more appropriate to refer to the discrimination *potential* than the specificity of an antibody. Lower-affinity antibodies are more discriminatory and therefore more specific than higher-affinity antibodies (Day, 1990). In the case of low affinity binders, the loss of fewer hydrogen bonds is needed to lower the affinity constant to below threshold such that binding is not detected. Some antibodies, therefore have to be used at high dilutions to result in more 'specific' reactions. Any single antibody can only be specific for an epitope, not for antigens, viruses or cells all of which possess many different antigenic determinants. Of particular relevance to this thesis is the observation that antibodies have the potential to be multi-specific. This can be explained by the fact that fifty amino acid residues contribute to the CDR of an antibody of which only a third contributes to each paratope. Thus one antibody can have two 'unrelated' paratopes. These usually overlap so binding to one epitope prevents binding to the second. Thus these are

fuzzy boundaries which need to be taken into account in most antibody-antigen interactions. This should not, however, hinder the development and application of immunochemical assays (van Regenmortel, 1998).

1.1.4. Epitope mapping.

It is useful to identify the actual regions on proteins that react with antibodies. This information helps in the design of recombinant vaccines and diagnostic reagents. Peptides that mimic epitopes on pathogens can elicit protective antibodies and thus induce a limited, but directed immune response. For example, an epitope on the highly variable loop on the virus protein 1 of foot-and-mouth disease virus elicits neutralising antibodies (Pfaff *et al*., 1982). Vaccination with a synthetic peptide based on this determinant was successful in mice, guinea pigs and a natural host, swine. The peptide was made more resistant to proteolysis by reversing the peptide bonds and synthesising D-analogues which mimic the original L-peptides (Guichard *et al.*, 1994). Peptides can be used as diagnostic reagents, especially in autoimmune disease where this approach has been useful as markers for example in primary biliary cirrhosis and multiple sclerosis (see 1.2.3.5 and 6).

Epitopes can be mapped using a variety of approaches, but overall it is usually best to combine data obtained from different methods. First, the protein must be identified by standard methods such as polyacrylamide gel electrophoresis (PAGE), two dimensional PAGE and western blots. Fragments of the protein obtained by proteolytic or chemical cleavage can then be tested for immunoreactivity (van Regenmortel, 1988). Alternatively, if the gene is available, portions can be expressed to help identify the antigenic region (Harlow and Lane, 1988). Determining the immunoreactivity of chemically synthesised peptides (five to seven amino acids long) can narrow down the antigenic site and, by mutating single amino acids at a time, the residues critical for binding can be elucidated

(Geysen *et al.*, 1984). A strategy whereby large numbers of peptides are synthesised on plastic pins and subsequently screened with antibodies identifies amino acid sequences that mimic discontinuous epitopes (mimotopes). Comparison to the native protein sequence identifies the antigenic determinant on the protein (Geysen *et al.*, 1987). This is a rapid method to identify epitopes but is relatively expensive and the protein's amino acid sequence needs to be known.

Probably the most definitive way to determine which residues are involved in the interaction between an antibody and a protein is by X-ray crystallography. Pure forms of both the antibody and antigen are required. Crystals are formed of the immune complexes which are then studied (van Regenmortel, 1988). Knowledge of the three dimensional structures of the antibodies and antigens is not enough to predict this interaction since accessibility is not the only criterion. Antibody binding can cause conformational changes that 'expose' residues which would not be accessible in the native form of the protein and which could play an important part in the binding (van Regenmortel, 1995). Nevertheless, many computer programs have been developed to predict antigenic regions on proteins from their amino acid sequence. They are based on, amongst other, hydrophilicity (Hopp and Woods, 1983), static accessibility (exposed regions; Novotny, 1987), segmental mobility (Westhof *et al.*, 1984) and amino acid residues occurring in experimentally determined epitopes (Kolaskar and Tongaonkar, 1990). Some programs combine several criteria for increased accuracy (Odorico and Pellequer, 2003). None of these are one hundred percent reliable yet. Experimental data still remain essential since the immunogenicity of peptides depends on many intrinsic factors and regulatory mechanisms of the host which are not amenable to molecular design, while experimental use of antibodies has proven to successfully identify immunogenic and antigenic regions of proteins (van Regenmortel, 1988).

1.2. IDENTIFYING PROTEINS AND THEIR ENCODING GENES.

Organisms consist of many proteins and before the epitopes that are responsible for an immune response can be identified, the proteins on which they are situated need to be identified and characterised. Western blotting is a method to identify antigenic proteins, in addition their molecular weights can be determined. Further characterisation can be done by a variety of techniques such as amino acid sequencing (see 1.2.1) or using antibodies to screen libraries (see 1.2.2 and 3) that express fusion proteins derived from the DNA of the protein/organism under study. Potential antigenic proteins can also be predicted from the genome sequence if available (see 1.2.4).

1.2.1. Protein sequencing.

To unequivocally characterise proteins, their attributes such as their isoelectric point and apparent protein mass have to be defined. More information can be obtained from the protein sequence (Edman and Begg, 1967). The sequence of the N-terminal end of the protein can be determined. If the protein has been fragmented, internal sequence data can be obtained by matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS; Wilkins and Gooley, 1997). Protein sequencing has aided in the process of isolating specific genes of intracellular rickettsiae such as *map*1 of *Cowdria* (van Vliet *et al.*, 1994) and *msp4* of *Anaplasma marginale* (Oberle *et al.*, 1993). In the case of *map*1 of *Cowdria,* the DNA sequence was deduced from two peptide sequences, primers were designed and a portion of the gene was amplified by the polymerase chain reaction (PCR).

1.2.2. Expression libraries.

Escherichia coli (*E. coli*) is the most commonly used host for expressing recombinant proteins. It can grow rapidly to a high density in inexpensive media, a large number of cloning vectors are available and a variety of mutant *E. coli* strains have been optimised for expression. Despite the tools available, there is however, no guarantee that a specific protein will be expressed in *E. coli*. Potential problems can be that the protein is toxic to the host, the growth rate of the host is reduced or that post-transcriptional or translational modification or protein folding results in a product that does not resemble the native protein (Baneyx, 1999).

Expression libraries consist of cDNA made from total mRNA, usually cloned in an *E. coli* vector, that enables the inserts to be expressed as fusion proteins. Both plasmid and bacteriophage vectors can be used. The resulting bacterial colonies or phages expressing the proteins can then be screened by a functional or immunological assay (Helfman *et al.*, 1983; Young and Davis, 1983). The proteins are blotted onto filters and allowed to react with specific antibodies which are in turn recognised by an enzyme-conjugated secondary antibody, which can be detected by the enzymatic catalysis of a suitable substrate. The filters are used as a guide to locate the corresponding phages or colonies on the culture plates. Clones, such as some derived from *Cowdria*, are unstable in plasmid vectors. In bacteriophage vectors, the phage together with the expressed proteins are released from the bacteria and the foreign proteins seems to have a less deleterious effect on the bacteria, making the clones more stable (Brayton *et al.*, 1999). In both these systems, clones represented at a frequency lower than one in 10^5 in the library are easily missed since screening more than 10^6 colonies or plaques is not normally practical (Young and Davis, 1983; Devlin *et al.*, 1990). In addition, a large amount of antibody is required. Expression libraries made from eukaryotic cDNA enable only the protein encoding regions to be cloned, thus eliminating introns

and other non-coding stretches of the genome. An advantage of prokaryotic bacteria is that there are no introns which can disrupt reading frames, but bacterial mRNAs have no convenient poly A "tail" to facilitate the isolation of mRNA for subsequent cDNA production. With prokaryotes, genomic DNA can be digested by restriction enzymes and the resulting fragments cloned and expressed. With this approach, the DNA is fragmented randomly. Consequently, there is no way to ensure that only the protein encoding regions are cloned.

1.2.3. Phage display.

Phage-display technology has been aptly described as 'finding a needle in a vast molecular haystack' (Rodi and Makowski, 1999). The display of peptides and proteins on filamentous bacteriophages has created new possibilities for epitope mapping, identifying genes and generating new antigens for diagnostic tests (Parmley and Smith, 1988; Markland *et al.*, 1991).

1.2.3.1. Filamentous bacteriophage.

The filamentous bacteriophages f1, M13 and fd are all essentially the same, differing in only a few nucleotide substitutions. They have a single stranded (ss) circular DNA genome surrounded by a flexible tube consisting of several thousand copies of the major coat protein pVIII, forming a particle of approximately 6.5 nm in diameter and 930 nm long (Webster, 2001). Four minor coat proteins are located on the tips of the virion, pIII and pVI at one end and pVII and pIX at the other, all at approximately five molecules per phage particle (Figure 1.1). In addition, the genome codes for proteins required for replication, morphogenetic markers and transcriptional terminators.

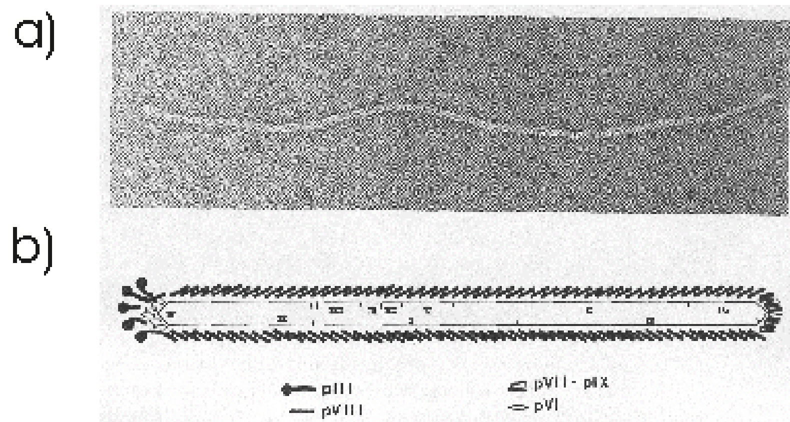


Figure 1.1. The Ff bacteriophage. Electron micrograph of a negatively stained Ff bacteriophage with a schematic representation below of the bacteriophage particle, showing the position of the capsid proteins (taken from Webster, 1996).

Being male specific, these phages only infect bacteria containing the F^+ episome that codes for the F pilus. Infection begins when pIII attaches to the tip of the F pilus (Figure 1.2). The ssDNA enters the cytoplasm and is converted into a double stranded replicative form (RF) by bacterial enzymes. The RF is maintained as a 'plasmid' at twenty to forty copies per host cell and serves as transcriptional template for the synthesis of phage encoded proteins. It also acts as a template for progeny ssDNA which is synthesized by the host's enzymes. Phage protein pII nicks one strand of the RF and the resulting free 3'-hydroxyl end acts as primer for the enzymes to make ssDNA by a 'rolling circle' mode of replication (Webster, 2001). When pV reaches an optimal concentration, it forms a complex with new ssDNA which is extruded through the inner membrane where it acquires the coat proteins to form mature phages. Each coat protein has a signal peptide that enables translocation through the bacterial membrane where phage particles are assembled and exported into the periplasmic space. This process is not lethal to the host which continues to multiply at a slightly slower rate. Phages are continuously exported from the host at a rate of 100 to 200 per hour (Scott, 1992; Jefferies, 1998; Cwirla *et al.*, 1990; Smith, http://www.biosci.missouri.edu/smithgp/index.
html; Dübel, http://www.mgen.uni-heidelberg.de/SD/SDscFvSite.html).

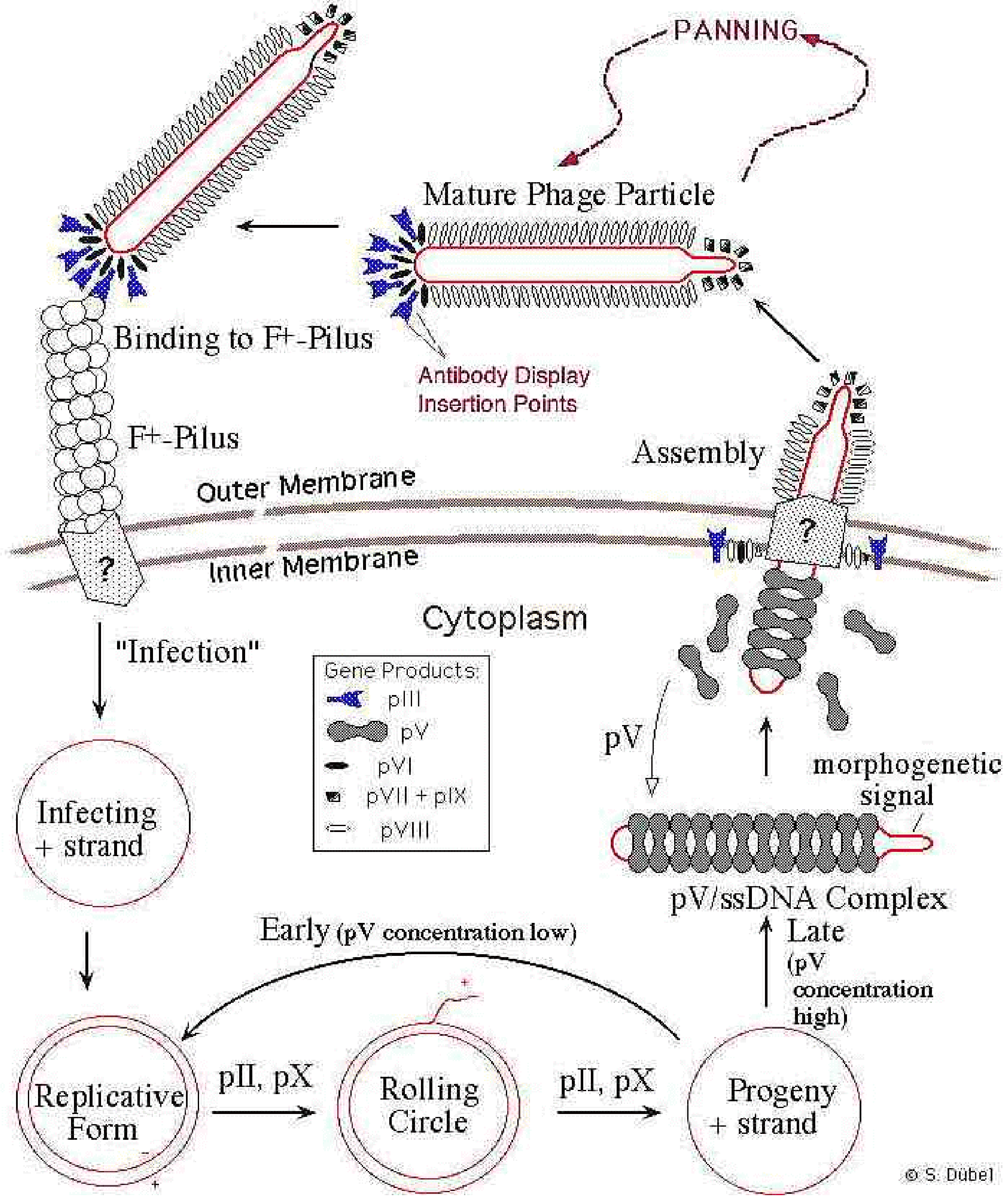


Figure 1.2. Life cycle of filamentous bacteriophages (Dübel web page: http://www. mgen.uni-heidelberg.de/SD/SDscFvSite.html).

1.2.3.2. The origin of phage display.

George Smith (1985) was the first to show that a foreign sequence can be inserted into pIII without disrupting its function. Approximately five molecules of this 42 kDa protein are attached to the tip of the virion (Figure 1.1; Webster, 1996). The protein contains glycine-rich regions resulting in an apparent molecular weight of 60-65 kDa as determined by SDS-PAGE (van Wezenbeek *et al.*, 1980). The C-terminal domain of pIII is involved in the assembly of phage particles while the N-terminal domain binds to the F pilus of the host, thereby initiating the infection process. Foreign DNA is cloned into the phage gene III, such that the pIII open reading frame remains intact. The coat protein which is essential for production of phage particles and *E. coli* infection, is still produced. The foreign sequence displayed as a pIII-fusion protein is in an immunologically accessible form on the surface of the infectious particle.

As a model, restriction fragments of the genes encoding the *EcoR* I endonuclease and methylase enzymes were cloned in between the C- and N-terminal domains of the pIII of phage f1. It was shown that by affinity selection or ‘panning’ (see 1.2.3.3.) with antibodies against *EcoR* I that fusion-phages could be isolated from a large mixture of non-fusion phages. Many improvements were made to the series of vectors (named fUSE vectors) which now form the basis of many of the cloning vehicles used in phage display technology (Parmley and Smith, 1988). The parent phage fd was used instead of f1 and the tetracycline resistance gene was incorporated into the genome to enable antibiotic selection of recombinant phages. These phages do not kill the host and can be propagated as colonies rather than plaques. The site for inserting foreign DNA was moved two to three amino acids downstream from the signal peptide in the N-terminal end, keeping pIII more ‘intact’ and apparently causing the foreign inserts to have a smaller effect on the function of pIII. Inserts of up to 335 bp (base pairs) were cloned into gene-III. Evidence of proteolytic breakdown of the recombinant protein was seen by western blot but this did not prevent the clones from being

affinity-selected. Parmley and Smith (1988) suggested that not only the size, but also the nature of the sequence will affect pIII function and this has indeed been shown experimentally (see 1.2.3.5). Instead of using the phage genome, the fusion-peptide can also be encoded by a plasmid (or phagemid) that contains a phage derived origin of replication enabling the 'plasmid' to be packaged into the phage particle. The wild type phage proteins are supplied by helper phage to produce complete phage particles containing the phagemid as genomic material (Scott and Barbas, 2000). Phagemids have the advantages of easily obtainable double stranded plasmid DNA for cloning and large DNA inserts are more stable in phagemids that phage genomes. The disadvantage is that a helper phage 'rescue' step is required for phage particle production.

In their 1988 paper, Parmley and Smith also suggested the concept of an epitope library; i.e. cloning short, synthetic random oligonucleotides would create a 'universal' reagent consisting of a repertoire of random phage displayed peptides. Panning this repertoire with a specific antibody would then give information about the epitope recognised by the antibody without knowledge of the natural antigen. This idea was later implemented by Scott and Smith (1990), Cwirla *et al.* (1990) and Devlin *et al.* (1990; see 1.2.3.4.). The work of Parmley and Smith (1988) truly laid the foundation for the general acceptance of phage display technology which has since been extended to the display of complete proteins and single chain antibody fragments. Ligands other than antibodies have also been used for the selection of peptides. The full impact of their work is now being realised.

1.2.3.3. Affinity selection.

A phage expressing a peptide that reacts with a specific ligand can be isolated from a phage display library by affinity selection or panning (Figure 1.3). Smith (1985) showed that this is a highly selective procedure. The amount of antibody required is much less than with the filter-based colony-

or plaque-screening methods. In the most used system, a ligand (often an antibody) is immobilised on a plastic surface by direct binding (Smith, 1985; Wang *et al.*, 1995). The phage library is then exposed to the affinity matrix. Non-binders are washed away and any bound phages released from the ligand are amplified in *E. coli* host cells. The resulting pool of amplified phages is usually subjected to further rounds of panning (Figure 1.3). Even if a phage with low yield is selected, the amplification steps ensure detection of the clones (Parmley and Smith, 1988; Devlin *et al.*, 1990). This iterative selection method is so powerful that one clone from a mixture of 10^9 clones can readily be isolated (Parmley and Smith, 1988). Since the genotype and phenotype are linked, once the DNA sequence of the insert is determined, the amino acid sequence of the peptide can be deduced. There are several variations on this basic principle. For example, to improve the efficiency of panning, the strong binding of biotin to streptavidin can be exploited (Parmley and Smith, 1988). Biotinylated antibody was allowed to react with phages in solution and the phage-antibody complexes were trapped with streptavidin immobilised on the plastic surface. Another approach was adopted by Cortese *et al.* (1996), in which anti-human Fc-specific antibodies were attached to magnetic beads that trapped the phage-IgG complexes.

The panning method used depends largely on the information required from the experiment and on the nature of the target molecule. In a phage library pool, there are non-binders, non-specific binders which bind to the plastic surface or the blocking agent and then, finally, a minuscule number of phages that bind to the target molecule either weakly or strongly. Repeated washing with a high detergent concentration (0.5 % Tween; Parmley and Smith, 1988) usually removes the non-binders,

while high salt concentrations (150 mM) can further decrease background binding (Menendez *et al.*, 2001; Liu and Marks, 2000). Phages that bind to the plastic surface or to the blocking agent can be minimised by pre-absorbing the phage pool with all the components except the target molecule

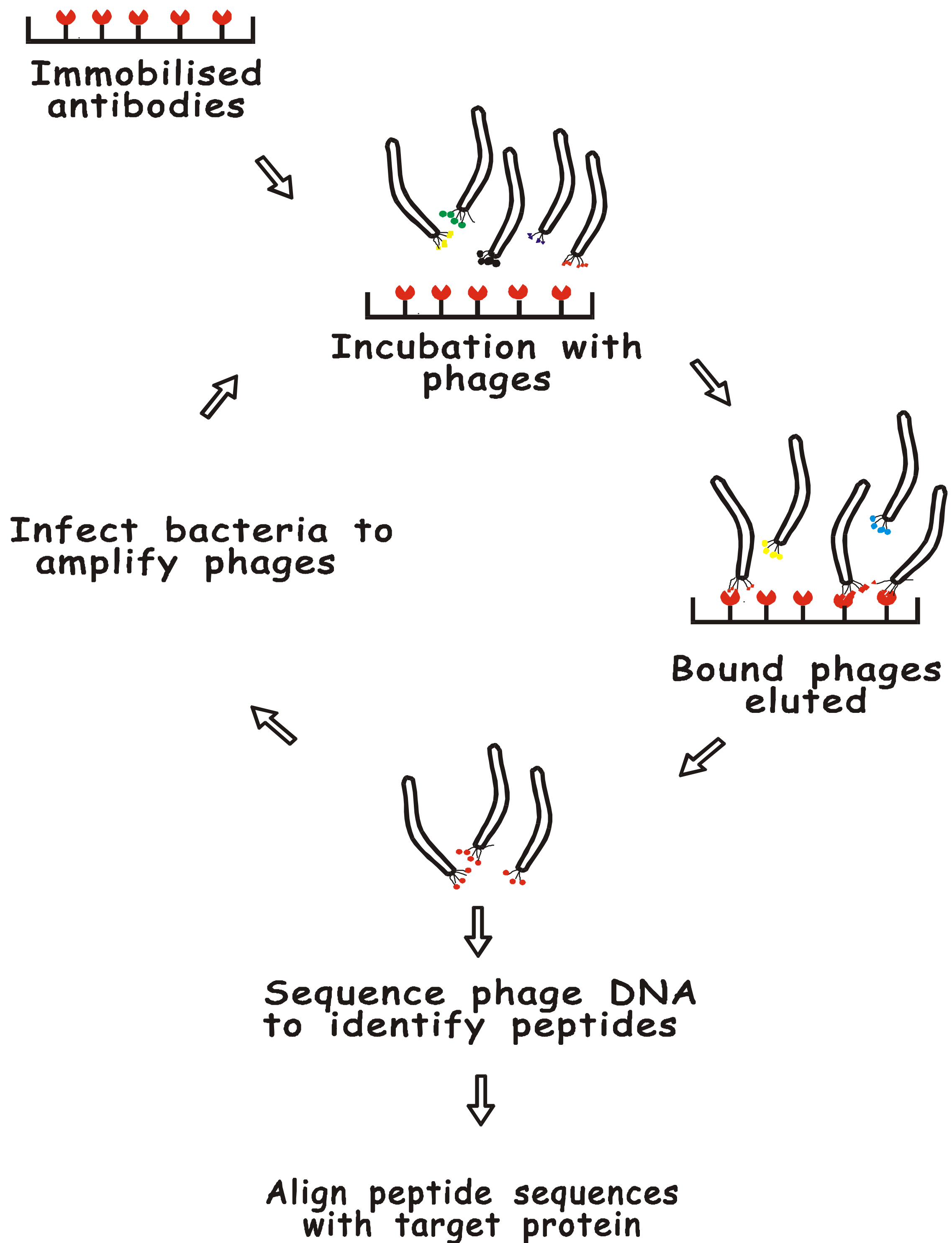


Figure
1.3. Schematic representation of affinity selection (panning) of bacteriophages. Antibodies are immobilised on a solid surface, a mixture of different phages are allowed to react with the antibodies, non-binders are washed away before binders are released. The phages are transfected into *E. coli* and the resulting pool of amplified phages subjected to another round of selection.

before panning. If the aim is to find the strongest binder, stringent conditions must be used, but if the aim is to obtain sequence diversity, less stringent conditions are appropriate. This is usually the case if one expects more than one peptide binding site on the target or if the target is a mixture of molecules, such as polyclonal antibodies. It is thought that by immobilising the target directly onto the surface, the phage will attach multivalently since there are multiple copies of the expressed peptide on the phage. This will theoretically increase the avidity and enable the weak but relevant binders to be isolated (Sparks *et al.,* 1996; Menendez *et al.*, 2001). To increase the surface area and thus the target concentration, Sepharose or magnetic beads can be used (Sparks *et al.,* 1996). Another approach is to allow the target and phage to react in solution. The complex is then captured in subsequent steps. In this case, if the target is monovalent, multivalent binding will be avoided and high affinity binders will be selected (Sparks *et al.,* 1996). Scott and Smith (1990) also suggested that by decreasing the target concentration in successive panning rounds one can 'force' the selection of high affinity binding peptides. The multiple copies of the expressed peptide on a single phage particle will bind to two or more neighbouring target molecules and may dissociate from the solid surface slowly, even if the monovalent affinity is modest. As the density of the target is decreased, this 'avidity effect' is reduced, monovalent attachment predominates and selection of high affinity binders occurs (Smith, http://www.biosci.missouri.edu/smithgp/index.html). Usually when starting an experiment, it is best to follow the conservative approach by solid-phase panning and, if binders are obtained, selections can be optimised with in-solution panning and target concentration (Menendez *et al.*, 2001). An advantage of panning as opposed to screening a λ library is the small amount of antibody usually required. For instance, Parmley and Smith (1988) and later du Plessis *et al.* (1995) showed that enough antibody can be affinity-purified from an immunoblot for panning purposes.

1.2.3.4. Random peptide libraries.

Determining the actual amino acids (epitope) which bind to an antibody is of cardinal interest in the understanding of molecular recognition. Hence, phage libraries can replace the more expensive chemical synthesis method (see 1.1.4). Large epitope libraries have been made by cloning short random DNA sequences, resulting in the expression of millions of disparate peptides on phage surfaces (Scott and Smith, 1990; Cwirla *et al.*, 1990; Devlin *et al.*, 1990; Scott,1992). Together with the efficiency of affinity selection, this is a potentially powerful and inexpensive approach to map epitopes on proteins (du Plessis *et al.*, 1994; Fack *et al.,* 1997). Libraries expressing random peptides from six amino acids, the average size of a continuous epitope (van Regenmortel, 1988), up to thirty eight amino acids have been successfully used for mapping antigenic determinants (Scott and Smith,1990; Devlin *et al.*, 1990; Scott, 1992; Cwirla *et al.*, 1990; Stephen *et al.*, 1995; Bonnycastle *et al.*, 1996; Kay *et al.*, 1993). The longer peptide sequences in effect enlarge the size of the library since each peptide potentially embraces several short epitopes (Bonnycastle *et al.*, 1996). After several rounds of panning, individual phages are usually analysed by ELISA, western blot or some other binding or competition assay to confirm binding. Consensus sequences are identified by sequencing the insert DNA. Alignment of the consensus or motifs with that of the native protein sequence allows the epitope recognized by the corresponding antibody or ligand to be identified (Scott, 1992; du Plessis *et al.*, 1994).

As mentioned earlier, the first papers describing large random peptide libraries were by Scott and Smith (1990), Devlin *et al.* (1990) and Cwirla *et al.* (1990). These three libraries theoretically expressed 4 x 10^7 different hexapeptides, 2 x 10^7 different fifteen residue peptides and 3 x 10^8 different hexapeptides on pIII respectively. Scott and Smith (1990) used two well characterised monoclonal antibodies against myohaemerythrin to select binding sequences from their peptide

library. Sequences similar to the native epitope were selected as well as one with no obvious similarity, but which also bound strongly to the antibody. This was clearly a mimic of the native antigenic region. Devlin *et al.* (1990) investigated whether the phage-displayed peptides could bind to ligands other than antibodies and selected streptavidin binders. The vector used by Cwirla *et al.* (1990) was manipulated to contain a frame shift that resulted in no phage replication. When the stuffer fragment was removed and the new insert (random oligonucleotide) cloned, the restored reading frame allowed phage replication. This means that no wild type phages were included in their library. When a monoclonal antibody against β-endorphin was used to pan the library, specific ligands were isolated without any prior structural information of the native protein. A phagemid system expressing random octapeptides as pIII fusions was used to map the epitope of the C5a neutralizing monoclonal antibody (Kola *et al.*, 1996). The fusion-peptide was encoded by a phagemid. The assumption was that on the rescued phages, one copy of pIII was recombinant while the remainder were wild type, thereby resulting in monovalent expression of the peptides. Thus polyvalent attachment by the phages would be reduced and high affinity binders should be isolated. The epitope was located, but high affinity binders were not found. Library size, and thus its diversity, plays an important role in determining the statistical chance of finding high affinity binders which are very rare. (Kola *et al.,* 1996).

1.2.3.5. Other phage display systems.

All the abovementioned work utilised pIII to display fusion peptides on the phage surface. Other phage proteins such as pVIII and pVI have also been used, the former most often. The bacteriophage major coat protein (pVIII) is fifty amino acids long. There are about three thousand copies per phage particle (Markland *et al.*, 1991). Constructing peptide libraries as pVIII-fusion proteins has the advantage that since pIII is not altered, there is no potential interference with

infectivity. As well as phages in which all pVIII molecules are expressed as fusion proteins, a phagemid system can be used in which case only thirty to sixty copies of the fusion protein are expressed per particle (Markland *et al.*, 1991). This multivalent display allows better capture of weak binders as the presence of more copies of the epitopes enables a 'tighter bind' in comparison to pIII display (Wells, 1996). Small peptides of six amino acids can be expressed on all copies of pVIII (Greenwood *et al.*, 1991), but bigger peptides must be interspersed with wild-type proteins in hybrid virions. The wild-type proteins are either supplied by a helper phage if a phagemid system is used or the foreign sequences are cloned into a second, synthetic copy of pVIII (Malik and Perham, 1996; Bonnycastle *et al.*, 1996). It was generally thought that small peptides would be present on the phage in a higher copy number than larger peptides. In a study to investigate the size of the insert and its effects on the phage particle, the foreign sequence itself proved to be a contributing factor. Some larger peptides (sixteen amino acids) could be expressed at higher levels than hexapeptides (Malik *et al.*, 1996). The insert also had an effect on the rate of processing of the pVIII pro-coat which then determined how efficiently it would be incorporated into the mature phage particle. Malik *et al.* (1996) predicted that it should in fact be possible to express large proteins (100kDa) on pVIII. This is indeed possible: the gene coding for a penicillin G acylase (86 kDa) has been expressed as both pIII and pVIII fusions displayed on the surface of the phage. Both these fusion proteins showed enzyme activity similar to the soluble native protein (Verhaert *et al.*, 1999).

The structural framework within which a random peptide is presented is important for optimal binding and can influence the sequences selected. Bonnycastle *et al.* (1996) therefore introduced structural constraints into the peptides they displayed on phages. Sequence-randomized peptide libraries (from six to seventeen amino acids long) were constructed with fixed cysteine residues at certain positions enabling the peptides to form single or multiple disulphide-bridged loops. The

peptides were displayed as pVIII-fusions using a vector that encoded both the wild type and a synthetic pVIII. Thus, only a small percentage of the pVIIIs were expressing the peptides. The libraries were screened with a variety of monoclonal and polyclonal antibodies against peptides, proteins and carbohydrates. No one library performed 'best' and different libraries gave different consensus sequences. The varied peptide sequences selected from the different libraries could of course also reflect the polyspecificity of some antibodies. Comparing the results obtained from screening a panel of libraries with a particular antibody can yield information on its structural specificity; most antibodies directed against peptides, carbohydrates and linear epitopes on folded proteins recognised peptides displayed on phages, whereas antibodies against discontinuous epitopes on proteins often do not. This could be ascribed to the fact that antibodies bind peptides in a groove or cleft, whereas they mainly bind discontinuous epitopes on folded proteins over a large shallow surface (Bonnycastle *et al.*, 1996).

Jespers *et al.* (1995) investigated a phagemid system that could display peptides on pVI, a twelve kDa protein. Like pIII, there are about five copies of this minor coat protein present per phage particle. The peptides were fused at the C-terminus of the protein. Bovine pancreas trypsin inhibitor was used as a model fusion gene with a glycine-rich sequence at the fusion to allow conformational flexibility. This clone was compared to the same gene cloned as an N-terminal fusion protein with pIII. The pIII-fusion proteins were incorporated into mature phage particles to a much greater extent than the pVI-fusion proteins. They suggest that the foreign sequences may interfere with the incorporation of pVI during phage morphogenesis and this protein is for this reason not a good choice for displaying foreign sequences.

A system where the attachment of a leucine zipper pair connected phenotype and genotype instead of utilising a fusion to the N-terminus of the phage proteins was developed by Light *et al.* (1996). One part of the zipper was fused to pIII and the second part to the protein/peptide to be expressed

(see Figure 1.4). The proteins were exported to the periplasmic space where they associated and were incorporated in the phage particle. Light *et al.* (1996) used phage display to select an 'ideal' leucine zipper pair and then incorporated the pair into a random peptide library (also see 1.2.3.8.).

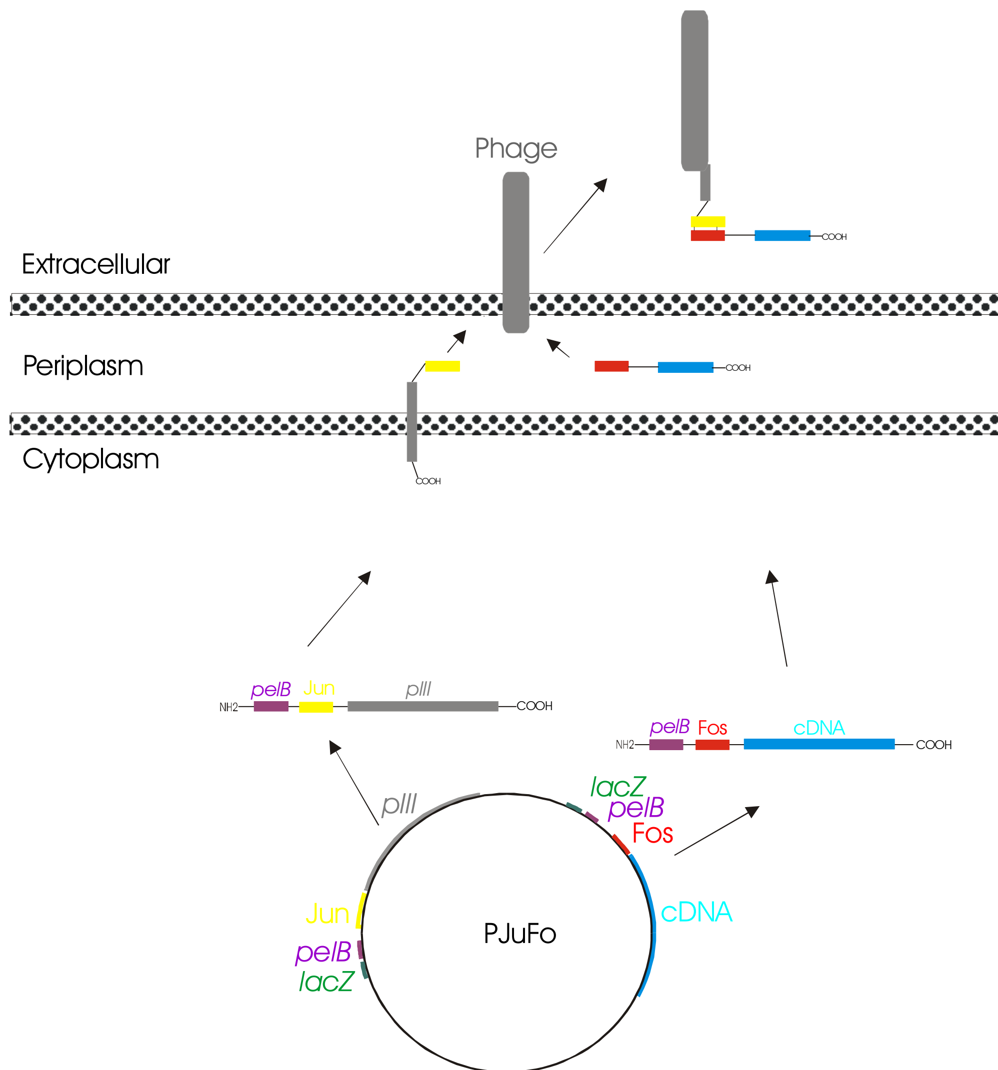


Figure 1.4. Schematic representation of the pJuFo system: a leucine zipper approach to link genotype and phenotype. Based on figure from Crameri and Suter (1993).

Maruyama *et al.* (1994) developed surface expression vectors based on bacteriophage lambda (λ). The advantage of this system is that the coat proteins are assembled in the cytoplasm whereas the fusion proteins of filamentous phages have to translocate across the plasma membrane. Using this

λ system, Kuwabara *et al.* (1999) mapped the minimal binding domain of inhibitor alloantibodies within the human fVIII protein. This antigenic determinant was discontinuous, since binding was abolished after treatment that disrupted the secondary structure of the protein.

In general, the choice of vector and type of fusion protein used will be determined by the study to be undertaken. Multivalent display of peptides by pVIII allows better capture of weak binders in comparison to pIII display. The leucine zipper approach avoids C-terminal fusions and is useful for expressing cDNA clones which contain stop codons. The λ system shows promise for identifying discontinuous epitopes.

1.2.3.6. Epitope mapping with random display libraries.

There are many examples where random peptide libraries were used to determine the binding sites of both monoclonal and polyclonal antibodies, as well as other ligands. In early studies, the epitopes of two monoclonal antibodies directed against the viral capsid protein 7 (VP7) of bluetongue virus (BTV) were mapped using random peptide libraries and overlapping synthetic peptides. The sequences obtained from the two different methods correlated well and the epitopes could be precisely located on the protein (du Plessis *et al.*, 1994). A motif that mimics the epitope recognised by a neutralising monoclonal antibody against gp120 of the human immuno-deficiency virus (HIV-1) was identified by panning constrained random peptide libraries expressing nine and fifteen amino acid peptides. Peptide-specific antibodies were made in rabbits that reacted with the native gp120 and neutralised HIV-1 infection *in vitro*. This was not achieved with antibodies raised in mice (Lundin *et al.*, 1996). The binding site of a monoclonal antibody against a peptide derived from the surface glycoprotein of feline immuno deficiency virus was mapped with the use of a hexapeptide phage display library (Sibille and Strosbreg, 1997).

The antigenic determinants of two neutralizing antibodies of *Theileria annulata* could be identified by using a phage library displaying peptides of twelve amino acids (Prickett and Hall, 2000). To characterize peptides binding to HLA-DR13 MHC class II molecules, a random peptide library was panned on purified MHC molecules. The peptides identified by phage display had motifs in common with naturally processed peptides that were eluted directly from the MHC molecules (Davenport *et al.*, 1996).

By panning random peptide libraries with polyclonal antibodies, disease-related epitopes can be identified without knowledge of the aetiological agent (Folgori *et al.,* 1994; Kouzmitcheva *et al*., 2001). Random peptide libraries expressing nonapeptide pVIII-fusions were made with two cysteines flanking the inserts. Luzzago *et al.* (1993) proposed that the constrained environment of the peptides would strengthen their binding compared to flexible linear peptides. Mimics of human hepatitis B virus envelope protein were identified which in turn reacted back with sera from many infected individuals, showing the potential of the mimics for use as diagnostic reagents. In addition, antibodies made against the phages bearing these mimotopes reacted with the native envelope protein. Random peptide libraries were also used to identify specific ligands for antibodies that were concentrated in the cerebrospinal fluid of multiple sclerosis patients. Cortese *et al.* (1996) found that these selected peptides mimicked natural epitopes and that the antibodies in the fluid were individual-specific and were due to nonspecific immunodysregulation. They did therefore not result from a reaction to a specific disease-related antigen. Sera from patients with Lyme disease were used to pan twelve large random epitope libraries. Seventeen peptides with eight common motifs were identified, none of which could be matched to proteins from the disease causing pathogen, *Borrelia burgdorferi*. These mimotopes have the potential to be used in a diagnostic ELISA which should be cheaper than existing assays (Kouzmitcheva *et al*., 2001).

A DNA-based selection system was used to find ligands from random peptide libraries that were associated with hepatitis C virus (HCV) infections in humans. To simplify the screening of many sera and to compare the results obtained from sera of infected and normal individuals a DNA hybridisation step was incorporated. The peptide-encoding DNA inserts derived from panning 'A' were biotinylated and made single stranded to act as capture probes. From panning 'B', single stranded phage DNA was isolated and allowed to react with the 'A' probe. If common sequences were selected in 'A' and 'B', they hybridised and the biotin on probe 'A' was used to trap the ssDNA which was then electroporated into *E. coli.* If the unique sequences were required, the ssDNA molecules that did not hybridise to the probe were electroporated into *E. coli*. In this manner pools from different pannings were 'compared' (Bartoli *et al.*, 1998). These selected peptides were affinity-matured with random mutagenic oligonucleotides and after more selections, peptides with an increased specific reactivity with the serum were identified. These peptides could discriminate between different components of the polyclonal HCV response in serum. This approach can be used to develop better diagnostic reagents without knowledge of the natural antigens (Urbanelli *et al.*, 2000).

Phage display has been used to analyse protein-protein interactions. Peptides with similar motifs that bind to streptavidin were found by two independent studies (Devlin *et al.*, 1990; Kay *et al.*,1993). Peptides have been selected that mimic the binding region of Taxol. Sequence similarity analysis enabled the identification of a human protein, Bcl-2, that contains the peptide motifs and is also a Taxol binder (Rodi *et al.*, 1998).

1.2.3.7. Fragmented-gene phage display.

Following an approach proposed by Smith (1985), Wang *et al.* (1995) in a joint South African-

Australian collaboration, showed that epitopes could be mapped on a BTV protein by cloning random DNA fragments of a gene of interest into a phage display vector. Since the sequence of the gene coding for the protein (VP5) was known, direct alignment of the selected peptide sequences to the gene sequence quickly identified the actual antigenic region. Wang and co-workers used DNase-generated fragments that were first made blunt ended. Linkers were then added to make the ends compatible with the cloning site of the vector. In a similar study, Blüthner *et al*. (1996) used direct blunt-ended cloning to construct a similar fragmented-gene library. Both these libraries were made as pIII-fusions. They did not have the diversity and universal application of random synthetic peptide libraries, but since the native peptides were represented, cognate sequences could be identified instead of mimotopes. This approach has been especially successful with viral epitopes. In addition to the initial work on VP5 of BTV (Wang *et al.*, 1995), two antigenic regions on NS1, a non-structural protein, were identified by this approach (du Plessis *et al.*, 1995). Polyclonal antibodies were used, but these had to first be affinity-purified from eletroblotted of NS1 to minimise non-specific selection. This study showed the potential for identifying peptides that could be developed into highly specific diagnostic assays (du Plessis *et al.*, 1995). In yet another study focussed on orbiviruses, epitopes with the potential to discriminate between different African horsesickness virus serotypes were identified by panning a phage library derived from the gene encoding the viral coat protein 2 (VP2). A combination of a monoclonal antibody, polyclonal chicken antibodies and horse serum was used (Bentley *et al.*, 2000). This work is significant since it showed that discontinuous epitopes could be found using a combination of display-based approaches. A genus-specific epitope of the major envelope glycoprotein (E2) of pestiviruses has been mapped with a monoclonal antibody, using a combination of fragmented-gene and random peptide phage display libraries, combined with analysis of overlapping synthetic peptides (Yu *et al.*, 1996). The complementary use of these techniques allowed fine mapping and confirmation of the position of the antigenic determinant. In an attempt to improve the efficiency of constructing

fragmented-gene libraries, Nagesha *et al.* (1996) incorporated a PCR step using oligonucleotide linkers as primers. Only fragments with linkers on both ends were amplified which resulted in more cloneable fragments and thus the cloning efficiency was increased. An additional potential advantage is that recombined DNA fragments can also be cloned, thereby creating greater diversity. This technical refinement might be useful in defining discontinuous epitopes, but has not been much used.

The concept of fragmented-gene phage display was shown to be applicable to proteins other than those of viral origin (Petersen *et al.,* 1995). With random DNA fragments (50-200 bp) cloned into a pIII-fusion display vector, monoclonal antibodies were used to map known epitopes on RNA polymerase, p53 and cytokeratin 19. After one round of panning an RNA polymerase-derived display library, positive clones in a colony blot assay all contained epitope sequences. In the case of the cytokeratin 19 library, two rounds were necessary to yield 10/48 positive clones by colony blot. One round was also adequate for the p53 library where 31/48 were positive. Petersen *et al.* (1995) found that ELISA was a more sensitive method than colony blots to confirm the binding of single clones. They also confirmed that phage display is a cheaper (and quicker) way to map antigenic determinants than using chemically synthesised peptides, especially for large proteins like RNA polymerase (215 kDa). When libraries with different linker sequences were constructed, it was shown that the linker-encoded amino acids can cause some restrictions on the optimal presentation of fusion peptides. Libraries made with linkers encoding alanines and glycines were in general more successful than those encoding cysteines which could make the peptides form loops through disulphide bridge formation. This was in contrast to the Bonnycastle *et al.* (1996) and Luzzago *et al.* (1993) findings with random peptide libraries where the conformational restrictions were important for optimal binding and could strengthen the interactions compared to flexible linear peptides.

Fragmented-gene libraries have been exploited to identify disease related epitopes. Autoantibodies associated with polymyositis / scleroderma (PM/Scl) overlap syndrome in humans were used to identify epitopes on the 100 kDa PM/Scl protein (Blüthner *et al.*, 1996). These polyclonal antibodies had to be affinity-purified from immunoblots before panning. Linear epitopes were identified and mapped to high resolution. In another study, affinity-purified antibodies were used to pan a fragmented-gene phage display library derived from the gene encoding a nuclear protein, sp100 (Blüthner *et al.*, 1999). Two antigenic regions of sixteen and twenty amino acids were identified. To avoid background clones being selected, it was essential to preabsorb the antiserum against *E. coli* proteins, since the antibodies were eluted from a recombinant protein. These antigenic regions were confirmed and narrowed down to six and nine amino acids using overlapping, synthetic peptides.

In a study aimed at developing a vaccine for malaria, a fragmented-gene library derived from the gene encoding the glutamate rich protein (GLURP) of *Plasmodium falciparum* was made using a pVIII phagemid vector. Affinity-purified human anti-GLURP IgG, which may play a role in development of clinical immunity to malaria, was used to select clones from the library. Since the peptides identified were able to block intra-erythrocytic parasite growth, they may be useful when designing a multi-component malarial vaccine (Theisen *et al.*, 2001).

Fack *et al.* (1997) compared epitope mapping of four monoclonal antibodies using fragmented-gene and random peptide phage display libraries. In this case, the fragmented-gene display approach yielded better results in that binders were selected after the first round of panning from the fragmented-gene library with all the antibodies they tested, while with the random peptide libraries, epitopes could be located with only two of their antibodies (Fack *et al.,* 1997). Ligands other than antibodies have also been used to select binding peptides from a fragmented-gene library. For

instance, a library derived from the gene encoding the multi-functional chicken oocyte receptor for yolk deposition was panned on the receptor associated protein allowing the minimal binding domain to be delineated (Bajari *et al.*, 1998).

1.2.3.8. Fragmented-genome and cDNA phage display.

The fragmented-gene phage display approach can be extended to genomes. Thus in theory, the resulting library will display different parts of all the proteins encoded by the genome of a specific organism. Using this approach, prokaryotic receptor genes have been isolated without prior knowledge of the receptors. A library derived from the genome of *Staphylococcus aureus* was made using 100-700 bp genomic DNA fragments expressed as pIII fusions and used to find ligand-binding domains (Jacobsson and Frykberg, 1995). Both known and novel sequences with binding domains for human IgG and fibronectin were isolated. The reading-frame entering the inserts was correct, but it was out of frame when entering the vector. This was surprising since intact fusion proteins (i.e. intact pIII) are required for phage production and infection. The authors proposed that ribosome slippage caused successful expression of the inserts. In an attempt to improve the system, a *Staphylococcus aureus*-derived library was made in a pVIII phagemid vector (Jacobsson and Frykberg, 1996). Stronger binders to IgG, fibronectin and fibrinogen were isolated. 'Wrong' i.e. out of frame clones were again found, but fewer than with the pIII phage library. The position of the apparent ribosome slippage differed between the pVIII and the pIII vectors. Jacobsson and Frykberg (1996) propose that the phenomenon of slippage keeps the proteins expressed at low levels that would otherwise be detrimental to the host cell.

Not only single viral genes, but whole viral genomes have been used to construct libraries for epitope identification. A phage display library derived from the HCV genome was panned with sera

from HCV positive patients, thereby enabling the identification of antigenic regions on three proteins. A combination of these peptides used in ELISA showed potential for use as diagnostic reagents (Pereboeva *et al.*, 2000).

Screening cDNA libraries in bacteriophage λ or plasmid vectors may not be successful since the gene products do not always assume their native three-dimensional structure. They can also be denatured when transferred to the solid support. Expression of cDNAs by phage display would overcome some of these problems and would have the added advantage of selection instead of the cumbersome screening of clones on solid supports. Of course, cDNAs have stop codons at the ends and cannot be used directly by gene fusion in phage display systems. To allow cloning of cDNAs, a leucine zipper approach (pJuFo, see Figure 1.4.) was developed that covalently links the cDNA to the phage surface by a leucine zipper pair. The Jun portion was fused to pIII and the gene of interest fused to the Fos portion of the leucine zipper. Both are encoded on the same plasmid under a *lac* promoter and contain a *pelB* signal sequence for processing to the periplasmic space. To test the system, the *E. coli phoA* gene was expressed and the phage-displayed enzyme was shown to be active (Crameri and Suter, 1993). Next, a cDNA library derived from *Aspergillus fumigatus* was made in the pJuFo display system. *Aspergillus fumigatus* is known to produce allergenic proteins recognised by IgE from sensitised patients. Phages encoding allergenic proteins were selected from the library with IgE antibodies (Crameri *et al.*, 1994). The leucine zipper approach was also used to construct an *E. coli* fragmented-genome phage displayed library. The *E. coli* inserts were fused to the 3' end of the *fos* gene and the *jun* zipper partner was fused to pIII (Palzkill *et al.*, 1998). By making a 3' fusion, only one end of the gene needed to be in the correct reading frame. This meant that one in six clones were meaningful clones, instead of one in eighteen as with pIII 5' fusions. To show the functionality of the system, epitopes on the Rec-A protein were mapped by panning the library with polyclonal antibodies against the protein. This universal tool can now be used to search

for any binding interaction between virtually any *E. coli* protein and any ligand, such as DNA, RNA, proteins or carbohydrates (Palzkill *et al.*, 1998).

Phage display was used to 'create' native-like proteins by combinatorial segment assembly from nonhomologous proteins, showing that novel domains can be generated *in vitro* (Reichmann and Winter, 2000). The coding sequence for the N-terminal half of the β-barrel domain of the heat shock protein, CspA, was fused to fragments of the *E. coli* genome and cloned into a pIII phagemid vector. The aim was to identify novel protease resistant domains. The rescued phages were therefore selected for their ability to survive proteolytic treatment. Of the resistant domains isolated, some were formed from portions of the original gene (CspA) whilst others were from unrelated sequences. Novel protein domains had therefore been generated (Reichmann and Winter, 2000).

To take advantage of the superior selection methods of phage display, an existing cDNA library of the parasite *Ancylostoma caninum* was subcloned into a phagemid system, generating fusion proteins with pVI. Two serine proteases (trypsin and factor Xa) were used as ligands for panning. Factor Xa binding clones identified by panning were presented in 0.01% of the original cDNA library and could easily be missed with other screening methods (Jespers *et al.,* 1995).

A great challenge following the human genome project is to ascribe functions to the many proteins defined by open reading frames. The function of many genes cannot simply be deduced from sequence similarity and therefore biochemical methods are required (Palzkill *et al.*, 1998). A method called reverse identification of transcriptional effectors (RITE) employs cDNA phage display to directly measure interaction of proteins with a known component of the transcriptional machinery (Mazzarelli and Ricciardi, 2001). Novel sequences were found with RITE proving better than the classical two-hybrid-yeast system.

1.2.3.9. Phage display: a tool for vaccine development.

Identifying antigenic mimics that can induce an immune response which resembles that induced by native proteins would be very useful in developing new vaccines. Immunogenic peptides could be designed without prior knowledge of the amino acid sequence of the target antigen or even information about the aetiological agent. In a study aimed at identifying disease specific epitopes, a random peptide library was panning with immune sera from patients suffering from hepatitis (Folgori *et al.,* 1994). Mimotopes of the human hepatitis B virus (HBV) envelope protein were obtained and could be used as diagnostic reagents, since they reacted with serum from many infected individuals. When the mimotopes were tested for their immunogenicity, the phage-displayed peptides proved very efficient in inducing antibodies that reacted with the native envelope protein (Meola *et al.*, 1995).

Peptides are generally poor immunogens and usually need to be coupled to protein carriers to stimulate an adequate humoral immune response (van Regenmortel, 1988). In contrast, phages are highly immunogenic. This is an obvious advantage over bacterial expression vectors (Jefferies, 1998). Phages can in fact recruit T-cell help and elicit B-cell antibodies without the need for adjuvants. Peptide repeats of the *Plasmodium falciparum* circumsporozoite protein have been expressed on phages, the NANP and NVDP repeats as pIII fusions in a phage f1 based phage vector (de la Cruz *et al.*, 1988) and NANP and NDVP as pVIII-fusions in a phagemid system (Greenwood *et al.*, 1991). Both pIII and pVIII type fusion-phages induced a strong antibody response to the peptide sequences (de la Cruz *et al.*, 1988; Greenwood *et al.*, 1991). In another study, mimotopes of the NANP repeat were selected from a hexapeptide library with a monoclonal antibody directed against the repeat. One of the mimotope peptides was synthesized and used to immunise mice. The resulting antibodies recognized *Plasmodium falciparum* sporozoites (Stoute *et al.,* 1995). Zhong

et al. (1994) investigated a chlamydial epitope known to react with a neutralizing monoclonal antibody. Random conformational constraints were introduced into this determinant before it was displayed on phages. The aim was to empirically make it 'immunogenically fitter' for use in a vaccine. Peptides were selected with a conformation-sensitive monoclonal antibody. Phages displaying the selected peptides were used to immunise mice and the resulting antibodies reacted with the native protein. This peptide is a candidate for a peptide-based vaccine (Zhong *et al.*, 1994). Recently, Matthews *et al.* (2002) showed that peptides selected from 'natural peptide libraries' (ie. gene- or genome-derived libraries) were 'immunogenically fitter' than those selected from random peptide libraries.

Since phages are good immunogens, a number of pVIII-based vectors have been developed for use in vaccine studies (Malik and Perham, 1997). A system whereby hybrid virions expressing a mixture of two foreign peptides interspersed with wild type pVIII was devised. For vaccine purposes, two different epitopes of a pathogen can be expressed, or alternatively effector and targeting sequences can be expressed simultaneously (Malik and Perham, 1997). When an MHC class I binding peptide of HBV surface antigen (HBs_{28-39}) was expressed as a pVIII fusion and mice immunised without adjuvant, a specific CTL response to HBs_{28-39} was detected (Wan *et al.*, 2001). Mice immunised with the recombinant phages displaying a peptide (RT2) of the reverse transcriptase of HIV-1 together with a T-helper epitope, led to the activation of an effective, specific anti-HIV-RT2 CTL response (de Berardinis *et al.*, 2000). Craig *et al.* (1998) using polyclonal antibodies to screen random peptide libraries, identified linear epitopes that structurally mimicked the antigenic determinants on folded proteins. The structure was analysed by comparing the linear peptide and the epitope on the folded protein. Critical binding areas were identified and selected peptides were found to have some conformational preferences which allowed them to functionally mimic epitope structures (Craig *et al.,* 1998). These studies show that phage-displayed peptides are

indeed suitable tools for vaccine development, both for identifying epitopes or mimotopes able to induce the appropriate immune response (Folgori *et al.,* 1994; Meola *et al.*, 1995) and as a vaccine delivery system (Wan *et al.*, 2001; de Berardinis *et al.*, 2000).

1.2.3.10. Antibodies expressed on phage.

Antibodies play a major role in research and diagnostics. Polyclonal antibodies can be raised in laboratory animals while monoclonal antibodies can be made by hybridoma technology. Both approaches require the use of experimental animals and are limited by several biological constraints. Displaying single chain antibody fragments on phages can greatly reduce the use of animals (Burton, 1995) since many different antibodies can be selected from one donor animal. Large repertoires of antibody fragments have been expressed as phage display libraries. These are derived either from naïve (e.g. Marks *et al.*, 1991; Vaughan *et al.*, 1996) or immune donors (e.g. Clackson *et al.*, 1991). They can be made semi-synthetically (Hoogenboom and Winter, 1992; Nissim *et al.*, 1994) or entirely synthetically (Knappik *et al.*, 2000). In most such libraries, antibody variable light and variable heavy regions are joined by a synthetic flexible linker. Most of these libraries are made in phagemid vectors. This technology makes it possible to obtain human 'monoclonal' antibodies and to genetically manipulate the antibodies. 'Immune' libraries have the advantage that the library need not be as large. Naive libraries, however, display a representation of antibodies which have not come into contact with antigen and have therefore not undergone somatic mutation. Semi-synthetic libraries utilise known antibody frameworks with the CDRs (usually CDR3 of the heavy chain) randomised by inserting degenerate oligonucleotides. These libraries can be regarded as 'universal reagents' and have to be extremely large (about 10^9 to 10^{10} different clones) to provide a statistical probability of obtaining high affinity antibodies directed against virtually any antigen (Viti *et al.*, 2000). Useful single chain antibodies reactive with a vast

array of disparate antigens such as chemokine receptors (Osbourn, *et al.*, 1998), human-CD30 (Klimka *et al.*, 2000), tumour cell surface antigen (Pereira *et al.*, 1997), virus glycoproteins (Sanna *et al.*, 1995) and bacterial proteins of *Streptococcus suis* (de Greef *et al.*, 2000) have been selected from universal antibody libraries.

1.2.4. Genomics.

The DNA sequence of the genomes of many organisms is becoming available at an unprecedented rate. This makes it possible to predict the role of putative open reading frames (ORFs). The first step is to do homology-based assays, but about forty precent of the ORFs identified in newly sequenced genomes have no obvious homologues (Saunders and Moxon, 1998). Nevertheless, it is difficult to predict protein dynamics, structure or potential interactions purely from DNA sequences. Also, there is usually no direct correlation between gene activity and protein abundance. In genomics, together with proteomics (the study of all the proteins expressed by the cell), strategies are being developed for high throughput screening and to assign a biological function to newly identified proteins (Cahill, 2001). For example, Holt *et al.* (2000) have screened protein arrays with single antibody fragments from antibody libraries. With their approach 27 648 proteins could simultaneously be screened with the same antibody. The proteins identified in this way still need to be tested in experimental systems to determine or verify their function. Potential ORFs identified from the sequence of *Cowdria* are at present being tested in several different vaccine delivery systems (see 1.3.7). Sequence data is also useful to design DNA probes or PCR primers for use in diagnostic tests. These can be aimed at either unique or common regions between strains or serotypes depending on the type of test required.

It should be clear from the many examples given above that it is possible to use phage display to identify epitopes and/or mimotopes that could be useful for diagnosis or vaccine development, even without much prior knowledge of the organism or antigens. Accordingly, it was decided to apply phage display technology to the rickettsial organism, *Cowdria ruminantium.* Its genome sequence is not yet known. It causes a veterinary disease of economic importance in Africa and, as will be seen in the next section (1.3), both a better vaccine and improved diagnostic tests are required.

1.3. *COWDRIA RUMINANTIUM.*

1.3.1. The disease, heartwater.

Heartwater is caused by the rickettsia, *Cowdria ruminantium* (*Cowdria*). This disease of domestic and wild ruminants is endemic to sub-Saharan Africa. It also occurs on some Caribbean islands. Heartwater is one of the most important tick-transmitted diseases in southern Africa, causing economic losses due to mortality and the costs of controlling the disease (Uilenberg, 1983). The first documented case could be as far back as 1838, when Louis Trichardt reported in his diary a fatal disease amongst his sheep after coming in contact with ticks (Provost and Bezuidenhout, 1987). The disease causes accumulation of fluid in body cavities, especially in the fibroserous sac that surrounds the heart, hence the name heartwater (Prozesky, 1987b).

The distribution of heartwater generally corresponds to that of the tick vector, ie. several *Amblyomma* species*. A. hebraeum* and *A. variegatum* are most important in sub-Saharan Africa (Provost and Bezuidenhout, 1987; Barré *et al.,* 1987) while *A. maculatum* and *A. cajennense* have both been shown to transmit heartwater. Since both occur on the mainland of North America, there

is concern in some quarters that heartwater will in time spread to the Americas from the Caribbean (Barré *et al.*, 1987). It has also been shown that wildlife such as buffalo, blesbok, black wildebeest, springbok, impala, sable and tsessebe are all susceptible or can be carriers of the disease (Provost and Bezuidenhout, 1987; Peter *et al.*, 1999a; Peter *et al.*, 1999b; Kock *et al.,* 1995).

1.3.2. Classification.

Cowdria is an obligate intracellular rickettsial pathogen which stains negatively with Gram's stain (Winkler, 1990). Initially, the organisms were classified as *Rickettsia ruminantium* by Cowdry (1925) and were later renamed *Cowdria ruminantium* (Moshkovski, 1947). Phylogenetic analysis of the16S ribosomal RNA (rRNA) sequences confirmed a close relationship between the genus *Cowdria* and members of the genus *Ehrlichia*. *Cowdria* formed a tight cluster with *Ehrlichia canis* and *E. chaffeensis*, canine and human pathogens respectively (Dame, *et al.*, 1992; van Vliet *et al.*, 1992). Sequences from additional *Cowdria* strains have clarified its position within the ehrlichiae, particularly in genogroup III *Ehrlichia*, whereas the other closely related bovine pathogen, *Anaplasma marginale,* was in group II (Allsopp *et al.*, 1997). Recently, it has even been suggested that *Cowdria ruminantium* should be reclassified as *Ehrlichia ruminantium* (Dumler *et al.,* 2001) but in this thesis '*Cowdria*' will still be used. Stocks of *Cowdria* have been isolated from various geographical areas in southern and West Africa (du Plessis *et al.*, 1989; Uilenberg, 1983). These isolates differ in virulence and cross-protective capabilities. Analysis of antibody responses to various stocks has shown antigenic diversity between isolates (Jongejan and Thielemans, 1989; Rossouw *et al.*, 1990; see 1.3.6). Diversity was also evident from 16S rRNA sequence analyses (Allsopp *et al.*, 1997), from the random amplified polymorphic DNA (RAPD) assay (Perez *et al.,* 1997) and from macrorestriction fragment profiles (de Villiers *et al.*, 2000a). Because of these

differences there is a need for diagnostic tools to identify strains present in specific areas so that the appropriate vaccine control strategy can be implemented.

1.3.3. Symptoms of heartwater.

Heartwater can cause mortalities of up to 90% in sheep, goats and cattle (du Plessis and Malan, 1987) and as a consequence constrains livestock production. After tick transmission, the average incubation period is less than 2 weeks before a fever develops. This time depends on the animal and the heartwater isolate. Age, immune status and breed of animal also has an influence on the outcome of the disease which can be mild to peracute. Indigenous goat, sheep and cattle show innate resistence to heartwater. In peracute cases of the disease, animals develop pyrexia, sometimes show convulsions and die within a few hours. Lung oedema causes respiratory distress. Acute heartwater is characterised by fever of 40°C and a drop in temperature before the animal dies. Loss of appetite, loss of co-ordination, convulsions, hypersensitivity to bright light and noise are some of the symptoms (van de Pypekamp and Prozesky, 1987). The mild form of heartwater is difficult to diagnose because symptoms are not obvious and animals recover after a few days. Sheep generally show less severe symptoms compared to cattle (van de Pypekamp and Prozesky, 1987). Increased permeability of blood vessels plays a major role in the pathogenesis of heartwater, but the cause of this is unclear. Products of arachidonic acid metabolism, complement and immune complexes or a toxin may play a role (du Plessis *et al.,* 1987).

1.3.4. The life cycle of *Cowdria*.

It was shown in sheep that after tick transmission, *Cowdria* first infects the lymph nodes, where replication takes place. Electron-dense elementary bodies are then released into the blood stream

and enter endothelial cells. Organisms are taken up by phagocytosis and multiplication is by binary fission to form colonies. The colonies are enclosed by a thin membrane in phagosomal vacuoles in the cytoplasm (du Plessis, 1970). In the vertebrate host, colonies are pleomorphic and are divided into small, medium and large sizes ranging from 0.49µm to 2.7µm (Prozesky, 1987a). Elementary bodies are released following rupture of the endothelial cells. *Cowdria* infects macrophages and neutrophils in addition to endothelial cells (Logan, 1987). In the tick host, *Amblyomma sp, Cowdria* colonies are also pleomorphic but elementary bodies are predominantly seen. Organisms develop initially in gut epithelial cells and later in salivary gland cells, from where they are transmitted to the vertebrate host (Kocan and Bezuidenhout, 1987).

1.3.5. *In vitro* cultivation of *Cowdria*.

Some strains of *Cowdria* are pathogenic to laboratory mice, which provides a relatively cheap animal model for studying the organism (du Plessis and Kümm, 1971). The *in vitro* cultivation of *Cowdria* in calf endothelial cells was first achieved by Bezuidenhout *et al.* (1985). Improvements have subsequently been made by using a chemically defined, serum-free culture medium (Zweygarth *et al.,* 1998). *In vitro* cultivation of *Cowdria* has greatly helped with research on heartwater. Since *Cowdria* is intracellular and pleomorphic, purifying it from the host cell is still problematic. Organisms have been purified by differential centrifugation (Neitz *et al.*, 1987; Rossouw *et al.*, 1990), Percoll gradient purification (Mahan *et al.*, 1995) and on immuno-affinity columns (de Villiers *et al.*, 1998) but all these preparations still contain some host cell DNA and proteins.

1.3.6. Control of heartwater.

Heartwater is endemic in South Africa. It is normally controlled by the chemical control of ticks

(Bezuidenhout and Bigalke, 1987) and the infection-and-treatment method of vaccination. Vaccination, followed by tick infection in the field, maintains the immune status of the animals and ensures endemic stability (Howell *et al.*, 1981). After observing that survivors of heartwater were protected against reinfection, the control strategy of infection and treatment for heartwater was developed (Neitz and Alexander, 1945) and is still used in South Africa (Oberem and Bezuidenhout, 1987). The animals are vaccinated with a blood vaccine and treated with tetracycline as soon as a febrile reaction is observed. This vaccine consists of viable, virulent organisms (Ball 3 isolate) in sheep blood and has several shortcomings; a cold chain is required for delivery, there is a risk of transmitting other blood-borne pathogens and vaccination-associated deaths do occur. The vaccine is only suited for endemic areas (Oberem and Bezuidenhout, 1987). Cross protection occurs only between antigenically related strains (Audu *et al.,* 1995).

1.3.7. New approaches to heartwater control.

Due to the shortcomings of the current vaccine, improved ways of vaccination against heartwater are required. Several approaches have been investigated. Jongejan (1991) developed an attenuated vaccine. The Senegal isolate was attenuated by serial passage in endothelial cells. After eleven passages the *Cowdria* cultures failed to cause heartwater in goats and sheep which were immune to subsequent homologous challenge. Unfortunately, attenuation could not be achieved with other strains. This is thus not a general strategy applicable to heartwater control.

In another approach, inactivated *Cowdria* elementary bodies (Gardel isolate) were shown to protect animals against heartwater (Martinez *et al*., 1994; Martinez *et al*., 1996). Chemically inactivated cell-cultured organisms were combined with an adjuvant (complete Freund's) to vaccinate goats, sheep and cattle and resulted in protection against disease. Limited cross protection against other stocks was achieved. The vaccination did not prevent infection, since replicating organisms were

detected after challenge, but death was prevented. The safer Montanide ISA50 produced the same results as Freund's adjuvant and immunity lasted for up to 17 months. Martinez *et al*. (1996) postulate that this vaccination procedure together with exposure to field challenges will provide long-lasting protection against heartwater. Different levels of protection was observed, depending on the dose of challenge and experimental animals used; for example, goats were more susceptible than sheep. Another group (Mahan *et al.,* 1995; Mahan *et al.,* 1998a) have vaccinated sheep with culture-derived inactivated Crystal Springs and Mbizi organisms (Zimbabwean isolates) together with Freund's adjuvant and achieved protection against homologous and heterologous challenges. It was also observed that the vaccine prevented mortality, but not infection, as organisms were detected in brain biopsies after challenge. Protection was adjuvant-dependent and was associated with high levels of *Cowdria* antibodies. Large amounts of *Cowdria* organisms need to be cultured for this vaccination strategy and despite great improvements in the culture system (Zweygarth *et al.,* 1998), it is still quite a difficult and expensive task.

An appropriate recombinant vaccine could eliminate some of the problems associated with the abovementioned vaccine strategies, but for this, the actual proteins responsible for a protective immune response as well as their encoding genes need to be identified. The *map*1 gene (see 1.3.8 and 9) has been tested as a DNA vaccine and showed the potential to induce a protective immune response in mice (Nyika *et al.*, 1998). Variable results were obtained, 23 to 88% of mice immunised with the *map*1 construct were protected against lethal challenge and many showed a delayed mortality. No correlation between seroconversion and protection was observed, but a T helper cell type 1 immune response was observed in *in vitro* cell culture proliferation assays. Interferon gamma (IFN-γ) and interleukin-2 (IL-2) were detected. They postulate that *in vivo* $CD8^+$ T cells, IFN-γ and possibly major antigenic protein 1 (MAP1)-specific antibodies acting together could have delayed or prevented death in the mice. These studies need to be repeated in the natural hosts i.e. sheep, cattle and goats and other antigens which induce similar immune responses will probably need to

be identified to achieve more consistent protection.

One approach to find genes encoding antigens that could induce a protective immune response has been to screen *Cowdria* genomic libraries in animal models (Brayton *et al.*, 1998). Libraries were made in a *Salmonella* vaccine delivery system. Mice were immunised with pooled clones and then challenged to identify pools containing protective clones. These were then dissected into smaller groups and retested. One pool produced 14% protection and is to be studied further (Brayton *et al.*, 1998). *Cowdria* clones have also been screened with the so-called expression library immunisation system, whereby the plasmids containing *Cowdria* DNA inserts are administered to the animals and then subsequently challenged (N. C. Collins, personal communication). A genome sequencing project has been initiated to identify *Cowdria* genes and to predict from the DNA sequence which open reading frames could encode potentially protective antigens. These will then each need to be expressed and tested in animal models (Collins *et al.*, 1998; B. A. Allsopp, personal communication). The number of genes to be tested will depend upon the criteria used for selection. Nevertheless, this promises to be a daunting task.

Barbet *et al.* (2001) recently identified 34 plasmid clones encoding proteins which reacted with sheep immune sera (see 1.3.9). Some of these proteins stimulated the proliferation of peripheral blood mononuclear cells (PBMC) from immune cattle. Pools of bacterial lysates expressing the proteins were tested in mice. One showed 58 to 89% protection in two different experiments. The clones were tested separately, but no protection was achieved.

1.3.8. Antigenic characterisation of *Cowdria*.

Animals with heartwater develop antibodies. Immune serum has been used to study strains of *Cowdria* isolated from various geographical areas (du Plessis *et al.*, 1989; Jongejan *et al.*, 1988). The various stocks show antigenic diversity and also differ in virulence and cross-protection capabilities. Goat antisera against nine stocks of *Cowdria* (Senegal, Um Banein, Kiswani, Ifé, Gardel, Ball 3, Kümm, Kwanyanga and Welgevonden) were analysed on western blots to identify immunodominant proteins. Infected tissue from goat choroid plexus was used as antigen and a 32 kDa immunodominant protein was identified as being common to all stocks (Jongejan and Thielmans,1989). Rossouw *et al*. (1990) used cell-cultured *Cowdria* (Welgevonden, Ball 3 and Kwanyanga) and identified a 31 kDa and a 27 kDa protein to be immunodominant and common to the stocks tested (Welgevonden, Ball 3, Breed, Mara, Germishuys, Kümm, Kwanyanga, Mali, Comoro). The 31/32 kDa protein was found to vary in size between isolates (Barbet *et al*., 1994). Due to its immunodominance it was named the major antigenic protein 1 (MAP1; Barbet *et al*., 1994). Rabbit and goat antibodies were made against both MAP1 and the 27 kDa protein. Immunoblotting indicates that these two proteins share common epitopes, with those on MAP1 being dominant (van Kleef *et al*., 1993). Electron microscopy using an immunogold labelled monoclonal antibody specific for MAP1 revealed that it is a surface exposed protein (Jongejan *et al.,* 1991). A goat injected with polyacrylamide gel-separated MAP1 developed antibodies, but was not protected against disease when challenged (van Kleef *et al.*, 1993).

Mahan *et al.* (1994) identified immunogenic proteins by immunoprecipitation of radio-labelled *Cowdria* proteins with sheep hyperimmune serum. Proteins of 21, 32, 40, 46, 58, 85 and 160 kDa were found to be both immunogenic and antigenic. In this assay the antibodies and proteins were allowed to react in solution, the immune complexes were collected, and the selected proteins were

then separated by PAGE. The relative recognition of the proteins by the antibodies differed in the immunoprecipitation and western blot assays. Thus the strong reaction with MAP1 (compared to other proteins) by western blotting was due to the immunodominance of MAP1 and not due to its abundance (Mahan *et al.*, 1994).

Cross-protection between antigenically diverse isolates differs. This may indicate that the antigens responsible for protective immunity are polymorphic. In an attempt to identify any such polymorphic antigens, serum from goats rendered immune to the Gardel and Senegal isolates was cross-adsorbed and tested for immuno-reactivity (Perez *et al.*, 1998). Goat serum raised against the Gardel isolate was adsorbed with Senegal elementary bodies and tested for residual reactivity by ELISA and western blot using homologous and heterologous antigens. The same was done with a serum directed against the Senegal isolate. Distinct serogroups were revealed by ELISA. A group of variable proteins of 23 to 29 kDa identified by western blot shared antigenic determinants between isolates. In addition, a 26-27 kDa protein specific for the Gardel isolate was identified. The study showed that Welgevonden and Gardel are antigenically closer to each other than to the Senegal stock. This correlates with the genetic analysis by RAPD (Perez *et al.,* 1997).

1.3.9. Molecular characterisation of *Cowdria*.

Cowdria has a circular genome estimated to be between 1546 and 1692 kilo bases (kb) in size (de Villiers *et al.*, 2000a and b) and a has high adenine and thymine (AT) content (70%; van Vliet *et al.*, 1994). A physical map of the genome was constructed from data obtained from pulsed field gel electrophoresis of DNA restriction fragments, and several genes such as *map*1, pCS20 and GroEL (see below) were mapped by hybridization (de Villiers *et al.*, 2000b). This provides useful basic molecular genetic information, such as the required size of DNA libraries to represent the complete

genome.

Attempts to clone the gene encoding MAP1 from genomic libraries were not successful (van Vliet *et al*., 1993). The amino acid sequence of the N-terminus and an internal sequence of MAP1 were obtained from which two oligonucleotides were derived. PCR with these two primers amplified a 99 bp fragment of the *map*1 gene from Senegal genomic DNA which was used to identify a genomic *Hind* III fragment. This DNA was subsequently cloned and contained most of the *map*1 gene. A genomic *Pst*I-*Sau*3A fragment was cloned to obtain the rest of this 854 bp gene. Low levels of truncated MAP1 were expressed from its own promoter, indicating that *Cowdria* promoters were active in *E. coli* (van Vliet *et al*., 1994). Sequence comparison of the *map*1 genes from six isolates showed an overall similarity of between 86 to 99.4% and that there were three variable regions which are spread across the gene (Reddy *et al.,* 1996). Southern blot analysis showed that *map*1 was a single copy gene (van Vliet *et al*., 1994; Reddy *et al.,* 1996). This has however been disputed by Sulsona *et al.* (1999) who found by sequencing the flanking regions of *map*1 that two members of the *map*1 gene family were located in tandem in the genome. The second open reading frame (*orf*2) was highly conserved between isolates (99 to100% identity) in contrast to *map*1 which varied between isolates (89 to 100% identity). The levels of identity between the six *map*1 and *orf*2 regions were between 43 to 46% (Reddy *et al.,* 1996).

Antiserum from a sheep infected with the Crystal Springs isolate was used to screen a *Cowdria* genomic library made in a plasmid vector. A gene encoding a 21 kDa protein was isolated which was named the major antigenic protein 2 (MAP2). The molecular weight was conserved between seven additional isolates and antiserum from animals infected with the isolates all contained antibodies recognising MAP2 (Mahan *et al*., 1994). This protein had 55.5% sequence identity with the major surface protein 5 (MSP-5), a 19 kDa protein of *Anaplasma marginale* which is used in

a serodiagnostic assay for anaplasmosis (Visser *et al.,* 1992). Cloning of four additional *map*2 genes revealed that this was a highly conserved protein. It had 83 and 84% identity with similar proteins from *E. chaffeensis* and *E. canis* and was thus not suited for heartwater diagnosis (Bowie *et al.*, 1999).

Several other *Cowdria* genes have been isolated, not all encoding proteins with known functions. The *Cowdria groE* operon, encoding heat shock proteins (HSP), was isolated by screening a λZAPII library (Welgevonden isolate) with goat antiserum against the Ball 3 isolate. The *Cowdria* GroES and GroEL proteins (also known as HSP10 and HSP60) showed 83.2 and 93.5% similarity respectively, to those of *E. chaffeensis* (Lally *et al.*, 1995). The expressed proteins reacted with antiserum against several *Cowdria* isolates, but not with serum against *Ehrlichia* species. The potential use of these proteins in an immunodiagnostic test remains to be investigated, as does whether it plays a role in protective immunity or pathogenesis of heartwater. DNA clones pCS20 and pCR9 were isolated from a pUC19 library with the aim of finding genes for use as DNA probes to detect *Cowdria* DNA (see 1.3.11). Both clones contained open reading frames for proteins of unknown function; nevertheless pCS20 was useful in diagnostics (Waghela *et al.*, 1991).

Recently, Barbet *et al.* (2001) constructed *Cowdria* libraries in plasmid vectors from two different strains; one from the Highway (Zimbabwe) and one from the Gardel strain (Guadeloupe). Using antiserum from a sheep immune to the Crystal Springs strain (Zimbabwe) to screen the libraries identified 34 clones that expressed proteins reactive with several other sheep immune sera. DNA hybridisation showed that eight clones contained either the previously identified *groES/EL* or the *map*2 genes. Interestingly, no *map*1 containing clones were isolated from these plasmid libraries. The remainder of the clones were sequenced (73 kb in total) and 33 complete and 23 partial genes were identified. The sequences were compared with sequences in the database to identify potential

proteins. In addition, *in vitro* transcription and translation followed by immunoprecipitation was used to analyse protein expression from the clones. Two proteins were of special interest, one of 37 kDa, with similarity to a *Brucella abortus* outer membrane protein (OMP) and one of 28 kDa, with similarity to a *Coxiella burnetii* OMP. The two open reading frames were cloned into an expression vector and the resulting recombinant proteins were recognised by immune sheep sera. Another gene of interest was one encoding a repeated peptide sequence (VTSSPEGSV), similar to that found in other rickettsial organisms. A synthetic copy of this peptide reacted with immune sera in ELISA. Bacterial lysates containing all the above mentioned clones were tested for their potential to induce a protective response against heartwater (see 1.3.7), but neither the 28 kDa and the 37 kDa proteins nor the one containing the repeat sequence induced protection in mice.

1.3.10. Serodiagnosis.

To control a disease such as heartwater there has to be a good understanding of the epidemiology of the causative organism. Definitive diagnosis of heartwater was traditionally made by the brain squash technique (Purchase, 1945) in which organisms were detected in endothelial cells of stained brain smears obtained by postmortem or biopsy. Heartwater-infected animals develop antibodies and many attempts have been made to develop tests for easier diagnosis of the disease using serum from animals. An indirect fluorescent antibody (IFA) test was developed by du Plessis and Malan (1987) using as antigen peritoneal macrophages of mice infected with *Cowdria*. This test was used in epidemiological studies, to evaluate immunization and to detect antibodies against *Cowdria* in wildlife. Problems such as false positive reactions from the serum of animals in heartwater-free areas were encountered with this technique. This was as a result of cross-reactions due to the presence of *Ehrlichia* species (du Plessis and Malan, 1987; Mahan *et al.*, 1993).

A competitive ELISA based on a monoclonal antibody directed against MAP1 was developed to detect *Cowdria* antibodies present in animal sera (Jongejan *et al.,* 1991). This test was positive for 79% of the sera from goats experimentally infected with *Cowdria*. The sera that were negative with ELISA were shown to react with *Cowdria* proteins on western blots indicating that false negative results were obtained with this ELISA. Nevertheless, there was good correlation with results obtained with the IFA test (using cultured organisms as antigen) but the ELISA was more sensitive. No cross-reaction was detected with *E. phagocytophila* antibodies. This test could also be used to detect the presence of antibodies in the sera of wildlife species (section 1.3.1).

Another ELISA, based on a fragment of MAP1 called MAP1-B which encodes amino acids 47 to 152 was developed. This recombinantly expressed polypeptide reacts specifically with sera from *Cowdria*-infected animals, but not with sera from *E. bovis-, E. ovina-* or *E. phagocytophila*-infected animals. These *Ehrlichia* species normally infect ruminants and thus in this test antibodies directed against these organisms would not cause false positives. Cross-reactivity was still, however, observed with *E. canis* and *E. chaffeensis* antibodies (van Vliet *et al.,* 1995). The problem of false positives seems to have been solved with this test, but most of the studies were performed with goat and sheep serum. When this test was used on bovine serum, the sensitivity was very low (Mahan *et al.*, 1998b; Mondry *et al.*,1998). Thus, with the current serodiagnostic tests, false negative results are obtained, especially with bovine serum.

1.3.11. DNA-based diagnosis.

Detection of *Cowdria* DNA in material from infected animals (e.g. blood) is potentially a direct and sensitive method of diagnosis. DNA-based diagnosis is in fact the method of choice to detect carrier animals and determine the isolate with which the animal is infected. Waghela *et al.* (1991) isolated

a DNA clone (pCS20) from a *Cowdria* (Crystal Springs isolate) pUC9 library. This 1500 bp DNA fragment was very specific for *Cowdria* and has been used to detect *Cowdria* DNA in ruminant blood and tick material. To improve on the sensitivity of the test, PCR primers were designed to amplify a 279 bp genomic fragment which was then hybridised with the pCS20 probe. This test could specifically detect 1 to 10 *Cowdria* organisms (Peter *et al.*, 1995). This level of sensitivity is required to detect low level carriers of heartwater. A PCR assay based on the *map1* gene has been used to identify carriers in wild life (Kock *et al.*, 1995). A panel of 16S rRNA gene probes has been developed to detect different *Cowdria* genotypes (Allsopp *et al.*, 1999). The *map*1, pCS20 and rRNA probes were compared using PCR amplification followed by hybridization with the specific probes. The *map*1-based test was the least sensitive and specific of the three while the pCS20-based test was best suited for large scale diagnosis. rRNA probes are not as sensitive, are technically demanding to use, but give useful phylogenetic information (Allsopp *et al.*, 1999).

1.3.12. Immunity to heartwater.

Animals exposed to heartwater develop specific antibodies, but they do not seem to confer protection. In 1931 Alexander administered hyperimmune bovine antiserum to animals prior to, simultaneously with, and after infection with *Cowdria*, all with no effect on the course of disease. In another study, antibodies from immune and non-immune animals failed to inhibit the development of heartwater in sheep. Both the antibody preparations prevented clinical heartwater when allowed to react with the blood prior to infection (du Plessis, 1970). This, however, was obviously not a specific immunological reaction. One of the factors that apparently causes this non-specific resistance is conglutinin (du Plessis, 1985). The addition of complement to immune serum was also shown to have a cowdricidal effect (du Plessis, 1993). Immune serum from mice could neutralize *Cowdria* in bovine endothelial cells, but this inhibitory effect has not been shown *in vivo*

(Byrom *et al.*, 1993). New born calves have an innate immunity to heartwater for up to 4 weeks. Colostrum from immune dams have been shown to play a role in this resistance. Deem *et al.* (1996) suggest that antibodies in the colostrum might be the cause, but colostrum consists of various components such as T and B cells as well as soluble proteins, all of which could have beneficial effects.

Protection against heartwater is nowadays thought to be predominantly cell-mediated. Adoptive transfer of immunity to susceptible mice with spleen cells from immune mice provided the first supportive evidence for this. When the cells were depleted of $CD8^+$ T cells (cytotoxic T-cells), they failed to confer immunity to the recipients (du Plessis *et al.*, 1991). $CD8^+$ T cells play an important role in acquired immunity having the functions of suppression, cytotoxicity and serving as a source of lymphokines (du Plessis *et al.*, 1991).

Recombinant bovine interferon gamma (rBoIFNγ) inhibits the growth of *Cowdria in vitro.* rBoIFNγ usually induces expression of MHC class II molecules on cells such as endothelial cells. In *Cowdria*-infected endothelial cells, MHC class II expression was inhibited (Totté *et al.,* 1996). This was confirmed in another study that showed both MHC class I and MHC class II surface expression by endothelial cells was reduced in a dose dependent manner after *Cowdria* infection (Vachiéry *et al.,* 1998). Accumulation of MHC class I molecules in intracellular compartments was observed, probably due to disruption of transport of the molecules to the cell surface. Degradation or inhibition of synthesis may have occurred in the case of MHC class II molecules (Vachiéry *et al.,* 1998). It seems that in *Cowdria* infections, the endothelial cells are not able to present antigens to T cells in an MHC class II restricted manner. *Cowdria* may therefore have developed mechanisms to evade the immune recognition by somehow reducing the capacity of endothelial cells to express MHC class II molecules (Totté *et al.,* 1996).

Cattle immunised with inactivated *Cowdria* can be immune against virulent challenge for up to 10 months after vaccination (Totté *et al.,* 1997). T-cell lines generated from these cattle were more than 95% $CD4^+$, MHC class II restricted and produced interferon gamma (INF-γ). It was proposed that protection is conferred by the effect that helper T lymphocytes have on antibody maturation and the stimulatory factors they secrete to help with the generation of $CD8^+$ T cells. The INF-γ produced could also inhibit the growth of *Cowdria*.

The proliferative immune responses of cattle immunised by infection and treatment with *Cowdria* were characterised by the expansion of $CD4^+$, $CD8^+$ and γδ T cells. Production of cytokines IFN-γ, tumour necrosis factor alpha (TNF-α), TNF-β and interleukin-2 (IL-2) was shown by reverse transcription-PCR (Mwangi *et al.*, 1998a). The authors suggest that endothelial cells and monocytes may present *Cowdria* antigens to specific T lymphocytes during infection and in this way play a role in inducing protective immunity.

Interferons elevate the production of nitric oxide, which in turn can be microbiocidal and have cytotoxic effects on intracellular parasites. Endothelial cells infected with *Cowdria* showed signs of elevated nitric oxide production, which was further stimulated by rBoIFNγ, resulting in reduced viability of *Cowdria*. The production of nitric oxide could be a mechanism by which animals reduce the spread of infection (Mutunga *et al.*, 1998).

Specific $CD4^+$ T-cell lines were generated from cattle immunised with inactivated *Cowdria* (Gardel isolate). They responded to both soluble and membrane fractions of the organism in proliferation assays. The soluble fractions were fractionated by fast performance liquid chromotography (FPLC) and the fractions that contained proteins ranging from 22 to 32 kDa induced proliferation. These fractions also induced the production of INF-γ, indicating that these are immunogenic proteins that

could conceivably induce a protective immune response in host animals. The cell lines did not respond to recombinant MAP1, the well characterised, 31 kDa protein. Thus the proteins that induced the proliferative responses are still unknown and need to be further characterised (Totté *et al.*, 1999). These results were in contrast to experiments in which cattle were immunised by infection and treatment and proliferative responses to recombinant MAP1 and also MAP2 were detected (Mwangi *et al.*, 1998b)

In another study, *Cowdria* (Welgevonden isolate) proteins were separated by continuous flow electrophoresis (CFE) and the fractions tested for their ability to stimulate lymphocyte proliferation *in vitro* (van Kleef *et al.,* 2000). PBMC from animals immunised by infection and treatment and by inactivated *Cowdria* were studied. Several protein fractions were identified that stimulated proliferation in all the animals tested. The sizes of these proteins were 11, 12, 14 to 17 and 19 to 23 kDa. This range differs from what was found by FPLC separation (Totté *et al.*, 1999), but could be explained by the better resolution provided by CFE or due to the different isolates used in the two studies. These protein fractions have obvious potential to play a role in protection against heartwater. They need to be further characterised and their encoding genes identified. This thesis addresses one of the ways in which the identification of the genes that code for these proteins can potentially be achieved utilising phage display technology.

There is thus still a need for good, sensitive serodiagnostic tests. To develop this and also for recombinant vaccine development, the immunologically relevant genes need to be identified. The work that follows was aimed at using the B cell responses to defined regions on particular proteins and exploiting phage display technology, with its advantages of coupling genotype and phenotype, to search for genes that cannot readily be located by other techniques.

CHAPTER 2

CONSTRUCTION AND ANALYSIS OF A LAMBDA ZAPII LIBRARY.

2.1. INTRODUCTION.

A better understanding of a disease and facilitation of the development of diagnostic tests and recombinant vaccines are often obtained by characterising those proteins of pathogens that interact with components of the immune system. In an attempt to identify genes encoding *Cowdria* antigens, a lambda (λ) bacteriophage expression library was constructed and analysed. Using this approach, many genes from a wide variety of organisms have been isolated by antibody screening (Young and Davis, 1983; Helfman *et al.*, 1983; Skilton *et al.,* 1998). These include the heat shock proteins from *Cowdria* (Lally *et al.*, 1995).

The bacteriophage λ particle has an icosahedral head which contains its DNA. In addition, there is a 'tail' attached to the head (Figure 2.1). It's linear, double stranded DNA genome of 50 kb has complementary single stranded cohesive ends (*cos*) at the 3' and 5' termini. This enables circularisation of the genome after linear DNA is injected into *E. coli*. Replication is by the rolling circle model, which results in a linear molecule of many copies of the genome, linked by the *cos* sites. This molecule is cleaved at the *cos* sites into single 50 kb units as they are packaged into the phage heads. Upon infection, the λ tail adsorbs to receptors involved in maltose transport on the surface of *E. coli*. The DNA is then transferred through the cell wall into the bacterium, where the bacterial enzymes replicate the phage DNA and translate the genes. The phage particles then assemble and are released from the cell. The host cell machinery is switched off slowly during this

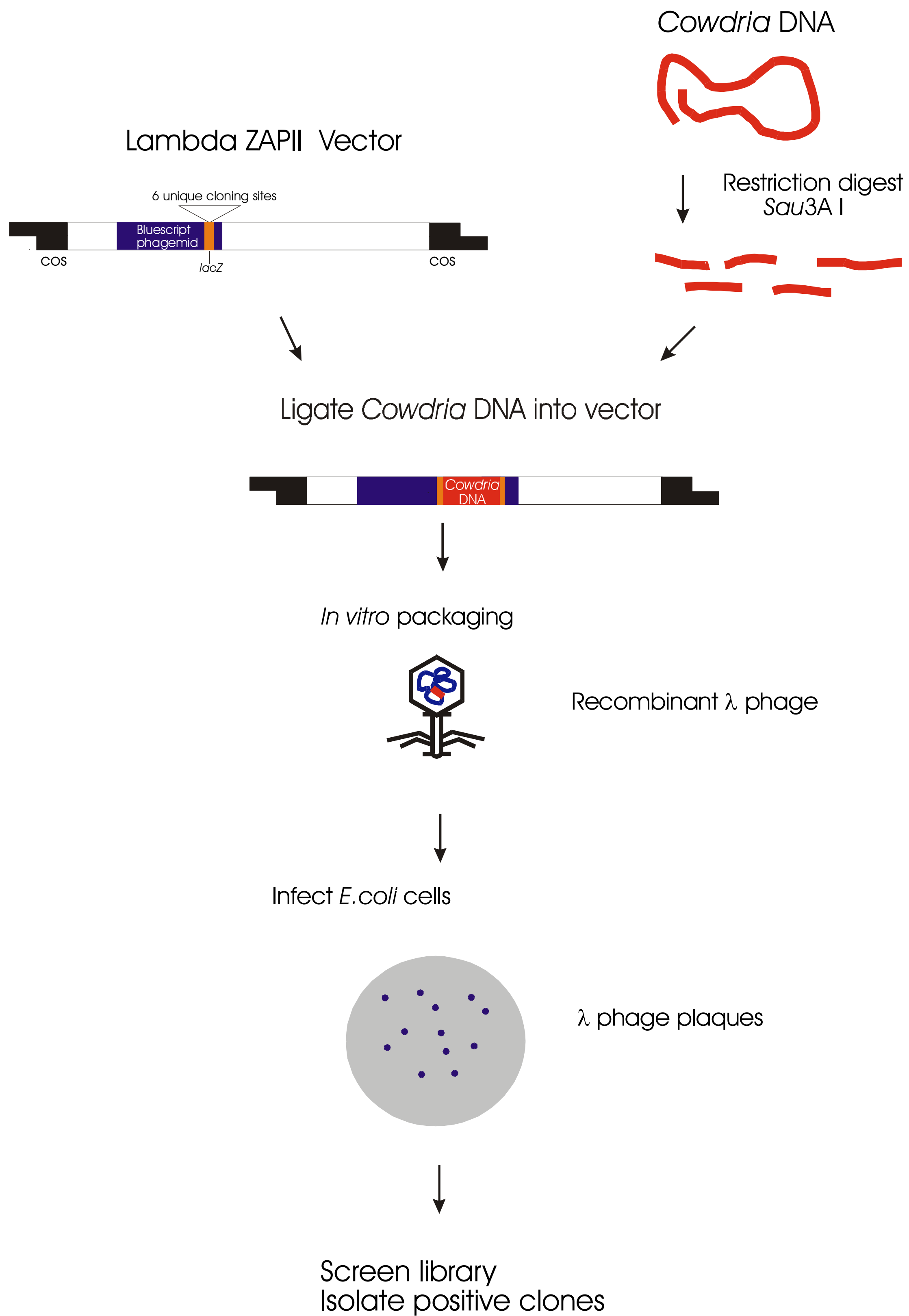


Figure 2.1. Schematic representation of the construction of the *Cowdria* λ ZAP II library.

process and eventually the cell dies, but only after the phage has had enough time to produce sufficient progeny (Freifelder, 1983).

The bacteriophage λ genome has been modified to enable foreign DNA inserts to be cloned. There are nonessential genes in λ that can be replaced with up to 10 kb foreign DNA. Bacteria usually rapidly degrade foreign proteins, but it has been shown that a fusion protein with a small part of β-galactosidase can stabilise such a protein (Young and Davis, 1983). Thus, a multiple cloning site within the *lac*Z gene was inserted into λ. DNA cloned into this site will inactivate β-galactosidase and allows for blue/white selection of recombinant plaques (Young and Davis, 1983).

A commercially available vector, λ ZAPII (Stratagene) was chosen to construct the *Cowdria* library since it combines the high efficiency the λ *in vitro* packaging system with the versatility of a plasmid system. Once λ clones are selected, the DNA inserts can be excised *in vivo* into a phagemid (or plasmid) vector from the bacteriophage λ genome. It is more convenient to sequence and subclone the inserts from the phagemid DNA (3 kb plus insert) than from the large λ genome (50 kb). The steps involved in constructing the library are illustrated in Figure 2.1. The resulting library was screened with both DNA and antibody probes.

2.2. MATERIALS AND METHODS.

2.2.1. *In vitro* cultivation of *Cowdria*.

The Welgevonden isolate of *Cowdria* (du Plessis, 1985) was cultured in an E5 endothelial cell line (Bezuidenhout *et al.,* 1985). Cell-culture material was initially supplied by E. Horn and C. Yunker and later by E. Zweygarth and A. Josemans of the Parasitology Division, Onderstepoort Veterinary

Institute.

2.2.2. Purification of *Cowdria* organisms.

Crude infected and uninfected protein extracts from cell cultures were prepared as described by Rossouw *et al.* (1990). Cell-culture material was usually frozen. Thawing the mixture therefore resulted in the lysis of most of the endothelial cells. This solution was centifuged at 1000 g for 10 minutes to remove cell debris and the supernatant was centrifuged at 30 000 g for 30 minutes. The pellet containing the *Cowdria* organisms was resuspended in a minimal volume of phosphate-buffered saline (PBS) and stored at -20°C. A rough estimate of the protein content was determined by spectophotometry, using the formula (1.55 x A_{280}) - (0.76 x A_{260}) = mg/ml (Harlow and Lane, 1988). This antigen preparation was used for SDS-PAGE (Laemmli, 1970; Appendix B) and western blots (Towbin *et al.*, 1979; Appendix B). Immunoaffinity-purified *Cowdria* was supplied by Dr. M. van Kleef (Brayton *et al.,* 1997) from which DNA was isolated and used for constructing the λZAPII library.

2.2.3. Isolation of *Cowdria* DNA.

Cowdria genomic DNA was isolated from immunoaffinity-purified organisms by treatment with SDS and proteinase K (Ausubel *et* al., 1998). Organisms were resuspended in TE (10 mM Tris, pH 7.4 and 1 mM EDTA, pH 8) and lysed in the presence of 0.5% SDS and 100 µg/ml proteinase K for 1 hour (h) at 37°C. A sixth volume 5 M NaCl was added and mixed well before adding 8.75 volumes CTAB/NaCl solution (Appendix A) followed by incubation at 65°C for 10 minutes. An equal volume of chloroform/isoamyl alcohol (24:1) was added and centrifuged at 13 000 g for 5 minutes to precipitate the CTAB-protein/polysaccharide complexes. The DNA was then extracted from the aqueous phase with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1),

precipitated with 0.6 volume isopropanol and finally resuspended in TE. The DNA concentration was determined by spectophotometry (1 A_{260} unit = 50 µg/ml; Sambrook *et al.,* 1989).

2.2.4. Determining the bovine DNA content in the *Cowdria* DNA preparation.

The amount of bovine DNA in the starting material used for constructing the library was estimated by slot-blotting tenfold serial dilutions of known quantities of purified *Cowdria* DNA and bovine DNA (calf thymus DNA, Roche). The blots were then hybridized with a radiolabelled bovine DNA probe and the levels of hybridization were compared. Prior to blotting, the DNA was denatured by incubation at 65°C for 1 h in the presence of 0.3 M NaOH in TE and subsequently neutralised with an equal volume 2 M ammonium acetate. The samples were transferred to a membrane (Magna Charge Nylon filters, MSI) with a slot blot apparatus (Hoefer Scientific Instruments). The DNA was bound to the filters by baking at 80°C for 2 h. For probes, bovine DNA or *Cowdria* DNA were digested with *Sau*3A I and denatured by heating at 65°C for 10 minutes before probe synthesis. DNA was labelled with the Multiprime labelling system (Amersham) and [α-^{32}P]dCTP (10 Ci l^{-1}, 3 Ci µmol^{-1}, 3.3 µM, Amersham). Filters were prehybridized in 0.5 M sodium phosphate buffer (pH 7.4), 7% (w/v) SDS (Church and Gilbert, 1984) for at least 1 h and hybridised overnight at 65°C in the same buffer with the addition of 2 x 10^6 cpm/ml of the probe. The next day, the excess probe was removed by two washes in 2 x SSC, 0.1% (w/v) SDS for 10 minutes at room temperature and one wash in 0.1 x SSC, 0.1% (w/v) SDS for 20 minutes at room temperature, followed by two stringent washes in 0.1 x SSC, 0.1% (w/v) SDS for 15 minutes at 65°C. The hybridization results were visualized by autoradiography (Cronex film, Du Pont).

2.2.5. Bacteriophage λ library construction.

The λZAPII genomic library was constructed as specified by the supplier of the phage vector

(Stratagene). The vector was prepared as follows: the *cos* sites of the λZAPII arms were ligated to prevent filling in of the ends in later procedures, before digestion with *Xho* I. Thirty micrograms of λZAPII arms were ligated in 30 µl containing Promega ligase buffer, 1 mM ATP and 12 U (unit) T4 ligase (Promega) overnight at 4°C. The ligase was heat inactivated at 65°C for 15 minutes and left to cool at room temperature, before digestion with 50 U *Xho* I (Roche) in a volume of 75 µl. The DNA was extracted once with phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated. To make the *Xho* I site of the vector compatible with the *Sau*3A I sites of the inserts to be cloned, the *Xho* I ends were filled in with 1 mM dTTP and 1 mM dCTP using 5 U DNA polymerase (Klenow fragment) in Klenow buffer (Appendix A) for 30 minutes at room temperature, heat inactivated, phenol extracted and precipitated as above. The DNA pellet was resuspended in 10 µl TE.

Cowdria DNA inserts were prepared from DNA isolated from immunoaffinity-purified *Cowdria* organisms (2.2.2 and 3). The *Cowdria* DNA was partially digested with *Sau*3A I to obtain fragments of two to 20 kb in size (conditions optimised by E. P. de Villiers). *Cowdria* DNA (3 µg in 30 µl) was digested with 0.15 U *Sau*3A I (Promega). Half of the reaction mixture was removed after 45 minutes and the remainder after 60 minutes, in each case the reaction being terminated by the addition of 15 µl ice cold 30 mM EDTA. The reactions were pooled, phenol extracted, ethanol precipitated and the resulting DNA pellet was resuspended in 6 µl TE. The *Sau*3A I ends were partially filled in using 1 mM dGTP, 1 mM dATP and DNA polymerase (Klenow fragment) as above.

Vector DNA (4 µg) was ligated to *Cowdria* insert DNA (0.4 µg) at 13°C overnight in a volume of 20 µl. The entire ligation reaction was packaged into Gigapack II or XL packaging extracts (Stratagene) in six separate reactions. To determine the titre of the library, tenfold dilutions were made in SM (50 mM Tris pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$, 0.01% gelatin) of which 10 µl was

added to 200 µl of XL1 Blue cells (Appendix C) and incubated at 37°C for 15 minutes. Top agar (3 ml) was added to the mixture which was then poured onto Luria-Bertani (LB) plates (Appendix A) and incubated overnight at 37°C. Blue-white selection of recombinants was achieved by addition of 7 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and 12,5 mg/ml X-gal (5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside) to the top agar.

2.2.6. Antibodies used for screening the λ ZAP II library.

Goat antiserum directed against PAGE-separated *Cowdria* MAP1 was provided by Dr. M. van Kleef (van Kleef *et al.*, 1993). Goat antiserum was also raised against sonicated *Cowdria* (Welgevonden) organisms. In an attempt to drive the immune response to proteins other than the immunodominant protein, MAP1, passive immunisation was used (Thalhamer and Freund,1985). This was done in order to prevent any immunodominant epitopes from being recognised by B cell receptors. Five minutes prior to injecting the antigen, 1 ml of the above-mentioned anti-MAP1 serum was injected into the goat. Immunoaffinity-purified *Cowdria* organisms (500 µg) were disrupted by sonication (MSE), at full power, for a total of 8 minutes, mixed with an equal volume of ISA 50 (Seppic, France) and injected subcutaneously. Serum was collected 2 weeks post immunisation.

Antibodies reacting with *E. coli* proteins often occur naturally in polyclonal antiserum and could cause false positive results. Thus, these antibodies were removed by pre-absorption with *E. coli* lysate as described in the *pico*Blue Immunoscreening protocols (Stratgene). *E. coli* (XL1 Blue) was grown overnight and collected by centrifugation, resuspended in PBS and sonicated to lyse the bacteria. Nitrocellulose filters were soaked in the lysate (diluted 1/10 in PBS) for 30 minutes and air dried. The filters were washed 5 times in PBS before incubation in PBS containing 1% BSA (PBS1%BSA) for 30 minutes. These filters were used to absorb the antisera by incubating a filter

in the serum (1/5 dilution in PBS1%BSA) for 10 minutes at 37°C, the filter was removed and this process repeated with 3 more filters. The resulting antiserum preparations were used for screening.

2.2.7. λ library screening with antibodies.

To screen the bacteriophage λ library with antibodies, expression of fusion proteins was induced from the *lacZ* promoter on the vector. Phages were plated out as above (2.2.5) at a density of 1000 pfu per plate and incubated at 42°C for 4 h. The plates were then overlaid with nitrocellulose filters (Hybond-C, Amersham) which had been pre-soaked in 10 mM IPTG. The plates were incubated at 37°C for a further 3 h. The filters were removed and incubated in PBS1%BSA for 1 h or overnight before incubation with antibodies for 2 h. The absorbed antisera (2.2.6) were diluted in PBS1%BSA, anti-MAP1 (1/200) and anti-*Cowdria* (1/50). Excess antibodies were removed by 3 rinses in PBS and 3 washes for 5 minutes in PBS. The antibodies were detected with donkey anti-goat IgG peroxidase conjugate (The Binding Site), diluted 1/1000 in PBS1%BSA and incubated for 1 h. The excess secondary antibody was removed by washing in PBS as above, before addition of freshly prepared horse radish peroxidase substrate (4-chloro-1-napthol; Appendix A) for colour detection. Positive signals were located on the respective plates, agar plugs of each region was picked and stored in 500 µl SM. These phage stocks were plated out again and retested as above to confirm results of primary screening.

2.2.8. *Map*1 primers and PCR.

Primers were designed to amplify the *Cowdria map*1 gene of the Welgevonden isolate by PCR. This was used as a DNA probe. The sequence of the *map*1 gene of the Senegal isolate (Van Vliet *et al.*,

1994, GenBank accession number X74250) was used for this purpose. Primer CR32-1 (5'GATGTAATACAGGAAGAG) corresponds to positions 471-491 and primer CR32-2 (5'GCTTCCGGGTTTATTGAA) to positions 1315-1335 of the published sequence. The resulting PCR product would span the complete coding region. PCR was carried out with *Taq* DNA polymerase (according to instructions, Promega) in the presence of 1 pmol of each primer/µl, for 1 cycle at 94°C (5 minutes) and 55°C (1 minute 45 seconds (s)) followed by 40 cycles at 72°C (3 minutes), 94°C (45 s) and 55°C (1 minute 45 s) in a Thermal reactor (HYBAID). The PCR product was ethanol precipitated to concentrate the DNA and resuspended in TE ready for probe synthesis.

To test recombinant λ clones for the presence of the *map*1 gene, PCR was performed on clones with conditions as above. The DNA templates were prepared as follows: agar plugs from the areas of λ plaques that were positive were picked, 2 µl of the phage suspension was added to15.75 µl water, boiled for 5 minutes and cooled on ice for 5 minutes. This total sample was used as template in the PCR.

2.2.9. *Map*1 probe synthesis and hybridisation.

The *map*1 PCR product (2.2.8) was labelled with a Multiprime labelling system and membranes for hybridisation were prepared as before (2.2.4). Plaque lifts of the library were made from overnight plates with approximately 200 plaques per plate. Filters were placed on cooled plates and left for 5 minutes before careful removal. The DNA was bound to the filters by baking at 80°C for 2 h. Prehybridisation, hybridisation and washing was as above (2.2.4.).

2.2.10. Excision of pBluescript phagemids from the λZAP II genome.

In vivo excision of the pBluescript phagemids was performed according to the manufacturer's

instructions (Stratagene). Phages (10^5) were added to 200 µl XL1 Blue cells (OD_{600} = 1) and 10^6 of the ExAssist helper phage followed by 15 minutes incubation at 37°C. After addition of 3 ml LB broth, the cells were incubated for 3 h at 37°C and the bacteria were subsequently heat killed for 20 minutes at 70°C. The solution was cleared by centrifugation for 15 minutes at 4000 g. The phages in the supernatant (1 and 50 µl) were used to infect 200 µl SOLR cells (OD_{600} = 1, Appendix C) by incubation for 15 minutes at 37°C. After infection, 100 µl of each reaction was plated onto LB plates containing 50 µg/ml ampicillin, incubated overnight at 37°C. The resulting colonies contained the phagemids. The colonies were grown in 5 ml LB broth (containing 50 µg/ml ampicillin) overnight or as described below (2.2.11) for protein induction.

2.2.11. Protein expression from phagemid clones.

To determine whether the pBluescript phagemids expressed the correct proteins, the cells were induced with IPTG to produce fusion proteins and then detected on western blots. SOLR cells containing the pBluescript *map*1 phagemid were grown to an OD_{600} of 0.5 and 1 mM IPTG was added to induce the *lacZ* promoter. Aliquots (1 ml) were removed at 0, 1, 2 and 3 h after induction. The cells were collected by centrifugation, resuspended in 100 µl PBS and stored at -20°C. Equal volumes of PAGE sample buffer (Appendix B) were added to the samples, boiled for 3 minutes, centrifuged (10 000 g for 10 minutes) and the proteins separated by electrophoresis in a 12% polyacrylamide gel containing SDS (SDS-PAGE; Appendix B). Prestained molecular weight markers (Sigma) were used to estimate molecular weights. Proteins were transferred to Immobilon-P transfer membranes (Millipore) with a semi-dry transfer unit (Hoefer). After electrophoresis the gels were soaked in transfer buffer (Appendix B) for 10 minutes and the membranes were treated according to the manufacturer's instructions. The gels and membranes were sandwiched between buffer-soaked 3MM paper (Whatman) and the proteins were transferred (140 mA, 1 h) to the membranes. After transfer, the membranes were rinsed in PBS to remove adhering polyacrylamide

before detecting the expressed proteins with anti-MAP 1 antibodies as described above for library screening.

2.2.12. Sequencing of the *map*1 gene.

To confirm the identity of the *map*1 clone, it was sequenced by Dr. E. P. de Villiers (Molecular Biology Division, Onderstepoort Veterinary Institute).

2.3. RESULTS.

2.3.1. Genomic library construction.

Cowdria organisms were cultivated in bovine cells. The *Cowdria* DNA could thus still possibly have contained some host cell DNA. To estimate the amount of bovine DNA present in the DNA isolated from immunoaffinity-purified *Cowdria* organisms, serial dilutions of the DNA were compared with dilutions of bovine DNA, all of which were hybridised with a bovine genomic probe. The intensity of the hybridisation signal obtained with 1 ng of *Cowdria* DNA was less than half the intensity produced by 0.1 ng of bovine DNA (Figure 2.2.a), indicating that the bovine DNA comprised less than a tenth of the preparation. This was considered sufficiently pure for the construction of a library.

The *Cowdria* genomic DNA was partially digested by *Sau*3A I and ligated into the λZAPII phage vector. Ligation reactions were packaged into λ phage heads *in vitro* and transduced into *E. coli* XL1-Blue. The packaging efficiency was 6.8 x 10^6 plaque forming units (pfu)/µl, giving a total of 3 x 10^7 pfu. Analysis of ten random clones from this library by restriction digestion with *Bam*H I,

showed that the insert sizes averaged three kb (not shown). The theoretical size of a library required to represent the *Cowdria* genome can calculated (Sambrook *et al.*, 1989):

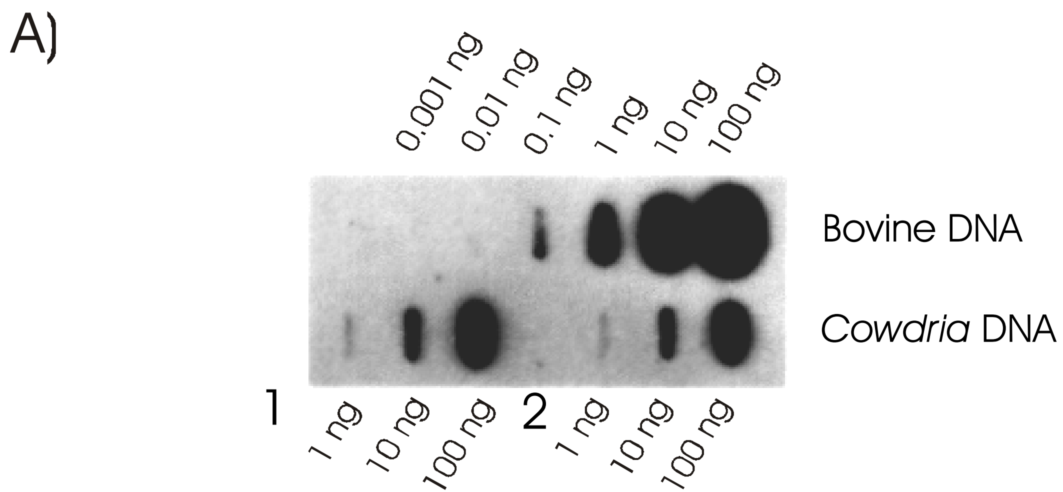
$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

P = (desired probability)

N = (necessary number of recombinants)

f = (fractional proportion of the genome in a single recombinant)

For a 99% probability, the theoretical required size to represent the approximately 1.7 x 10^6 bp *Cowdria* genome (de Villiers *et al.*, 2000a) with inserts of 3 kb is 2607 pfu, suggesting that the library contained at least 10^4 genome equivalents. Blue/white colour selection for recombinant clones showed that more than 98% of the clones contained inserts. Probing duplicate plaque lifts of the library with either the genomic *Cowdria* DNA or bovine DNA showed that roughly five out of 150 plaques contained bovine DNA (Figure 2.2.b). Since approximately three percent of the clones had bovine DNA inserts, this was regarded as representative of the starting material which had less than ten percent bovine DNA.



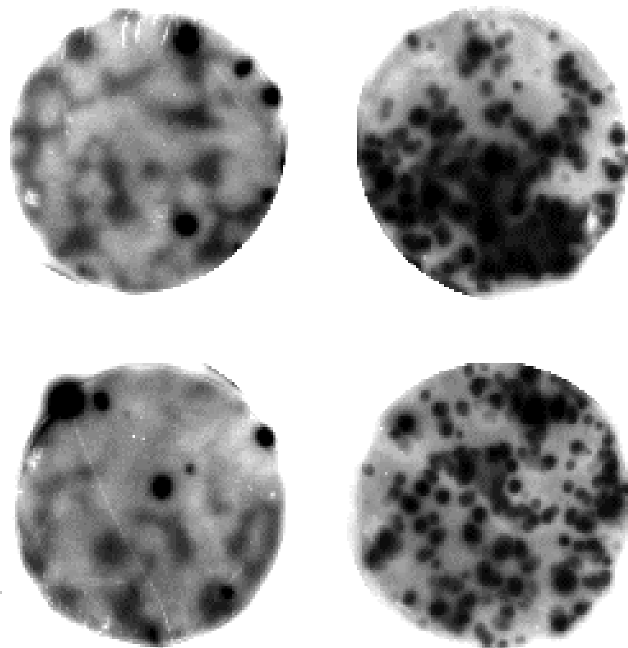


Figure 2.2. a) Determination of relative amount of bovine DNA in purified *Cowdria* DNA preparations. Tenfold serial dilutions of bovine and *Cowdria* DNA were probed with radiolabelled bovine DNA. The total amount of DNA per slot is indicated. Two separate *Cowdria* DNA purifications (1 and 2) are shown. b) Duplicate plaque lifts of the λ library hybridised with either a bovine or a *Cowdria* radiolabelled DNA probe.

2.3.2. Screening the library with antibodies.

λZAPII is an expression vector. In the case of prokaryotes like *Cowdria,* it is not feasible to isolate mRNA and make cDNA for expression libraries, since the mRNAs lack poly-A tails which allow easy cDNA synthesis. A library made from fragmented genomic DNA was therefore constructed instead. To determine whether any of the clones were expressing fusion proteins encoded by their inserts, the λ library was screened with antibodies after induction via the *lacZ* promoter. In addition to the available anti-MAP1 antibodies, anti-*Cowdria* antibodies were raised against disrupted, immunoaffinity-purified organisms. Figure 2.3 shows that both goat antisera, used at a dilution of 1/200, recognised *Cowdria* proteins. The anti-MAP1 serum (Figure 2.3.a) reacted with MAP1 and with additional proteins at this dilution, as reported by van Kleef *et al.* (1993). The anti-*Cowdria* serum (Figure 2.3.b) reacted strongly with MAP1, but also contained antibodies directed against a range of other *Cowdria* proteins. For antibody screening, 1000 plaques were plated per plate and at least 5000 plaques were screened to ensure coverage of the genome. In addition, duplicate lifts were made for easier detection of positive signals. No clones expressing proteins that reacted detectably with any of the antisera were isolated. Repeatedly, when potential positive plaques were picked, plated out and re-tested in a second round, all proved to be negative.

2.3.3. Isolation and characterisation of the *map*1 gene.

To establish whether the library did indeed contain *Cowdria* DNA, a *map1* DNA probe was used for screening. The PCR primers designed from the sequence of the *Cowdria* (Senegal) *map*1 gene amplified *map*1 from DNA from Welgevonden-infected cell culture material, as shown in Figure 2.4.a. The PCR product was radiolabelled and used to screen duplicate filters of the λ library.

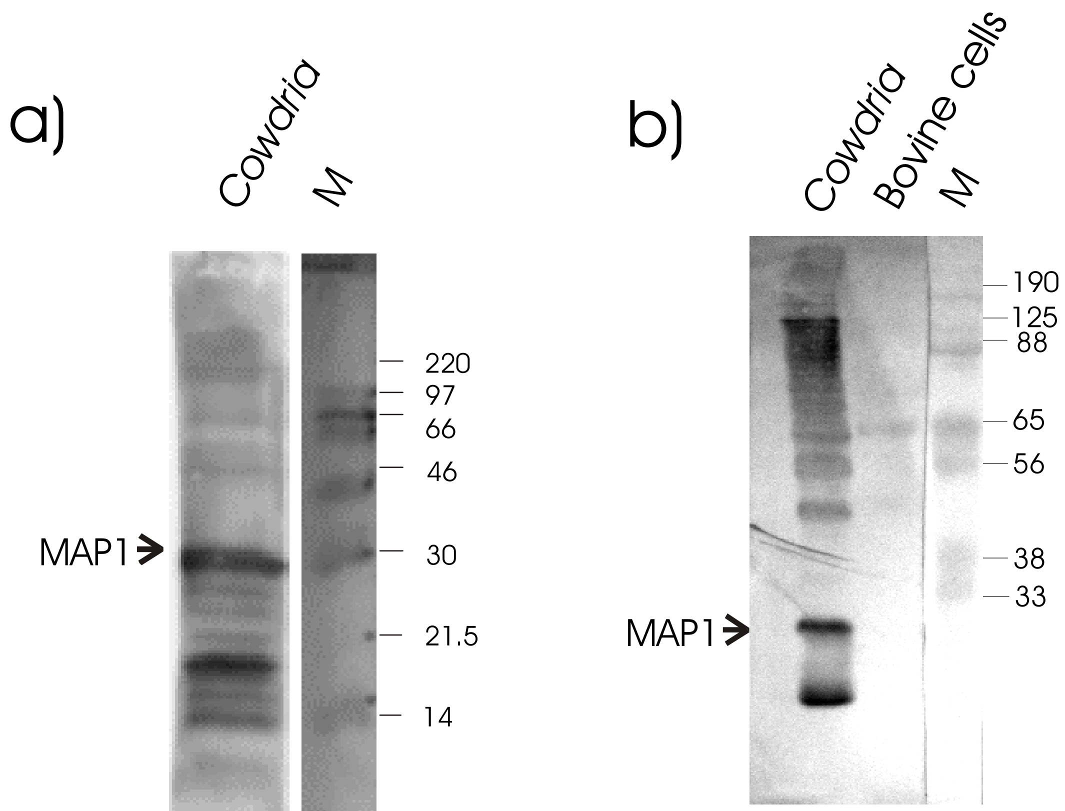


Figure 2.3. Reactivity of *Cowdria* proteins with goat antisera in western blots. a) Proteins reacting with goat antiserum (1/200 dilution) directed against PAGE-purified MAP1. Crude *Cowdria* extracts are in the left lane. b) Proteins reacting with goat antiserum (1/200 dilution) directed against sonicated, immunoaffinity-purified *Cowdria* organisms. Crude *Cowdria* protein extracts are in the left-hand lane and uninfected bovine cell proteins in the right-hand lane. Molecular weight markers (in kDa) are shown and the position of MAP1 is indicated by the arrow.

Seven positive clones were isolated from the 5000 plaques screened and from five clones the insert could be PCR-amplified using the above primers, indicating the presence of the full length *map*1 genes (Figure 2.4.b). Secondary screenings of these five clones with anti-MAP1 goat antibodies showed that one of the five clones was expressing MAP1. This was only evident because many copies of the same clone were present on the filter. This low level of expression was missed in the primary screening with antibodies (2.3.2).

To confirm that the clone was expressing MAP1, *E. coli* cell extracts containing the *map*1 clone were analysed by western blot (Figure 2.4.c). There was a reaction with a band similar in size to the native protein in the 31 kD region (lane 2), indicating that the expressed product of the *map*1 clone was not a fusion protein as might have been expected. IPTG induction did not enhance the level at which the gene product was expressed (lanes 4 to 8) since there was also a protein of the same molecular weight and intensity before IPTG induction (lane 4; 0 h). The slight increase in the signal is probably just due to the fact that after each hourly interval, more cells were present in the aliquot. This result suggests that the cloned insert included an in-frame *Cowdria* promoter which was driving the gene expression. Sequence analysis (not shown, done by Dr. E. P. de Villiers) confirmed that this was indeed the *map*1 gene and that it contained its own promoter.

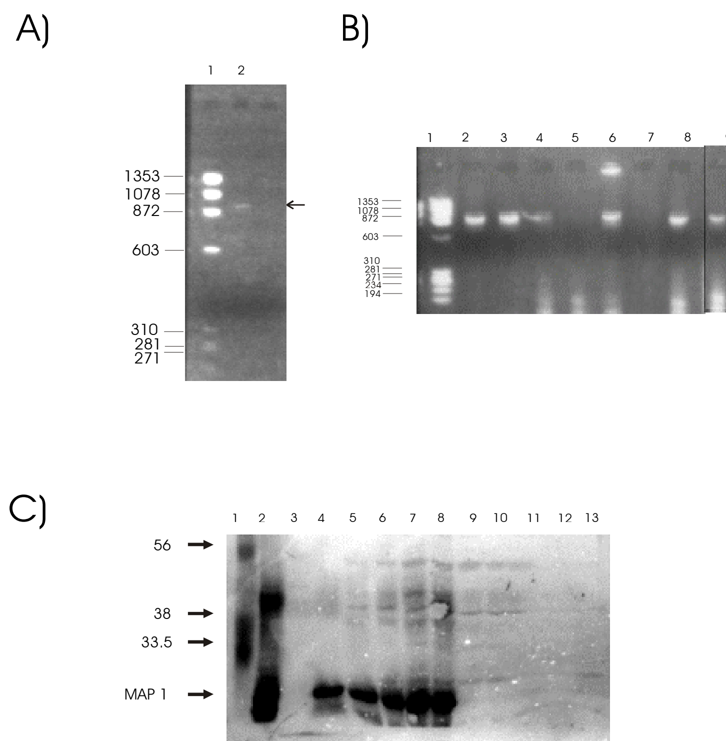


Figure 2.4. a) Agarose gel electrophoresis showing PCR amplification of the Welgevonden *map*1 gene (indicated by the arrow), lane 2. Molecular weight markers, lane 1 and the sizes in bp. b) The PCR products after amplifying the *map*1 positive λ clones. Lane 1: molecular weight markers, sizes in bp; lane 2-8: the seven λ clones; and lane 9: *Cowdria* genomic DNA as positive control. c) Western blot of the clone expressing MAP1. The proteins were allowed to react with anti-MAP1 antibodies. Lane 1: molecular weight markers in kDA; lane 2: native *Cowdria* proteins from infected bovine cells; lane 3: uninfected bovine cells; lanes 4 to 8: *E. coli* cells containing the *map*1 construct at 0, 1, 2 and 3 hours after IPTG induction; and lanes 9 to 13: control *E. coli* cells at 0, 1, 2 and 3 hours after IPTG induction.

2.4. DISCUSSION.

The molecular genetic analysis of *Cowdria* has been limited by factors such as the difficulty of obtaining *Cowdria* DNA free of host DNA contamination, the lack of a good expression library and because there are few probes or other molecular tools available. To address the first problem, a method to isolate *Cowdria* from the host cell using positive selection immunoaffinity chromatography with an anti-MAP1 antibody was developed (Brayton *et al.,* 1999). This time-consuming technique results in low and variable yields. Despite these drawbacks, DNA isolated from organisms purified in this way contained less than ten percent bovine DNA, a significant improvement over previous isolation methods (Ambrosio *et al.*, 1987) and similar to the preparations containing one percent bovine DNA obtained by Perez *et al.* (1997). Publications describing other *Cowdria* libraries, make no mention of the bovine DNA content (Lally *et al.,* 1995; Mahan *et al*, 1994; Barbet *et al.*, 2001). The purified material was used to make a λ expression library containing approximately 10^4 *Cowdria* genome equivalents. More than 98% of the clones contained inserts, with an average size of 3 kb and the level of clones with bovine DNA inserts being about three percent.

Antibodies were raised against sonicated organisms so as to obtain antibodies against all *Cowdria* proteins, not just those 'seen' by an animal's immune system when infected. Neither these, nor anti-MAP1 antibodies, were able to select any λ clones from the library. This was unexpected, since Lally *et al.* (1995) screened a similar λ library with goat antiserum against the Ball 3 isolate of *Cowdria* and in this way isolated the GroES- and GroEL-encoding genes. Nevertheless, these were the only protein-encoding genes isolated. In other studies, plasmid libraries made from the Crystal Springs, Highway and Gardel isolates were screened using immune sera and detected with ^{125}I-labelled protein G which enabled protein encoding-genes to be selected (Mahan *et al.*, 1994; Barbet

et al., 2001). Radiolabelled protein G is a very sensitive screening method compared to the colourimetric detection system which was nonetheless sensitive enough to detect *Cowdria* proteins in western blots. It should be noted, however, that the library constructed with DNA from the Gardel isolate, which is closely related to Welgevonden, yielded fewer clones than the Highway isolate-derived library.

Since no clones could be isolated with any of the available antibody probes, the λ library was tested by screening with a DNA probe. A *map*1 probe selected the *Cowdria map*1 (Welgevonden) gene. The λZAPII vector allows *in vivo* excision to form a phagemid containing the cloned insert. After *in vivo* excision, the *map*1 plasmid clone caused overnight cultures to lyse, suggesting that MAP1 is toxic to *E. coli* when it is produced intracellularly. It would appear that in the λZAPII phage system, the *map*1 gene was expressed at levels which were low enough so that host cell growth could proceed until lysis occurred as a result of normal phage activity. The outer membrane protein, rOMP B, of *Rickettsia rickettsii* was also found to be unstable in plasmids but could be cloned into a λ vector (Gilmore *et al.*, 1991). In an expression library, one would expect the cloned genes to be expressed as fusion proteins with β-galactosidase. In the case of the *map*1 clone, the gene was expressed as a non-fusion protein, transcription presumably being driven by its own promoter. Western blot analysis showed a protein of 31 kDa, the size of the native protein. IPTG induction had no effect on its expression levels. Sequence data also indicated that the clone contained a prokaryotic promoter sequence upstream of the *map*1 start codon. It has previously been found that promoters of cloned genes of the rickettsia *Anaplasma marginale* were active in *E. coli* (Barbet and Allred, 1991; Visser *et al.*, 1992). It is therefore perhaps not unexpected that the *map*1 *Cowdria* promoter was active.

The primary aim of making this library was to isolate genes encoding proteins that reacted with antibodies of interest. This goal was not realised. A possible explanation is that the inserts were not derived from mRNA (open reading frames) and very few clones actually expressed fusion proteins that could be induced to detectable levels. Some, like the *map*1 clone, were expressed from their own promoters, but too little protein was expressed to be detected in the λ plaque antibody screening system. Nevertheless, since the *map*1 gene could be isolated from the library, it remains a valuable source of cloned *Cowdria* DNA. It represents the entire *Cowdria* genome and contains few bovine DNA clones. This library is being used as a source of DNA for a *Cowdria* genome sequencing project (Molecular Biology Division, Onderstepoort Veterinary Institute) and in the work that follows, it provided good quality *Cowdria* DNA.

CHAPTER 3

CONSTRUCTION OF A FRAGMENTED-GENOME PHAGE DISPLAY LIBRARY.

3.1 INTRODUCTION.

The display of peptides on the surface of phages has proven to be a powerful method for selecting and identifying peptide sequences recognised by antibodies (1.2.3). The fragmented-gene approach (1.2.3.7) allows authentic epitopes on specific proteins to be located. It was decided to determine whether this approach could be extended to a genome, rather than just a single gene. Fragmented-genome phage displayed libraries had previously been made from the DNA of *Staphylococcus aureus* (Jacobsson and Frykberg, 1995) and *E. coli* (Palzkill *et al.*, 1998) and useful proteins identified (1.2.3.8) by screening these 'peptidomes'. Affinity selection (or panning, 1.2.3.3) is potentially much more sensitive than screening λ libraries. It is also less cumbersome and requires smaller amounts of antibodies. These can be affinity-purified, on immunoblotted proteins so that only a subpopulation of the antibodies in an antiserum is used (Parmley and Smith, 1988; du Plessis *et al.*, 1995).

This chapter describes the construction of a fragmented-genome phage library displaying peptides derived from *Cowdria* DNA (Figure 3.1). For this, the fUSE2 phage vector was used (Parmley and Smith, 1988). DNA cloned into the RF of the phage produces pIII N-terminal fusion proteins. Thus, all five copies of pIII are fused with the foreign peptide sequence. The open reading frame of pIII must remain intact for the coat protein to be produced since it is essential for phage particle production and *E. coli* infection. Some clones can maintain the open reading frame, but not encode the natural reading frame. These are termed 'meaningless' clones, in contrast to 'meaningful' clones

that encode the natural peptide sequence (Smith, 1985).

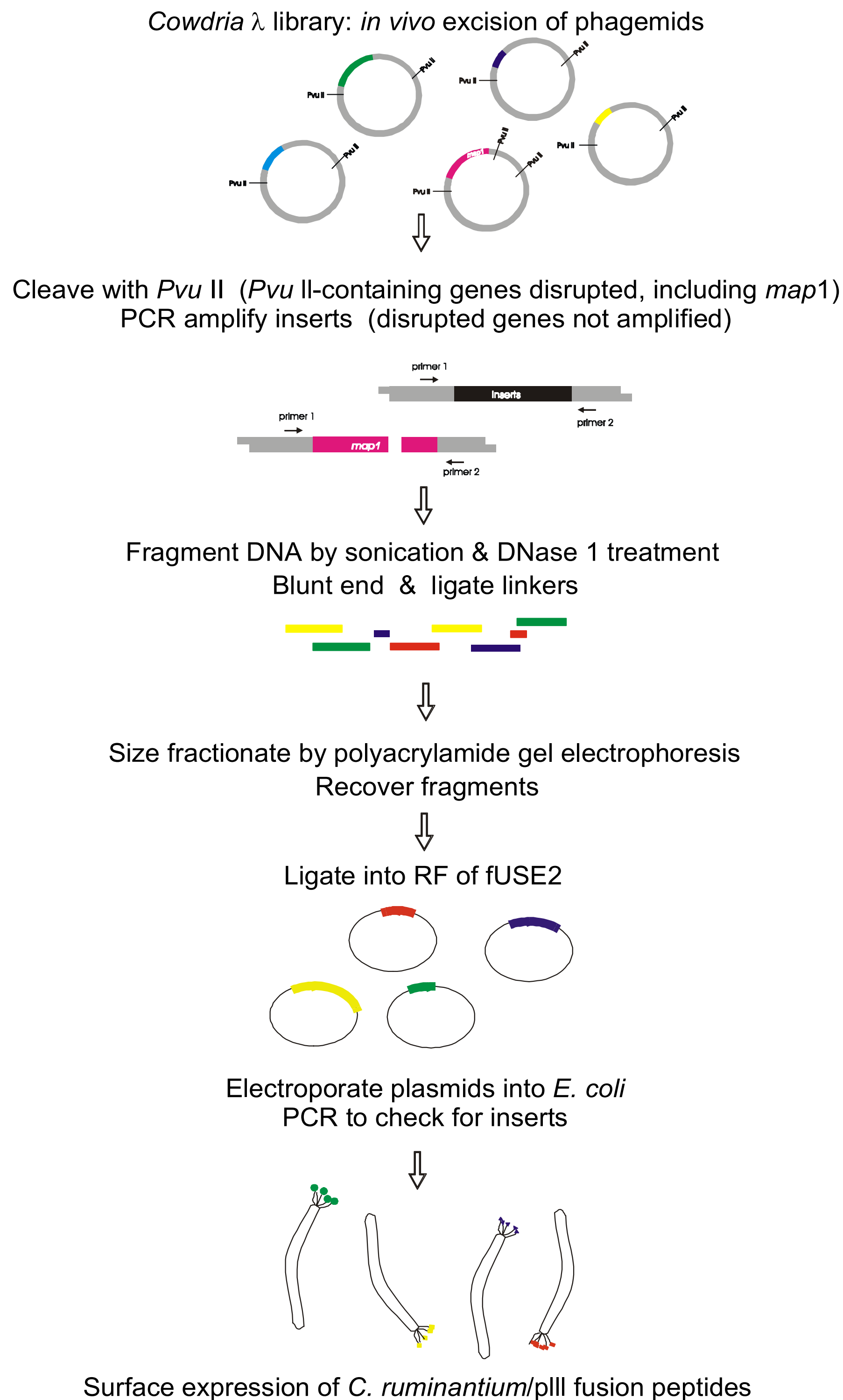


Figure 3.1. Schematic representation of the construction of the *Cowdria* fragmented-genome phage display library.

3.2. MATERIALS AND METHODS.

3.2.1. Preparation of the *Cowdria* DNA fragments.

Due to the limited availability of *Cowdria* cell culture material when this work was being done, the λ ZAPII library inserts were used as source of *Cowdria* DNA to construct the phage display library. An aliquot, equivalent to 6.8 x 10^7 pfu of the λ ZAPII *Cowdria* library (Chapter 2), big enough to represent the genome, was mass excised *in vivo* (2.1.10). The resulting phagemids were amplified according to the supplier's instructions (Stratagene) and purified by chromatography on Nucleobond columns (Macherey Nagel). The plasmid DNA was linearised with *Pvu* II (Amersham), before the inserts were amplified with the T7 (GTAATACGACTCACTATAGGGC) and the T3 (GAAATTAACCCTCACTAAAGGG) primers by PCR with the Expand HiFi enzyme mix (Roche) as instructed. Primers and free nucleotides were removed using the Prep-a-gene kit (Biorad).

The DNA fragments were prepared essentially as described by du Plessis and Jordaan (1996) with minor adjustments. The PCR amplified *Cowdria* DNA was first sonicated at full power for 100 s using a Branson sonicator to yield fragments of 200-1500 bp before DNase 1 treatment. Small pilot DNase 1 (Roche) digestions were performed with 1 µg sonicated DNA samples in DNase 1 buffer (50mM Tris pH7.6, 1mM $MnCl_2$, 0.1mg/ml BSA) containing 5 to 0.3 U/ml DNase 1 for 10 minutes at 15°C. The reactions were stopped with a sixth volume of stop solution (70% glycerol, 75mM EDTA, 0.3% bromophenol blue) and separated on a 2.5% agarose gel (SepRate, Amersham). The concentration of DNase 1 giving the optimal size was used to digest 4 µg DNA. These reactions were stopped with 25 mM EDTA and the DNase 1 was removed by phenol extraction (2.2.3). The termini of the fragments were made blunt-ended with 30 U T4 DNA polymerase (Promega) in repair buffer (40 mM Tris/HCl pH8, 10 mM ammonium sulphate, 10 mM β-mercaptoethanol, 5 mM $MgCl_2$, 0.5 mM EDTA, 100 µM of each dNTP) at 15°C for 1 h before 10 U Klenow (Promega) was

added and incubated for a further 30 minutes at 37°C. The enzymes were removed with Strataclean resin (Stratagene) as instructed and the DNA pellet resuspended in distilled water. To create enzyme sites on the inserts for cloning into the fUSE2 vectors, the fragments were ligated to phosphorylated *Bam*H I linkers (CGCGGATCCGCG), which were prepared by heating at 95°C for 5 minutes and allowing the mixture to cool down slowly to room temperature to anneal. Annealed linkers were ligated to the blunt-ended DNA fragments with a Rapid DNA Ligation Kit as instructed (Roche), followed by digestion with *Bam*H I to create compatible ends with the vector. The digested fragments were separated on a 5% polyacrylamide gel (Sambrook *et al.*, 1989), stained with 0.5 µg/ml ethidium bromide and fragments in the 50-300 bp range were recovered by crushing the gel and incubating in 2 volumes extraction buffer (0.5 M ammonium acetate,10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS) at 37°C overnight with shaking (Sambrook *et al.*, 1989). After centrifugation the supernatent was filtered (0.22µ, Dynagard) to remove traces of acrylamide. The extraction was repeated twice with fresh buffer with 2 h incubations each, before the DNA was ethanol precipitated and resuspended in dH_2O.

3.2.2. Phage display vector preparation.

The fUSE2 phage vector (gift from G. Smith; Parmley and Smith, 1988) was used to construct the library. *E. coli* K91 infected with fUSE2 was grown in LB broth containing 40 µg/ml tetracycline overnight and double stranded replicative form (RF) was isolated from the cells using the standard CsCl plasmid isolation method (Sambrook *et al*., 1989; Smith, 1992). The fUSE2 RF was linearised with *Bgl* II and dephosphorylated with calf intestinal alkaline phosphatase (CIAP, Promega). The enzymes were removed with Strataclean resin and the DNA pellet resuspended in distilled H_2O (d H_2O), ready for ligation with the *Cowdria* DNA fragments (Sambrook *et al*.,1989).

3.2.3. Ligation and electroporation.

Linearised fUSE 2 was ligated with the size selected *Cowdria* DNA fragments using T4 DNA ligase (Roche). The enzyme was removed using Strataclean resin, the DNA was ethanol-precipitated and resuspended in minimal amounts of dH_2O for electroporation (BTX *E. coli* TransPorator) at 1300V, into fresh electrocompetent MC1061 cells (Appendix C). The cells were incubated in LB broth containing 0.02 µg/ml tetracycline for 2 h before plating dilutions on LB agar plates containing 40 µg/ml tetracycline to titrate transformants and the rest amplified in LB broth containing 40 µg/ml tetracycline. Two separate ligations were electroporated into 30 aliquots of MC1061 cells (50 µl) to yield a library with a total of 2.6×10^7 colony forming units (CFUs).

3.2.4. PEG precipitation of phages.

Phages amplified in *E. coli* K91 or MC1061 in LB medium or on LB plates were purified by PEG precipitation (Smith, 1992). Colonies on plates were scaped off into Tris buffered saline (TBS). Overnight cultures were centrifuged at 3000 g for 10 minutes twice to remove the bacteria, 0.15 volume PEG/NaCl (16.7%/3.3 M) was added to the supernatant and mixed well. After at least 4 h at 4°C the phages were collected by centrifugation at 8000 g for 30 min. All the PEG was removed by discarding the supernatant and another brief centrifugation to collect remaining PEG and removing that with a pipette. The pellet was resuspended in TBS in approximately 10% of the original volume and precipitated a second time. The pellet was resuspended in a small, convenient volume in TBS containing 0.02% sodium azide (NaN_3). The titre of the phage suspension was determined (3.2.5) or the concentration of phage particles was determined by spectophotometry as described by Smith (1992); an absorbance of 30 at 269 nm corresponds to 1 mg/ml DNA, 6 mg/ml protein or 2×10^{14} phages/ml.

3.2.5. Titration of phage solutions.

Starved *E. coli* K91 cells or mid log cultures (Appendix C) were infected with phages to determine the titre (the number of viable phages) in a solution. Tenfold serial dilutions of phages were made in TBS of which 10 µl was added to 10 µl starved K91 cells (or 100 µl mid log cells) and allowed to infect for 10 minutes at room temperature. LB broth (1 ml) containing 0.02 µg/ml tetracycline was added and incubated for 1 h at 37°C and subsequently 100 µl was plated on LB agar (40 µg/ml tetracycline) and incubated overnight at 37°C. The colonies were counted the next morning.

3.2.6. PCR of fUSE2 bacterial colonies.

The size of inserts was determined by PCR using primers that flank the fUSE2 cloning site (Wang *et al*, 1995). Colonies were picked with a pipette tip into 20 µl dH_2O, boiled for 5 min and put on ice. The samples were microfuged for 5 min and 13.75 µl of the supernatant transferred to new tube. The PCR was performed as instructed for Taq enzyme (Promega) in the presence of 2.5 mM $MgCl_2$/µl and 0.4 pmol primer/µl. Primers pIII5' (GGTTGGTGCCTTCGTAGT) and pIII3' (CCATGTACCGTAACACTG) were used. The reaction conditions were 94°C for 30 s, 55°C for 30 s, 72°C for 30 s for 25 cycles and 1 cycle of 4 min at 72°C in a GeneAmp PCR system 2400 (Perkin Elmer). Products were analysed by electrophoresis on a 1.5% agarose gel.

3.2.7. Sequencing of fUSE2 clones.

Individual phage clones were grown in 5 ml LB broth containing 40 µg/ml tetracycline overnight at 37°C. The bacterial cells were removed by centrifugation for 5 minutes and the phages in the supernatent were PEG precipitated once for 15 minutes (3.2.4). The phage pellet was resuspended

in 100 µl TE, mixed with 200 µl phenol/chloroform/isoamyl alcohol (25:24:1), vortexed and left at room temperature for 15 minutes before vortexing again. The phases were separated by centrifugation for 5 minutes, the aqueous phase was removed and the phage DNA concentrated by ethanol precipitation. The pellets where resuspended in 7 µl TE and used in a sequencing reaction. The Sequenase 2 kit (Amersham) was used as instructed with the fUSE2 sequencing primer, CCCTCATAGTTAGCGTAACG (Parmley and Smith, 1988). The sequencing reactions were separated on 6% acrylamide gels (Sambrook *et al.* 1989), dried and exposed to Hyperfilm (Amersham). The sequences were read and analysed with the Staden (Staden, 1994) and Wisconsin GCG (Anon, 1994) software packages.

3.3. RESULTS.

3.3.1. Preparation of the *Cowdria* DNA fragments.

When the library construction was initiated, limited *Cowdria* cell culture material was available. The λ ZAPII library (Chapter 2) was therefore used as a source of *Cowdria* DNA. An aliquot of the λ phages, representative of the genome, was mass excised *in vivo* and the phagemid DNA isolated. Inserts ranged from 600 bp to 7 kb, while the phagemid itself was 2960 bp in size (Figure 3.2a and b). This meant that if the inserts were simply to be excised from the phagemid by restriction digestion, it would not have been possible to remove vector DNA from the *Cowdria* DNA (Figure 3.2a). To obtain DNA with minimal amounts of vector DNA, the inserts were thus amplified by PCR after the phagemids were linearised with *Pvu* II. When intact phagemids were used, the size of the inserts increased, possibly due to run-through amplification of vector sequences (not shown). The amplified DNA inserts were fragmented by sonication to obtain fragments in the range of 200

to 1500 bp (Figure 3.2c). Subsequently DNase I digestion was used to generate fragments of 50 to 200 bp (Figure 3.2c). These fragments were made blunt-ended, ligated to *BamH* I linkers and digested with *BamH* I to enable cloning into the *Bgl* II site of fUSE2.

3.3.2. Construction of a *Cowdria* fragmented-genome phage display library.

Ligation reactions were electroporated into MC1061 cells to yield a library of 2.6 x 10^7 CFUs. To confirm the presence of inserts, PCR was performed on individual colonies using primers that flanked the cloning site. A total of 62% of the colonies (25 out of 40) contained inserts (example of PCR, Figure 3.2d). Ten clones with inserts were sequenced, all of which contained different sequences ranging from 17 to 110 bp, with an average size of about 80 bp.

If a conservative estimate of insert size of 45 bp is taken, then 1,7 x10^5 clones are required for the 1,7 x10^6 bp *Cowdria* genome to be represented with a 99% probability (2.3.1). In the case of the fUSE2 vector, however, not all cloned inserts produce phages. The fragments were randomly fragmented and could therefore be cloned in one of three reading frames at either end of each fragment as well as in one of two directions. Thus the library needed to contain 18 times (3 x 3 x 2) more clones (18 x 1,7 x10^5 = 3,06 x 10^6 CFUs) to ensure the likelihood that a sequence would be expressed in the correct reading frame. The *Cowdria* fragmented-genome library contained 2.6 x 10^7 CFUs of which 62% contained inserts, thus in effect 1.67 x 10^7 CFUs. Consequently the resulting library was five times (1.67 x 10^7/3.06 x 10^6) that of the minimum required size to represent the genome.

It did not however, represent the entire genome, since together with any gene containing a *Pvu* II site, the probability of *map*1 sequences being present was greatly reduced by the cloning strategy;

the *map*1 gene has an internal *Pvu* II site and after mass *in vivo* excision, the target DNA was amplified by PCR from the resulting plasmid DNA after *Pvu* II digestion. Since the *map*1 gene has been cloned and well characterised, and the aim was to identify new epitopes or proteins, a library without this gene was not considered to be a limitation.

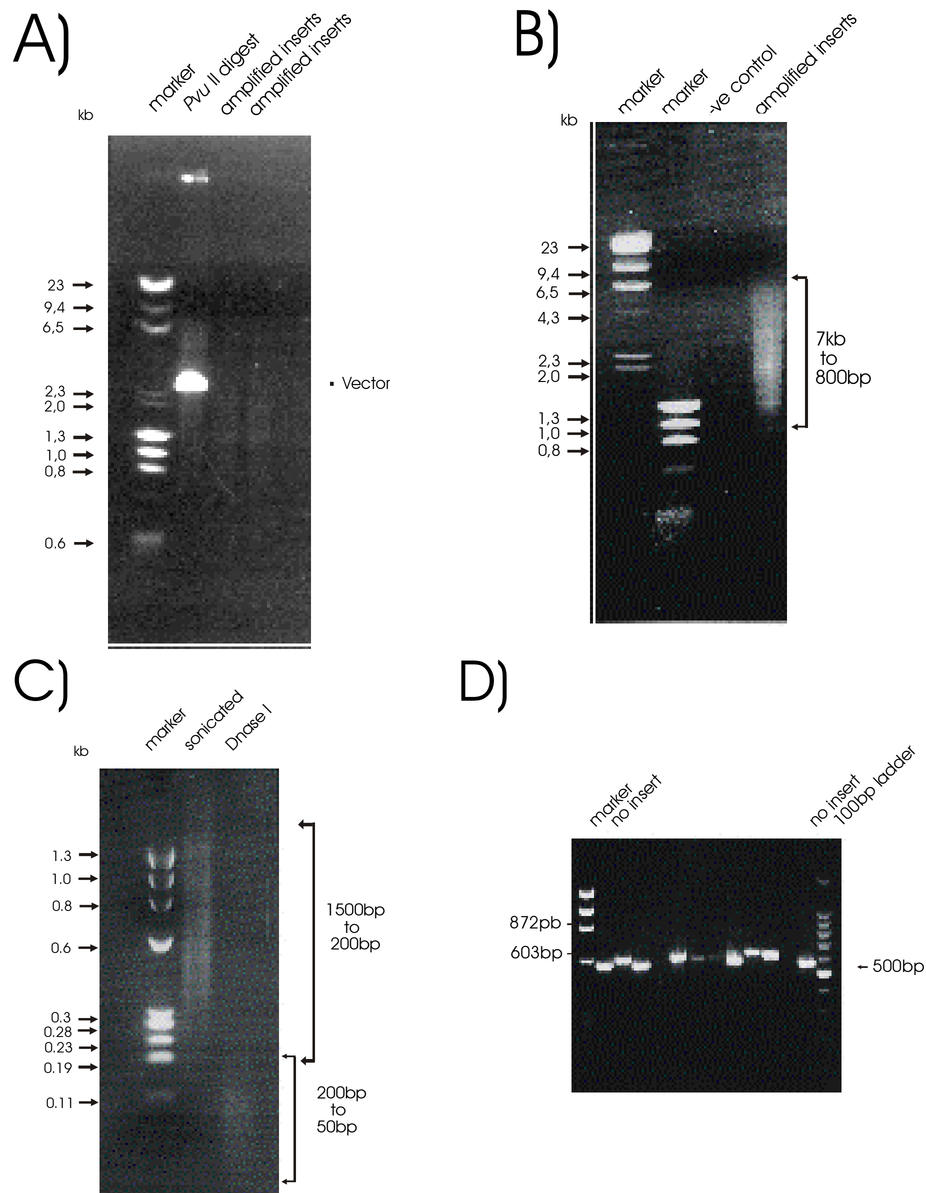


Figure 3.2. Agarose gel electrophoresis showing stages of *Cowdria* DNA fragment preparation. a) *Pvu* II digested phagemid DNA and *Cowdria* DNA inserts amplified from *Pvu* II digested phagemids, b) PCR amplification of the *Cowdria* DNA inserts from the λ ZAPII library, c) fragmentation of the DNA inserts (from b) by sonication and DNase 1 digestion and d) PCR of randomly chosen clones from the fragmented-genome phage display library to show the size of the DNA inserts.

3.4. DISCUSSION.

A filamentous phage library displaying peptides derived from random fragments of the *Cowdria* genome was constructed. It contained at least five times more clones than the theoretical size required to be representative of the genome. The aim of this study was to investigate whether the fragmented-genome approach would be useful for the identification of epitopes and/or proteins of organisms of which the genome sequence was unknown. If so, this would be an additional molecular tool to aid in *Cowdria* research.

In the fUSE2 vector, the cloned inserts generate pIII fusions. Only when the pIII reading frame is maintained, will phage particles be produced. One out of six clones that produce phages will express the authentic peptide sequence and are thus 'meaningful' clones (Parmley and Smith, 1988). The remaining five clones are 'meaningless' and are a source of additional peptide sequences that are not expressed in the organism from which the DNA was derived. The *Cowdria* phage display library contained 9.2 x 10^5 meaningful clones (1.67 x 10^7 CFUs x 1/18) and 4.6 x 10^6 meaningless clones (1.67 x 10^7 CFUs x 5/18).

With the work done earlier on fragmented-gene libraries, this extra diversity did not seem to have an effect on the clones selected, since authentic peptide sequences were always identified. This was probably because only target genes of 1,6 kb (NS1 of BTV; du Plessis *et al.*, 1995) up to 3kb (VP2 of African horsesickness virus; Bentley *et al.,* 2000) had been used as input DNA. The *Cowdria* genome was 566 times (1,7 x10^6 bp / 3 x10^3 bp) larger than the VP2 gene, its input DNA was therefore much more diverse. In the VP2 libraries, there was a total of 8.4 x 10^4 meaningless clones and none of these clones were selected (Bentley *et al.*, 2000). Petersen *et al.* (1995) also claimed to 'have never detected sequences derived from DNA inserts other than the epitope-containing

region.'

In papers describing the fragmented-genome libraries made from *Staphylococcus aureus* (Jacobsson and Frykberg, 1995) and *E. coli* (Palzkill *et al.*, 1998), no mention was made of the required size of the libraries for genome coverage, so comparisons with the *Cowdria* library could not be made. Both groups reported, though, that only authentic sequences were identified. These were IgG binding domains of protein A and known epitopes on the RecA protein respectively. While some of the clones containing protein A encoding DNA inserts were not in frame with pIII, they represented known binding regions. The authors suggest 'ribosome slippage' caused the correct protein to be expressed (Jacobsson and Frykberg, 1995). Even though the *E. coli* library was screened with polyclonal antibodies, no mention was made of irrelevant sequences being identified (Palzkill *et al.*, 1998). It was thus anticipated that authentic peptide sequences would be identified from the *Cowdria* phage display library, without prior knowledge of these proteins (Chapter 4). This assumption was to prove unfounded (Chapter 4 and 5).

CHAPTER 4

IDENTIFICATION OF *COWDRIA* EPITOPES AND PROTEINS.

4.1. INTRODUCTION.

The construction of a fragmented-genome phage library displaying *Cowdria* peptides was described in Chapter 3. This chapter describes how this library was used to identify epitopes on *Cowdria* proteins.

Panning exploits the interactions between molecules to select binding entities from an excess of non-binding entities (section 1.2.3.3, Figure 1.3). The ligand used for panning can be any molecule capable of interacting with a second entity, such as proteins, carbohydrates or antibodies (Smith, 1985; Wells, 1996; Devlin *et al.*, 1990). This study used antibodies to select binding peptides from the library. Antibodies are capable of binding a vast array of epitopes and are called specific when they can discriminate between two different antigens/epitopes (section 1.1). Antibodies made against several electrophoretically separated *Cowdria* proteins were used to identify the antigenic determinants contained on these proteins and then an attempt was made to identify the genes encoding the proteins. If the genome sequence is available, peptide sequences can in theory be compared with it to identify the genes that encode the proteins of which they form a part. When this work was initiated, the *Cowdria* genome sequence was not available. Accordingly, the sequences were 'compared' with the genome (in the form of the λ library; Chapter 2) by hybridising oligonucleotides encoding the peptide sequences. Subsequently, some *Cowdria* genome sequence data have become available (Molecular Biology Division, Onderstepoort Veterinary Institute). The peptide sequences could therefore be compared with this limited database to confirm and expand

on the results obtained with the λ library screenings. Antibodies from three sources were used in this study.

MAP1 is a major antigenic protein that was originally identified by western blot studies (section 1.3.8; Figure 2.3.a; van Kleef *et al.*, 1993). Its encoding gene has been cloned and characterised (section 1.3.9). This gene was, however, probably not represented in the filamentous phage library owing to the way in which it had been constructed (section 3.3.2). Panning with antibodies against MAP1 could thus be used to determine whether or not fragments of this gene were present. In addition, since *map*1 is thought to be part of a multigene family (section 1.3.9), such panning could therefore possibly identify common or cross-reactive epitopes on proteins in the MAP1 family. Perez *et al.* (1998) identified proteins from 23 to 29 kDa that varied both in size and antigenicity between different isolates of *Cowdria*. They were identified by cross-absorbing different antisera and antigens (section 1.3.8). Rabbit serum was raised against a range of such proteins obtained from the Gardel isolate. The Welgevonden and Gardel isolates are similar antigenically (Perez *et al.,* 1998) and since the rabbit antiserum reacted with Welgevonden isolate proteins, it was considered feasible to use it for panning the Welgevonden isolate-derived phage display library. Protein fractions of the latter isolate have been examined for their ability to stimulate lymphocyte proliferation *in vitro.* These studies identified a set of proteins in the size range of 11 to 23 kDa. They may thus play a role in protection against heartwater (section 1.3.10; van Kleef *et al.,* 2000). Accordingly, rabbit antibodies were also raised against these proteins (personal communication, Dr. M. van Kleef).

With aim of identifying antigenic determinants on the abovementioned proteins and their encoding genes, the *Cowdria*-derived phage display library was panned with the anti-MAP1 goat serum, the rabbit antisera against the Gardel isolate proteins of Perez *et al.* (1998) and the serum raised against

the 11 to 23 kDa Welgevonden isolate proteins identified by van Kleef *et al.* (2000).

4.2. MATERIALS AND METHODS.

4.2.1. Antisera.

Goat antiserum to *Cowdria* MAP1 that had been separated by SDS-PAGE was provided by Dr. M. Van Kleef (section 2.2.6). Rabbit serum against SDS-PAGE separated variable immunodominant proteins of *Cowdria* (Gardel isolate) was provided by A. Bensaid (Perez *et al.,* 1998) and against SDS-PAGE separated potential protective antigens of the Welgevonden isolate was provided by Dr. M. van Kleef (Immunology Division, Onderstepoort Veterinary Institute).

4.2.2. Affinity purification of antibodies.

Affinity purification of specific antibodies was carried out essentially as described (du Plessis *et al.*, 1995). *Cowdria* proteins were separated by 12% or 15% preparative SDS-PAGE (sections 2.2.2, 2.2.11 and Appendix B) and transferred to Immobilon P membrane (Millipore) in Towbin buffer (Appendix B). After blocking the membrane with PBS containing 2% milk powder (PBS2%MP), the appropriate band was located by immunostaining side strips of the membrane with the antiserum diluted in PBS containing 1% milk powder (PBS1%MP). The excess antibody was removed by washing three times with PBS containing 0.05% Tween (PBST) and detected with peroxidase conjugated secondary antibodies diluted in PBS1%MP for 1 hr. Excess conjugate was washed off as above and detection was as in section 2.2.7 or with SuperSignal chemiluminescence substrate (Pierce), in which case the bands were detected with a Lumi-Imager (Roche). The membranes were

cut to contain only the wanted protein bands and exposed to the antiserum. After washing the strips, the antibodies were eluted with 1 ml 0.2 M glycine-HCl buffer, pH 2.8 for 10 minutes at room temperature and neutralised with 50 µl 2 M Tris pH 8.9. For the MAP1 antibodies, the *Cowdria* proteins were separated on 12% gels, the antiserum was diluted 1/50 and the anti-goat IgG peroxidase conjugate (Zymed) diluted 1/2000 was used for detection. The rabbit antibodies against the variable immunodominant proteins of the Gardel isolate were also eluted from proteins separated on 12% gels, diluted 1/50. For the smaller range of Welgevonden isolate proteins, 15% gels were used to separate the separate the proteins, diluted 1/20. The anti-rabbit IgG peroxidase conjugate (DAKO) was diluted 1/2000. The concentration of the eluted antibodies was determined by spectophotometry (IgG: 1 A_{280} = 1.35 mg/ml; Harlow and Lane, 1988). The reactivity of the eluted antibodies with the protein used for elution was reconfirmed by western blotting.

Total IgG was isolated according to Clark and Adams (1977) with minor adjustments. Rabbit antiserum was diluted 1/10 in H_2O to a volume of 5 ml and an equal volume of saturated ammonium sulfate was added. After gentle mixing for 1 h at room temperature the precipitate was collected at 10 000 rpm for 10 minutes. The pellet was resuspended in 1 ml ½ strength PBS and dialysed against ½ strength PBS with three buffer changes. After dialysis the antibodies were batch purified with DE52 by adding 100 µl DE52 slurry (Appendix A) to 1,5 ml dialysed antibody solution, incubating for 10 minutes followed by centrifugation in a microfuge. The supernatant was transferred to a new tube and quantified by spectophotometry as above.

4.2.3. Affinity selection of phage displayed peptides (panning).

Affinity selection of fusion phages by panning was essentially as described (Smith, 1985; Wang *et al.,* 1995, Figure 1.3). In the first round of panning antibodies were added to a Petri dish (Falcon

1007) at a concentration of 10 to 20 µg/ml, diluted in 4 ml 0.1 M $NaHCO_3$ (pH unadjusted) and allowed to adsorb for 1.5 h at 37°C or overnight at 4°C. All further incubation steps were carried out at 37°C with gentle rocking. The dish was washed three times with TBS containing 0.05% Tween 20 (TBS-T.05) and blocked with 2% BSA in TBS for 1 h. The blocking solution was removed and approximately 10^{10} CFU heat inactivated, UV killed f1 phages (Smith, 1992) was added, to block those antibodies binding to phage proteins, and incubated for 1 h. The unbound f1 phages were removed by washing five times with TBS containing 0.5% Tween 20 (TBS-T0.5) and five times with TBS-T.05. Approximately 10^{11} CFUs of the library was added to the dish and incubated for 1h. The unbound phages were removed by washing five times with TBS-T0.5, five times with TBS-T.05 and once with TBS to remove all the Tween 20. The bound phages were eluted with 800 µl elution buffer (0.1 N glycine-HCl, pH 2.5 containing 1 mg/ml BSA) at room temperature for 10 min and neutralised by transfer to a tube containing 48 µl 2 M Tris pH 8.9. The eluate was concentrated to 200 µl by filtration (100 000 MW, Millipore), added to 200 µl starved K91 cells (Appendix C) and allowed to infect for 10 minutes at room temperature. LB broth containing 0.02 µg/ml tetracycline was added and incubated for 1 h at 37°C. Dilutions of this mixture was plated on LB agar (40 µg/ml tetracycline) to determine the number of phages (represented as a CFU) released after the round of panning. The rest were plated on a Bioassay agar plate (24 cm x 24 cm; Amersham) and incubated at 37°C overnight. The cells were scraped off the agar surface, the cells collected by centrifugation and the phages in the supernatent precipitated twice with PEG (section 3.2.4). One tenth of the output of the first round was subjected to a second round of panning in a 96 well polystyrene microtitre plate (Corning) in a volume of 50 µl, elution was performed with 40 µl elution buffer and neutralised with 280 µl TBS:2 M Tris (69:1 v/v). In some cases the antibody concentration was lowered for the third or fourth rounds of panning.

After the final round of panning, single colonies were picked, the DNA inserts sequenced (section

3.2.7) and analysed. The codon usage of the inserts was compared with the relative synonymous codon usage (RSCU) values determined by de Villiers (2001) for *Cowdria*. Codons with RSCU values of 1.0 and higher were considered positive and the expected codon usage was given as a percentage of positive codons in the insert.

4.2.4. Antigenic reactivity of the selected fusion peptides.

Recognition of the phage-expressed peptides by antibodies was confirmed in ELISA and western blots. Representative phages were amplified in 100 ml LB broth and PEG precipitated twice (section 3.2.4). The ELISA was performed in a microtitre plate (Corning Easy Wash), coated with 10 µg/ml phage diluted in TBS-T.05 at 37°C for 1 h, blocked with 5% milk powder in TBS-T.05 for 30 min before adding MAP1 antiserum diluted 1/200 in TBS-T.05 containing 1% milkpowder. The plate was washed five times with TBS-T.05 and incubated with HRP-conjugated-anti-goat immunoglobulin antibodies (Zymed) diluted 1/2000 in TBS-T.05 containing 1% milkpowder for 1h. After washing, H_2O_2 substrate and o-phenylenediamine chromogen was added and the absorbance read as described (du Plessis *et al.*, 1995).

Representative phages from the pannings were also tested on western blots. Phage proteins (20 µg) were separated by 10% SDS-PAGE (Appendix B) and transferred to Immobilon P membrane (Appendix B) and treated as in 4.2.2. The phage proteins were exposed to MAP1 antiserum (1/200) and detected with HRP-conjugated anti-goat immunoglobulin antibodies (1/2000) as described above. Or, in the case of the phages selected with rabbit antiserum, the blots were exposed to antiserum diluted 1/50, previously preabsorbed with *E. coli* lysate and fUSE2 phages and detected with HRP-conjugated anti-rabbit immunoglobulin antibodies (1/2000). Rabbit serum was absorbed by adding an equal volume of *E. coli* lysate (section 2.2.6) and approximately10^{10} CFU heat

inactivated, UV killed f1 phages to the serum and incubating at 37°C for 1 h. The immune complexes were removed by 5 minutes centrifugation in a microfuge, the supernatant transferred to a new tube and the process repeated with more *E. coli* lysate and f1 phages.

4.2.5. Screening the λZAP II *Cowdria* library.

To identify the region on the *Cowdria* genome that contains the phage displayed peptide-encoding sequences, deoxyoligonucleotides representing these sequences were synthesized. Suppliers were Genosys Biotechnologies (UK), MWG-Biotech GmbH (Roche) or GENSET (supplied by A. Bensaid, CMVT-CIRAD). The deoxyoligonucleotides were used as probes, either radiolabelled (section 4.2.5.1) or DIG-labelled (section 4.2.5.2), to screen the λZAP II *Cowdria* genomic library as described (section 2.2.9) using hybridisation stringency conditions as below.

4.2.5.1. Radiolabelled deoxyoligonucleotide probes.

Deoxyoligonucleotides were end labelled with terminal transferase (Promega) and [α^{32}P]dATP (Amersham) to 10^6 cpm/pmol. The specificity of hybridisation of the deoxyoligonucleotides was tested on DNA from *Cowdria* infected cultures (section 2.2.3), bovine DNA (section 2.2.4) and double stranded phage DNA (RF), isolated with a miniprep kit (Qiagen). Denatured DNA was slot blotted onto Magnacharge (MSI) membranes and hybridised as described in 2.2.4. Hybridisation was performed at 45°C. The membranes were washed three times at room temperature in 4xSSC and the stringency wash was once for 1 minute at approximately T_m -10°C. The filters were exposed to Hyperfilm (Amersham).

4.2.5.2. DIG-labelled deoxyoligonucleotide probes.

Deoxyoligonucleotides were labelled non-radioactively with digoxigenin (DIG) with the DIG Oligonucleotide Tailing Kit (Roche) according to manufacturer's instructions. Hybridisation was performed in DIG Easy Hyb buffer (Roche) at approximately T_m -10°C. Anti-Digoxigenin-AP, Fab fragments and the substrate CDP-Star (Roche) were used to detect the hybridisation. The chemiluminescent signal was captured by a Lumi-Imager (Roche). The specificity of hybridisation of the oligonucleotides was tested as above (section 4.2.5.1). The membranes were prehybridised for 1 h in Easy Hyb and hybridised overnight at the optimal temperature for each oligonucleotide. Membranes were washed as before with the stringency washes the same as the hybridisation temperature. Membranes were reused after the hybridised DNA was removed with 0.2 N NaOH, 0.1% SDS twice for 15 minutes at 37°C and a final rinse with 2 x SSC. DIG labelled probes were reused after denaturation at 68°C for 10 minutes.

4.2.6. Sequencing and analysing the λZAP II clones.

λZAP II clones isolated with the DNA probes were excised *in vivo* (section 2.2.10), the colonies were grown in 5 ml LB broth (containing 50 μg/ml ampicillin) overnight and the plasmids isolated using a Qiagen miniprep kit. The T7 and T3 primers were used for sequencing using an automatic sequencer (ABI Prism 377 DNA sequencer, Perkin Elmer) and the sequences were analysed as before (section 3.2.7). The predicted size and hydrophobic profile of the proteins encoded by the ORFs were determined with DNAMAN (version 2.6, Lynnon BioSoft). To determine whether sequences were similar to already known genes, BLAST® (Basic Local Alignment Search Tool), a set of similarity search programs designed to explore all of the available sequence databases was used (Altschul *et al.*, 1990; Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST/). BLASTN

was used to search for matches with DNA sequences in the DNA databases and BLASTP when amino acid sequences were used as the query sequence and the protein databases were searched. In addition, position specific iterated BLAST (PSI-BLAST) was used where no matches were found with the above options. This is an iterative search in which sequences found in one round of searching are used to build a score model for the next round. The strategy results in increased sensitivity. CD-Search compares a protein sequence against the Conserved Domain Database with the RPS-BLAST program allowing known functional and structural domains to be identified on protein query sequences. In addition, the proteins encoded by the identified ORFs were analysed with different programs: the Simple Modular Architecture Research Tool to identify protein domains (SMART; http://smart.embl-heidelberg.de), a program for the prediction of protein localisation sites in cells (PSORT; http://psort.nibb.ac.jp) and ProtScale to compute and present the profile produced by amino acid scales (http://us.expassy.org/cgi-bin/protscale.pl). Two different programs predicting continuous antigenic regions were used to analyse amino acid sequences. One uses a table based on amino acid residues of experimentally determined continuous epitopes called the Antigenic Site (Kolaskar and Tongaonkar, 1990; http://mif.dfci. harvard.edu/Tools/antigenic. html) while BEPITOPE is based on propensity scales based on hydrophilicity, accessibility, flexibility or secondary structure, either alone or in combination (Odorico and Pellequer, 2003; kindly provided by JL Pellequer).

4.2.7. *Cowdria* genome sequence database search.

An uncomplete, unedited version of the *Cowdria* genome sequence was made available (November 2001 version; Molecular Biology Division, Onderstepoort Veterinary Institute). The phage displayed peptide sequences were compared to the database as DNA sequences in a BLAST search. The resulting matches were analysed (section 3.2.7) for the presence of open reading frames (ORFs)

which encoded proteins that contained the peptide sequences. The ORFs were subjected to BLAST analysis to determine whether they were similar to any known proteins and analysed as before (section 4.2.6).

4.2.8. Bacterial expression of potential ORFs.

Two of the potential proteins identified previously by hybridisation were subcloned into expression vectors. To express the proteins encoded by the ORFs in λF and λH as fusion proteins, oligonucleotides were designed to amplify the coding regions such that an in-frame fusion with glutathione S-transferase (GST) protein of pGEX2-T (Pharmacia) was formed. A *Bam*H1 site was incorporated to enable directional cloning into the vector. In both cases the T3 primer was used to amplify the fragment and the *Eco*R1 site in the multiple cloning site was used to obtain the sites for cloning into pGEX2-T.

PCR was performed using a thermostable polymerase (TaKaRa Ex Taq), according to manufacturer's instructions in the presence of 0.2 pmol primers and 200 ng plasmid template. Twenty five cycles of 30 s at 94°C, 30 s at 40°C and 40 s at 72°C was followed by a 5 minute extension at 72°C. The product was purified with a Qiaquick PCR purification kit. Both the PCR product and the vector were digested with *Bam*H1 and *Eco*R1 (Pomega). The vector was purified by separation on an agarose gel, the band excised and cleaned up with Qiagen Gel Extraction kit. The insert was ligated into the vector with the Rapid ligation kit (Roche), according to instructions and 5 μl ligation mixture was transformed into TG1 TSS competent cells (Appendix C) by incubation on ice for 1 h, addition of 500 μl LB broth and shaking at 37°C for 30 minutes before plating on LB plates containing 50 μg/ml ampicillin and 2% glucose. Clones were sequenced with the pGEX sequencing primer (GGGCTGGCAAGCCACGTTTGGTG) to confirm in-frame cloning.

Cells were tested for protein expression by IPTG induction as described by the suppliers of the vector and separation by PAGE, followed by Coomassie blue staining or western blotting.

4.3. RESULTS.

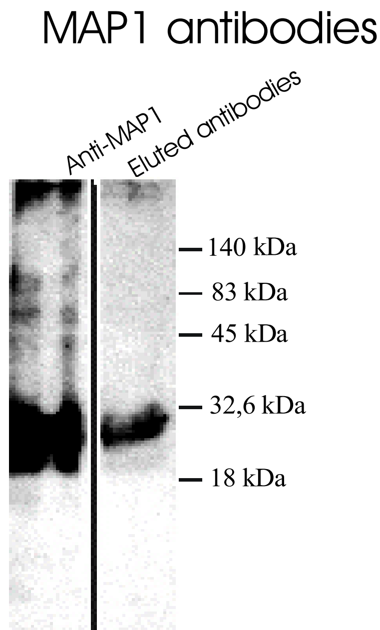
Polyclonal antiserum against *Cowdria* has been shown to contain apparently cross-reactive antibodies in addition to those directed against the original immunogen (van Kleef *et al.*, 1993). Antibodies were therefore affinity-purified from electroblotted proteins in order to obtain defined subsets of antibodies for panning (du Plessis *et al.*, 1995). Electroblotted proteins have been denatured by SDS and the disulphide bonds reduced with β-mercaptoethanol. In theory therefore, most of the eluted antibodies should be directed against linear epitopes (Blüthner *et al.*,1996). Since the *Cowdria* phage library displayed peptides derived from small fragments of DNA, these were likely to represent mainly linear epitopes.

4.3.1. Affinity selection with MAP1 antibodies.

The concentration of the antibodies eluted from the electrotransferred MAP1 was 56 µg/ml. They reacted with MAP1 as anticipated, recognising only the 31kDa protein. This was in contrast to the original preparation that recognised at least four additional proteins (Figure 4.1a). Affinity selection was performed at an antibody concentration of 10 µg/ml which was reduced to 1 µg/ml for the third round. The percentage yield (or output) was calculated from the number of phages introduced into the panning and the number of phages released after each panning. An increase in yield in subsequent rounds being taken as an indication that specific selection of phages had occurred. Significant enrichment was seen after round three when the percentage output was 178 times higher

than that from round two (Figure 4.1b). The pool of phages was thus enriched with phages capable of binding to the immobilised antibodies.

A)



B)

	Percentage Output
Round 1	1.3×10^{-5} %
Round 2	3.2×10^{-5} %
Round 3	5.7×10^{-3} %

Figure 4.1. Western blots and panning results obtained with MAP1 goat antiserum. a) Western blots showing *Cowdria* proteins reacting with MAP1 goat antiserum (diluted 1/200) and antibodies eluted from MAP1 (diluted ½). The positions of the molecular weight markers (kDa) are indicated. Super Signal chemiluminescent substrate was used and images were captured by a Lumi-Imager. b) The percentage output (number of phages released after panning / phages put into panning x 100) is given after each round of panning with the above antibodies.

The sequence of the DNA inserts of 40 clones from selection round three was determined and their amino acid sequences were deduced. This revealed that five different peptides had been recovered. They could be grouped into three distinct categories: Peptide types 1 and 2 contained the overlapping nonapeptide YLSDKYKIK (Y-9-K), peptide types 4 and 5 the sequence YISQKFMPQHNPLQD (Y-15-D), while the remaining peptide (type 3) was distinctly different to the other four (Figure 4.2a; Appendix E). None included amino acid sequences that were identical to any part of MAP1, indicating that the *map*1 gene was not present in the library as expected (section 3.3.2). Nevertheless, taking amino acid similarities into account (Bestfit,GCG), the overlapping sequence Y-9-K could be aligned with MAP1, as could part of Y-15-D (Figure 4.2b). The residues in positions 39 (Y), 40 (L/I), 41 (S) 43 (K) and 44 (Y) in particular were essentially identical. Similarly, the peptide Y-15-D showed seven identical residues (positions 39 to 41, 43, 45 and 46 and one conserved aromatic residue at position 44).

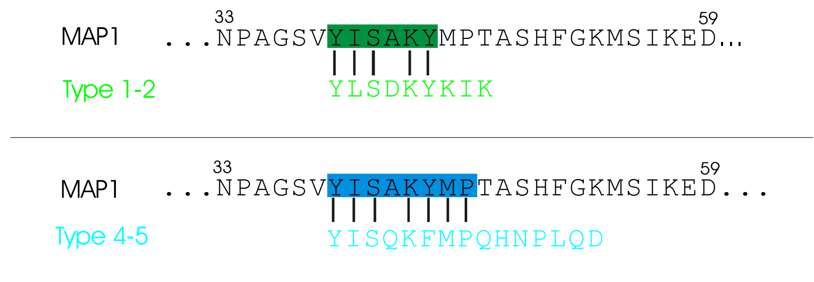
4.3.1.1. Antigenicity of the selected fusion peptides.

Selection by panning does not always imply that the binding peptides will be recognised with sufficient avidity to produce detectable signals in immunoassays such as ELISA or immunoblotting (du Plessis *et al.*, 1995). Accordingly, to confirm its ability to be recognised by MAP1 antibodies, each selected fusion peptide was tested for immunoreactivity in these assays. With the MAP1 antiserum (section 4.2.1), four of the five sequences produced convincing signals in both ELISA and immunoblotting (Figure 4.2c and d). These strong binders contained the overlapping sequences Y-9-K and Y-15-D. Peptide 3 with its distinctly different sequence was not recognised in either assay.

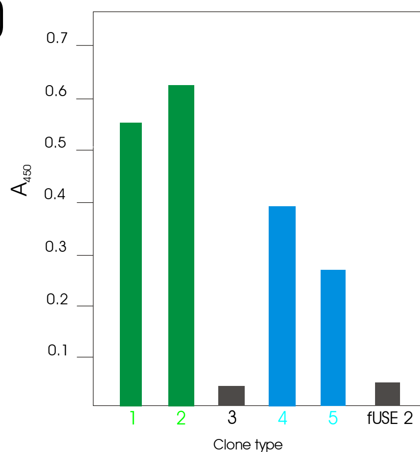
A)

Type	Sequence	No. of clones	Size
1	LKYLSDKYKIK	11/40	11 aa
2	YLSDKYKIKENLTY	15/40	14 aa
3	SKQCDEIREKIKKCNLRQGKKKSALSKFTDHF	2/40	31 aa
4	KSRLTEVNLYISQKFMPQHNPLQD	11/40	23 aa
5	YISQKFMPQHNPLQDVALI	1/40	19 aa

B)



C)



D)

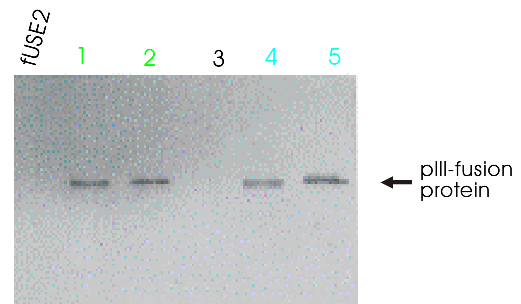


Figure 4.2. Sequence and immunoreactivity of fusion phage displayed peptides selected by affinitypurified MAP1 antibodies. a) Five peptides identified by panning the *Cowdria* phage display library with affinity-purified MAP1 antibodies. The amino acids in blue and green shows the overlapping sequences Y-9-K and Y-15-D. b) Alignment of overlapping regions Y-9K and Y-15-D with residues 33 to 59 of the authentic amino acid sequence of MAP1 (van Vliet *et al.*, 1994). Vertical lines indicate amino acid identity or near-identity. The shading illustrates a putative antigenic region on MAP1. c) Binding in ELISA of immobilised fusion phages with MAP1 antibodies. The negative control was non-fusion fUSE2 phages. d) Western blot of the pIII-fusion proteins reacting with MAP1 antibodies.

The MAP1 peptides indicate that the epitope identified by sequence comparisons (Figure 4.2b) is antigenic. Antigenic Site and BEPITOPE computer programs were used to predict continuous antigenic regions on the MAP1 amino acid sequence (Welgevonden isolate; see 4.2.6). The sequence YISAKY was identified by phage display as being an antigenic region. Both programs predicted that this sequence forms part of a continuous epitope extending upstream but including YISA, with BEPITOPE including only three of these residues (YIS). This was not the only predicted epitope (results not shown), but the *in silico* analysis shows that this region identified by phage display was slightly shifted from that which would have been predicted from sequence information alone.

4.3.1.2. Screening the λZAP II *Cowdria* genomic library.

In order to identify the regions on the *Cowdria* genome that contained the phage displayed peptide-encoding sequences, oligonucleotides based on the two peptides which were immunoreactive (i.e. type 1-2 and type 4-5; Table 4.1) were synthesised. Both hybridised specifically to *Cowdria* DNA as well as to the phage DNA that contained their specific inserts (Figure 4.3a and b). After two rounds of plaque screenings of the λ library, single plaques were excised *in vivo* and the DNA inserts of the phagemids were sequenced. Clones were isolated that contained either type 1-2 or type 4-5 DNA sequences. These regions did not, however, encode putative proteins (Figure 4.3a). The peptide sequences could nevertheless be located in the λ clones. Potential ORFs were identified in these clones, but they were either on the reverse strand or up/downstream in relation to the peptide sequences (Figure 4.4). Thus these peptides could *not* have represented cross-reactive epitopes on MAP1-related proteins. In the context of the phage display system, fragments of the *Cowdria* genome were probably expressed as fusion peptides which mimicked epitopes on MAP1 (Geysen *et al.*, 1986; see 1.1.2).

Table 4.1. Oligonucleotide probes derived from the selected phage displayed sequences. The length of each oligonucleotide is given in bases. The hybridisation temperature (T_m -10°C) is indicated. The theoretical T_m -10 is shown in brackets in cases where higher or lower temperatures were used as determined empirically.

Name	**Oligonucleotide sequence**	**Size (b)**	**Hybridization temperature (°C)**
Type1-2	GATAAGTATAAAATTAAGGAGAATCTCACG	30	55
Type 4-5	CTTTATATCAGCCAGAAATTTATGCCC	27	59
D25	GGGTTTGGAGATATGCTTATG	21	45
E3	CTCGATCTAGATACTGATGTACCCATTG	28	53
F12	GCCTTGTAAAATTTGATATGCATA	24	45
G4	TCTGATAGTAATTTTCTTAGCTGTTC	26	45
H1	GTGTTTTTTGGGTTTATGCTTGC	23	48
I8	GATATATGATGTTATGAAATAAGTG	25	(52) 48
J15	CAAATTTACTATTAGTGAATGAG	23	(35) 48
K18	CTTGAGTAAGGAACGTGTTG	20	48
L24	GTAAGCTTCCTGCAATTAAAGAAG	24	48
M37	CATTATGATTAGTAATCCTCTG	22	(35) 48
N49	CCTGAACTAGCCATTTTAATAC	22	48

4.3.1.3. *Cowdria* genome sequence database search.

Recent progress in *Cowdria* genome sequencing allowed a search analogous to the λ library screenings (previous section) to be made *in silico*. Since the sequence was incomplete, the database consisted of multiple sequences of the λ clones as well as some from additional clones. The same regions on the genome were identified as when the λ library was screened by hybridisation (section 4.3.2.3). In addition, a region matching the type 3 peptide was identified. This clone did not react with the anti-MAP1 serum. Surprisingly, this region contained a potential ORF encoding a protein that contained the phage displayed peptide sequence (Figure 4.3c, 4.4b). This 212 amino acid polypeptide was part of a larger protein. As it was selected with MAP1 antibodies, it was compared to MAP1 (Bestfit, GCG) but no similarity was found. Thus it is probably not a protein in the

MAP1-family. It did, however, have low levels of similarity with portions of known outer membrane proteins. For instance, there was 35% identity with a conserved hypothetical protein of 114 amino acids of *Streptococcus pneumoniae* and 30% identity with a translocase in the outer mitochondrial membrane of *Saccharomyces cerevisiae*, a 70 kDa protein. In addition, there were similarities with another *Saccharomyces cerevisiae* outer membrane protein of 617 amino acids, a 335 amino acid conserved hypothetical protein of *Schizosaccharomyces pombe* and to arrestin proteins which are 378 amino acids long. The polypeptide encoded by the type 3 ORF is likely to represent a new, and as yet undescribed *Cowdria* protein.

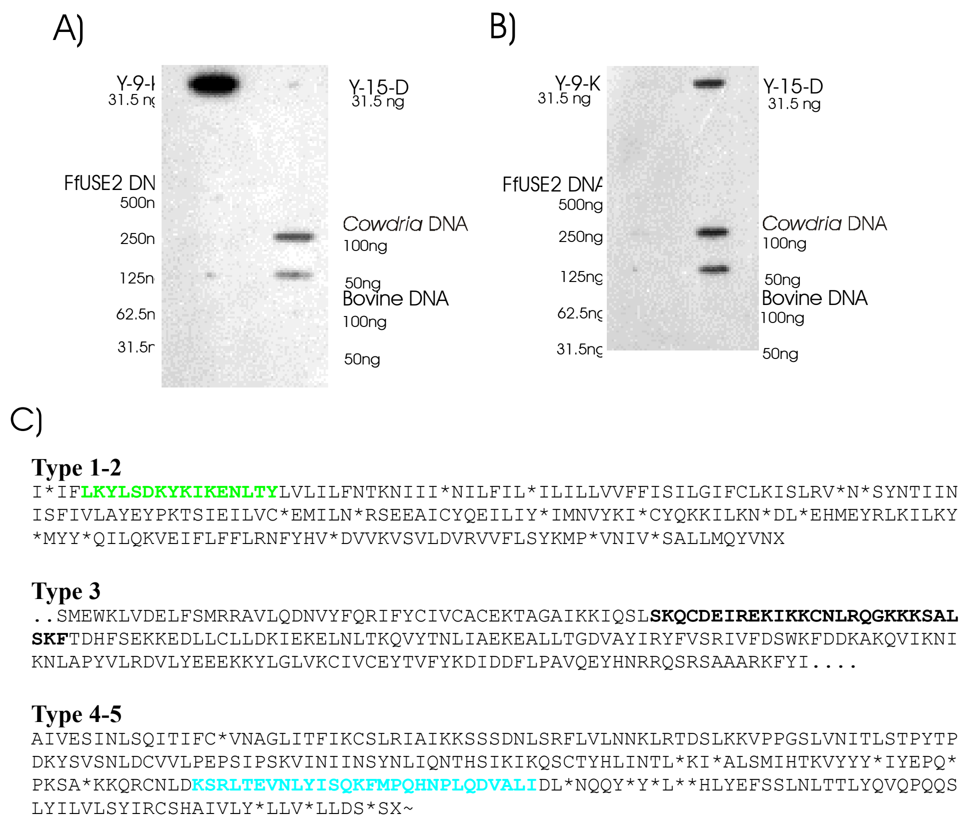


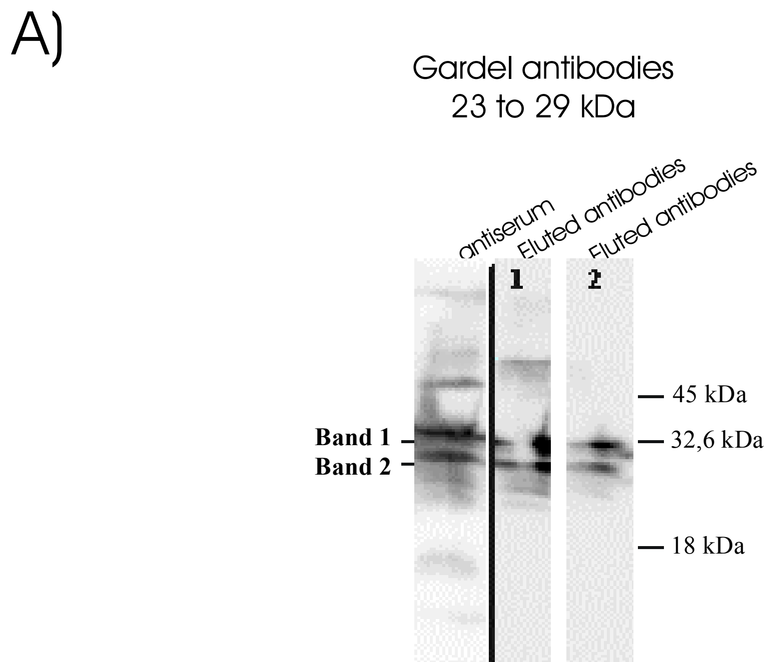
Figure 4.3. Specificity of hybridisation of the type 1-2 and type 4-5 oligonucleotide radiolabelled probes and the deduced amino acid sequence of the λ clones identified with the probes. a) and b) Autoradiograph of slot blots demonstrating the specificity of hybridisation, a) type 1-2 and b) type 4-5 probes. The amount of DNA per slot is shown. c) The deduced amino acid sequence (single letter code) of the λ clones identified by screening the *Cowdria* λ library with the above probes. The peptides (green and blue) were not in the context of ORFs. The deduced amino acid sequence of the genome match with the type 3 peptide (bold) was in the context of an ORF.

Figure 4.4. Schematic representation of the position of peptides on the λ clones or the *Cowdria* genome regions. Solid lines indicates the λ clone or genomic region, the open boxes the translated reading frames. a) The position of peptide types 1 and 2, upstream of the ORF, b) peptide type 3 as part of the ORF and c) the positions of types 4 and 5 on the reverse strand, spanning the carboxyterminus of the ORF.

4.3.2. Affinity selection with antibodies against variable immunodominant Gardel isolate proteins.

The rabbit antiserum raised against the 23 to 29 kDa range of Gardel isolate proteins cross-reacted on a western blot with a number of Welgevonden isolate proteins (Figure 4.5a). The major bands were 26 kDa (band one) and 22.5 kDa (band two), but several other *Cowdria* proteins were also recognised. Antibodies were eluted from bands one and two. The presence of specific antibodies was confirmed by western blot. Each preparation was expected to react with a single protein band. Antibodies eluted from band one reacted with the 26 kDa protein, but equally strongly with the 22.5 kDa protein. The same was observed with antibodies eluted from band two. In addition, the antibodies eluted from band one reacted with other *Cowdria* proteins. Even though they did not

recognise single protein bands, they were nevertheless used for panning the phage display library. In addition, total IgG was isolated from the antiserum. The aim was to compare peptides isolated by total serum antibodies with those selected by the affinity-purified antibodies.



B)

	Output (CFU)	Output increase
Round 1 IgG	1.98×10^3	-
Round 2 IgG	1×10^4	5x
Round 3 IgG	1×10^6	100x
Round 3 Band 1	1.7×10^5	17x
Round 3 Band 2	1.7×10^5	17x

Figure 4.5. Western blots and panning results obtained with antisera directed against Gardel proteins. a) *Cowdria* proteins reacting with rabbit antiserum made against the 23 to 29 kDa PAGE purified *Cowdria* (Gardel) proteins (diluted 1/50), as well as affinity purified antibodies (diluted 1/2). The positions of the molecular weight markers (kDa) are indicated. Super Signal chemiluminescent substrate was used and images were captured by a Lumi-Imager.b) The output after each round of panning, expressed as the number of colony forming units (CFU) and the output increase (roundn/roundn-1), with the above antibodies is given..

Three rounds of panning were performed with these antibodies. In the first and second rounds, total IgG was used at a concentration of 20 µg/ml. This IgG preparation was used since all the band one and band two antibodies (Figure 4.5a) were contained in this preparation. In the third round, band one, band two and IgG antibody preparations were used separately, each at the same concentration.

The total number of phages released after each panning was determined. Since no titres for input phages were available for calculating percentage output, an increase in output in subsequent rounds was taken as an indication that specific selection of phages had occurred. The round one output was 1.98 x 10^3 CFU. Rounds two and three showed five and 100 fold increase in outputs respectively (Figure 4.5b) for the IgG panning indicating that fusion phages had been specifically selected. The antibodies against the eluted bands were used separately in round three, both showing a 17-fold increase in output.

Peptide sequences of the inserts of 41 individual phages were determined after round three (Table 4.2; Appendix D). The peptide most frequently selected was TAKKITIRCDSITCY (G4). All three antibody preparations selected this peptide (Table 4.2). Seven phage clones had the sequence PDRVLLRIMNMQWVHQYLDRGGARYP (E3) which was selected by both the band one antibodies and the IgG preparation.

Eleven of the phage clones contained the sequence GYRAPP which was encoded by the pBluescript vector in the multiple cloning site. *Cowdria* DNA inserts had originally been amplified from this plasmid for use as starting material in the phage display library construction. The primers therefore amplified a small part of the vector DNA. Surprisingly, the sequence GYRAPP was recognised by the polyclonal rabbit serum. The sequences following GYRAPP, which were mostly short, differed between the various clones (Table 4.2), indicating that they were different and were probably selected due to the common GYRAPP. Only one of the 11 phage clones with this sequence was isolated with

antibodies that had been eluted from band two, showing that affinity purifying the antibodies had significantly removed anti-GYRAPP activity. With the exception of D25, these clones were not included in further studies. This clone had a long sequence following GYRAPP with the potential to contain additional epitopes.

Table 4.2. Peptide sequences selected with antibodies against the 23 to 29 kDa range of Gardel isolate proteins. Sequences from which oligonucleotide probes were designed are named D to N, others only have a number. The size of the peptides, the number of times each type of peptide was isolated and the antibody which selected the peptide is shown. Band one (b1) and band two (b2) eluted antibodies or total IgG. 'X' indicates an unknown amino acid due to uncertainty in the DNA sequence.

Type	Peptide sequence	Size (aa)	Number of clones	Selected by antibody
D25	GYRAPPRSLFVNISISPNPGXF	22	1	IgG
E3	PDRVLLRIMNMQWVHQYLDRGGARYP	26	7	b1, IgG
F12	GYAYQILQGIGRIKS	15	1	b1
G4	TAKKITIRCDSITCY	15	12	b1, b2, IgG
H1	EGNHQCKVVKFTDKCKHKPKKDRN	24	2	b1
I8	YKKIFTYFITSYINFADR	18	1	b1
J15	IISFTNSKFVYILKKXKSFN	20	1	b2
K18	FHIFHPLFYFQQILQHVPYSS	21	1	b2
L24	FLSNSEDLLSLDTLALSSLIAGSLLTRG	28	1	b2
M37	NILGRITNHNVSLSKA	16	1	IgG
N49	HKVLLSFSLSSIKMASSG	18	1	IgG
47	GYIQLRKGPNVVGPYGLLQP	20	1	IgG
19	GYRAPPRS	8	1	b2
33	GYRAPXRSLFVNIS	14	1	IgG
36	GYRAPPRSFR	10	3	IgG
41	GYRAPPRSMTLY	12	1	IgG
50	GYRAPP	6	3	IgG
53	GYRAPPRSMTAY	12	2	IgG

From the remaining 30 clones, 12 different peptides were identified (Table 4.2). They could be divided into three groups, each having a tendency to be rich in certain amino acids (Figure 4.6). Group A was rich in lysine (K) with the 'KKI' motif common between G4 and I8 and 'FT' occurring in H1, J15 and I8. Group B was rich in leucine (L) and serine (S) and group C rich in glutamine (Q), tyrosine (Y) and glycine (G). In addition, K18 and F12 in group C had YxxQILQ in common. No overlapping

peptides were selected as in the MAP1 panning (section 4.3.1). The peptides could be loosely aligned, but no real consensus was found. The total IgG preparation recognised peptides from all three groups. Group A and group C peptides were selected by both band one and band two antibodies. Group B peptides were recognised only by band two antibodies. These results indicated the presence of cross-reactive epitopes on the *Cowdria* proteins against which the antiserum was made. Western blotting (Figure 4.5a) also indicated that this might have been the case.

```
Group A
H1     EGNHQCKVVKFTDKCKHKPKKDR
J15          IISFTNSKFVYILKKRKSF
G4                      TAKKITIRCDSITC
I8                       YKKIFTYFITSYINFAD

Group B
L24     FLSNSEDLLSLDTLALSSLIAGSLLTRG
N49           HKVLLSFSLSSIKMASS
M37        NILGRITNHNVSLSK
D25      gyrapprSLFVNISISPNPG

Group C
E3         PDRVLLRIMNMQWVHQYLDRGGARYP
K18            FHIFHPLFYFQQILQHVPYS
F12                   GYAYQILQGIGRIK
47         GYIQLRKGPNVVGPYGLLQP
```

Figure 4.6. Peptides selected by the rabbit antibodies directed against the 23 to 29 kDa range of Gardel isolate proteins, arranged into three groups according to their amino acid content. Common amino acids are highlighted in blue, green or orange, sequences originating from the vector are in lowercase.

4.3.2.1. BLAST database analysis of peptide sequences.

To determine whether any of the abovementioned peptides formed a part of any known protein, the sequence databases were searched (BLAST). Only peptide 47 matched a known protein (Table 4.3), NADH dehydrogenase subunit 1, with only one amino acid mismatch (isoleucine in stead of methionine or valine) compared with the sequence from other organisms (Table 4.3). It is in a highly conserved region of the protein (Appendix E) confirming that the epitope is common between at least bacteria, humans, bovines and felines.

Table 4.3. Comparison of peptide 47 with the corresponding region of the known NADH dehydrogenase proteins (Genbank accession numbers in brackets). The differences are in bold.

Peptide sequence	**Organism**
GY**I**QLRKGPNVVGPYGLLQP	*Cowdria* peptide 47
..GY**V**QLRKGPNVVGPYGLLQP..	*Branchiostoma floridae* (AF035175)
..GY**M**QLRKGPNVVGPYGLLQP..	*Balanoglossus carnosus* (AF051097), human (P03886), bovine (P03887), feline (U20753), carp (S36002).

4.3.2.2. Antigenicity of the selected fusion peptides.

Representative clones (types D to N and 47) were tested on western blots for reactivity with the rabbit antiserum. Unfortunately the rabbit serum recognised some *E. coli*- and fUSE2-encoded proteins. Even after absorbing with these proteins, they were still recognised (Figure 4.7 a and b). It was therefore not possible to determine whether they recognised the expressed *Cowdria* peptide or pIII itself. Nevertheless, with clones D25, F12, M37 and 47 a protein larger than pIII was recognised, indicating that a fusion-pIII protein was reacting with the antiserum. No reaction was observed when the two affinity-purified antibody preparations were used. This could have been due to the low

concentration of the eluted antibodies or alternatively, the phage clones could have contained peptides that bound with low affinity, ie. strong enough binding to be selected by panning, but not detectable in immunoassays.

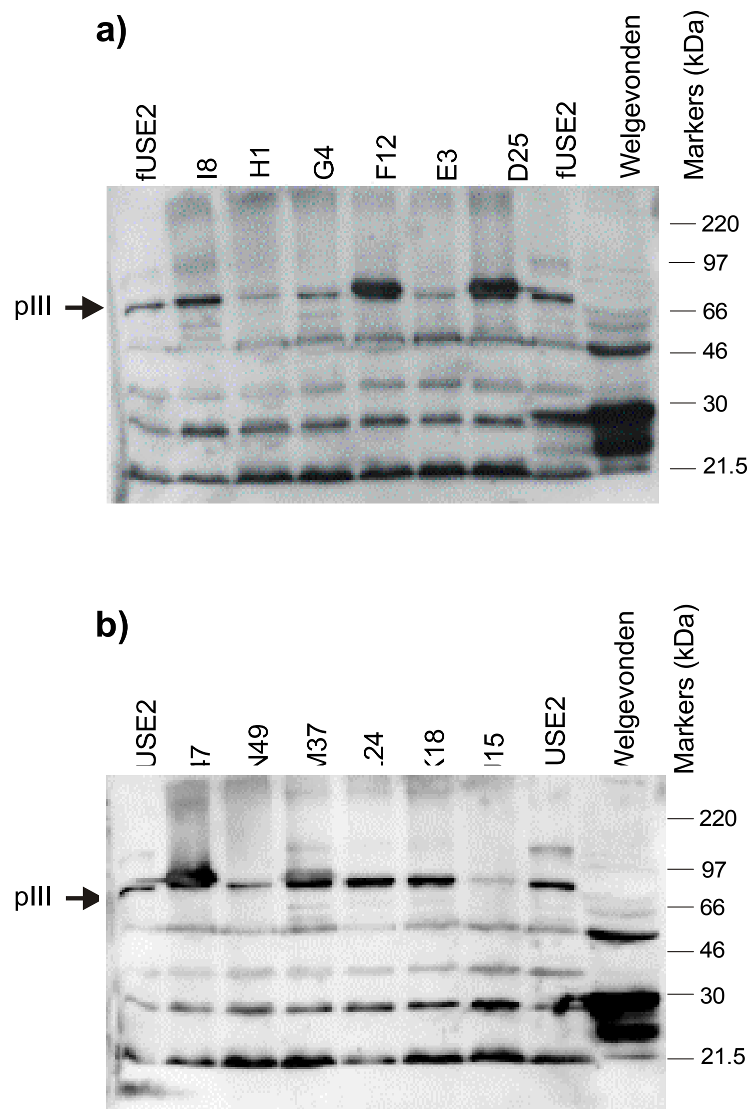


Figure 4.7. a) and b) Western blots of recombinant phage proteins (phage types D25, E3, F12, G4, H1, I8, J15, K18, L24, M37, N49 and 47) reacting with rabbit antiserum against 23 to 29 kDa Gardel isolate proteins. Welgevonden isolate and fUSE2 proteins are included as positive and negative controls respectively. The positions of the markers are indicated. The arrow shows the position of fUSE2 pIII.

4.3.2.3. Screening the λZAP II *Cowdria* genomic library.

The *Cowdria* λ library was screened with non-radioactively labelled (DIG) oligonucleotides derived from peptides D25, E3, F12, G4, H1, I8, J15, K18, L24, M37 and N49 (Table 4.1) to identify the clones that contained the phage displayed peptide encoding DNA sequences. As with MAP1 (Figure 4.3 a and b), the specificity of hybridisation of each oligonucleotide was tested on DNA from *Cowdria* infected cultures, bovine DNA and double stranded phage DNA. The results are summarised in Table 4.4. All the oligonucleotides, with the exception of I8, J15 and M37 hybridised specifically with double stranded phage DNA of its own type and with *Cowdria* DNA, but not with bovine DNA. Oligonucleotides D25, F12, G4 and H1 hybridised weakly with other types of phage DNA.

Oligonucleotide I8 did not hybridise with its own type of phage DNA at the theoretical T_m-10°C, but at 4°C lower the expected hybridisation pattern was obtained. Oligonucleotides J15 and M37 were not specific enough at their respective T_m -10°C, since both hybridised with bovine DNA. The hybridisation temperature was therefore increased by 13°C to achieve higher specificity (Table 4.4).

The λZAP II *Cowdria* genomic library was then screened these DIG-labelled probes. As an example, Figure 4.8 shows the results obtained with the oligonucleotide F12. It hybridised with *Cowdria* DNA and not with bovine DNA (Figure 4.8a) and to single plaques (1 to 7; Figure 4.8b) in the primary screening of the λ library. These plaques were picked, plated out and tested again. The positive plaques were excised *in vivo* and their phagemid DNA tested again before sequencing those which were positive (Figure 4.8c). Oligonucleotide types D25, E3, F12, G4, H1, L24 and N49 hybridised to λ clones (named λD to λN). No λ clones were identified after repeated screening with oligonucleotides I8, J15, K18 and M37. This was not surprising since oligonucleotides I8 and M37 showed weak hybridisation with *Cowdria* DNA. The reason no clones were identified with the J15

and K18 oligonucleotides is not clear. Either the DIG labelling system could have been less sensitive than radioactivity or the clones were in fact not present in the aliquot of library tested. In addition, the low AT content of *Cowdria* DNA made it difficult to design oligonucleotide probes with sufficiently high melting temperatures to achieve the desired specificity.

Table 4.4. Hybridisation profiles of oligonucleotides derived from phage displayed sequences selected by antibodies against Gardel isolate proteins (23 to 29 kDa). The temperature (+/- Tm-10°C) for each oligonucleotide is indicated. The hybridisation profile of each oligonucleotide with double stranded DNA of the phages of its own type (self) and other types, *Cowdria* and bovine DNA are summarized.

Name	Hybridization temperature (°C)	Oligonucleotide hybridization to			
		self	other phages	*Cowdria*	bovine
D25	45	✓	weak	✓	✗
E3	53	✓	✗	weak	✗
F12	45	✓	weak	✓	✗
G4	45	✓	weak	✓	✗
H1	48	✓	weak	✓	✗
I8	52	✗	✗	✗	✓
	48	weak	✗	weak	✗
J15	35	✓	✗	✓	✓
	48	✓	✗	✓	✗
K18	48	✓	✗	✓	✗
L24	48	✓	✗	✓	✗
M37	35	✓	✗	✓	✓
	48	✓	✗	weak	✗
N49	48	not done	✗	weak	✗

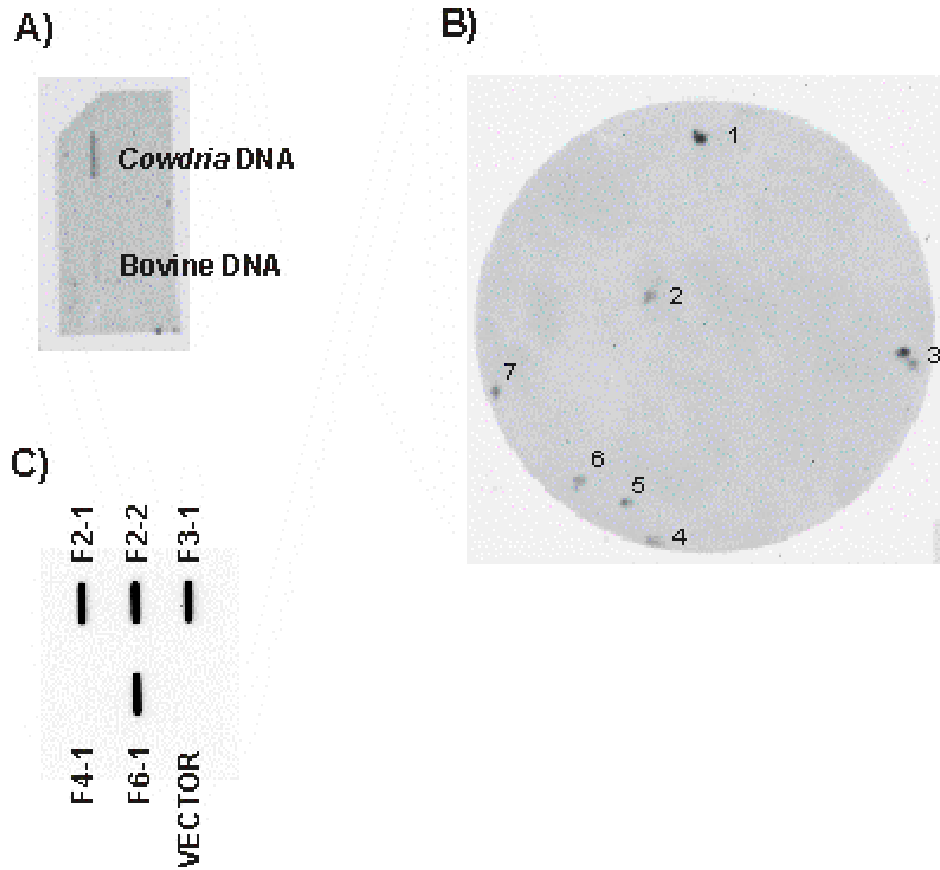


Figure 4.8. Hybridisation specificity and screening with the F12 DIG labelled oligonucleotide probe. a) Slot blot showing the hybridisation of the F12 probe with *Cowdria* DNA and not with bovine DNA. b) Primary screening of the λ library with this F12 probe, plaques 1 to 7 where picked, plated out again and rescreened. c) Slot blot shows hybridization of DNA of *in vivo* excised plasmids with the F12 probe.

The DNA inserts of individual λ clones were sequenced and analysed to confirm the presence of the oligonucleotide sequence within each clone. The DNA sequences of E, F, H and L were contained in λE, λF, λH and λL (Figure 4.9). The clones λD, λG and λN all had the same DNA insert and did not contain the D, G or N oligonucleotide sequences. The clone selected by these probes contained a long stretch of thymines which probably hybridised with the poly adenine tail created when labelling with the DIG Oligonucleotide Tailing Kit. When dCTP instead of dATP was incorporated into the labelled tail, no clones were identified. The MAP1-type probes were radioactively end-labelled with dCTP and

no clones with long stretches of thymines were identified.

Having been identified as containing sequences that represented immunoselected peptides, the DNA sequences of the inserts of clones λE, λF, λH and λL were each translated to determine whether any ORFs were contained in a particular clone. Peptide sequences F12 and H1 both formed part of predicted proteins (Figure 4.9). Neither clone contained complete ORFs. Peptides F12 and H1 were, however identified as being part of putative *Cowdria* proteins (section 4.3.3.5). This was not the case with peptides E3 and L24 (Figure 4.9). The DNA sequences of the phage clones were found in the inserts of the λ clones, but the peptide sequences did not form part of potential proteins.

The complete E3 peptide-encoding sequence was not present in insert of λE (Figure 4.9). This could be explained by the fact that it contained a *Sau*3A I site (GATC) and when constructing the λ library, chimeric clones of two unrelated *Sau*3A I fragments could have been cloned in tandem. Thus either λE or the clone encoding peptide E3 was derived from was chimeric. It was more likely to be the latter, since there was a potential ORF in λE, spanning across the *Sau*3A I site. The λE potential ORF was one frame out compared to the one containing the E3 peptide-encoding sequence and thus irrelevant to this study.

Multiple copies of the L24 peptide-encoding sequence were found in λL (Figure 4.9). These were not exact matches and not part of an ORF. There was also a potential ORF in the region, but because it was on the reverse strand, it is not relevant to this study.

λF clone translated (F12orf):

ELRLSKVMKEKDSLGEVNVPACHYWGAQTQRSIDNFKIGSEKMPKPLIRAMGIVKLAAARVNMKNG
DINEVIGNAICNAAAEVIDGKLDNEFPLVVWQTGSGTQTNMDVNEGIANRAIEILGGEKGSKVPVH
PNDHVNYSQSSNDTFPTAMHIATVSETENYLLPSLKNLYDALHSKSIAFQNIVKVGRTHLQDATPL
TLGQEFS**GYAYQILQGIGRIKS**ALSNLLELAQGGTAVGTGINSRKQFDVHIANEIKKITGFNFVSS
VNKFEALATHDALVEFSGALNVLAVSLMKIANDIRLLSSGPRCGIGEIILPANEPG...

λH clone translated (H1orf):

...SIMLLLPVIEQGQNREKYEKSIAEAAALLEAQVDLIIKQQQRRNSASIVEITDEEAEQIIREQ
KATQQTRQQSRSDLNASSGEVSIEFIQE**EGNHQCKVVKFTDKCKHKPKKHR**KHKDLPPSNSNENVA
LLSLTQYSSSSCLKKDDISTTNTTQATVSTAQDRTPSSNVFAPTVYQNPCITALTLRRHSI*

λE clone translated:

NM*WQNLWAKVPDVQEVKVALCICST*KNNSLEDMVS*VHKFL*VQELLLLTNTKKITTWFLHALE
MELLTKDKYMKLLIWLHSGNYQ*YM**LLRIMNMQWVHQYLD**HLILQIYIKKEKVLVYQDIK*MEWIY
LQ**KQQLMQQLIVENKTDLFY*R*RHTATE

λE-DNA	AGAATAATGAATATGCAATGGGTACATCAGTATCTA**GATC**ATCTTATATTACAG																
λE-translated	**R**	**I**	**M**	**N**	**M**	**Q**	**W**	**V**	**H**	**Q**	**Y**	**L**	**D**	H	L	I	L	Q
E3-peptide	**R**	**I**	**M**	**N**	**M**	**Q**	**W**	**V**	**H**	**Q**	**Y**	**L**	**D**	G	G	A	R	Y

λL clone translated:

*TSDLDVSSGKDLLLSSFI*TL**FLSNSEDLLS**V**D**N**LALSSLIAGSLLT**SDLDVSSGKDLLLSSFT*
TL**FLSNSEDLLS**V**DTLALSSLIAGSLL**ISNLDVSSGKDLLLSSFI*TLL**LSNSEDLLS**V**DTLALS**F
ST**A**SYFSST**G**TWSSSNFSLTSNFLLTLLLLTCFTFFLES

Figure 4.9. Deduced amino acid sequences of the λ clones identified by screening the *Cowdria* λ library with oligonucleotide probes derived from F12, H1, E3 and L24. In λF and λH the peptides (bold) were present in the context of proteins encoded by ORFs, whereas in λE and λL this was not the case. The DNA sequence of λE, with the translation shows the position of the *Sau3A* I site (GATC) at which point E3 and λE differs. '*' indicates the positions of the stop codons.

4.3.2.4. Characterisation of the potential ORFs.

A BLAST search showed that the partial ORF coding for a 321 amino acid protein contained in λF had a high similarity with fumarate hydratase (fumarase). Amongst others, it had 60% identity with the fumarases of both *Rickettsia prowazekii* and *Rickettsia conorii* and was thus highly likely to be a *Cowdria* fumarase. Interestingly, the epitope identified on this protein (F12) was in a variable region (Appendix F) and could possibly be a *Cowdria*-specific epitope. The abovementioned sequences are approximately 460 amino acids long and fumarases are known to be 50 kDa or larger (Woods *et al.*, 1986; Acuña *et al.,* 1991). This protein is thus larger than expected, since the antibodies used for panning were raised against and eluted from proteins of 23 to 29 kDa. Peptide F12 was selected with antibodies eluted from band one. There was a reactive protein larger than 45 kDa (Figure 4.5a), which could conceivably represent the *Cowdria* fumarase.

From the sequence of λF it was evident that the partial ORF (called F12orf) contained in this clone would not form an in-frame fusion protein with the β-galactosidase of the λZAPII vector (Appendix G) and would therefore not be expressed as an inducible protein. The F12orf was thus subcloned into pGEX2-T. Sequencing confirmed that the insert was cloned such that it would form an in-frame ORF with the GST-encoding sequence of the vector. IPTG was added to induce fusion protein expression of bacterial cultures containing the constructs and the resulting cell extracts were separated by PAGE. The expected size of the resulting fusion protein was 55 kDa (24 kDa GST plus 31 kDa, the predicted size of F12orf-encoded protein). A slightly larger protein of approximately 62 kDa was present, but the molecular weight estimation was not very accurate with the prestained markers used (Figure 4.10). This protein did however not react with the eluted antibodies and the antiserum recognised bacterial proteins, resulting in a high level of background (not shown).

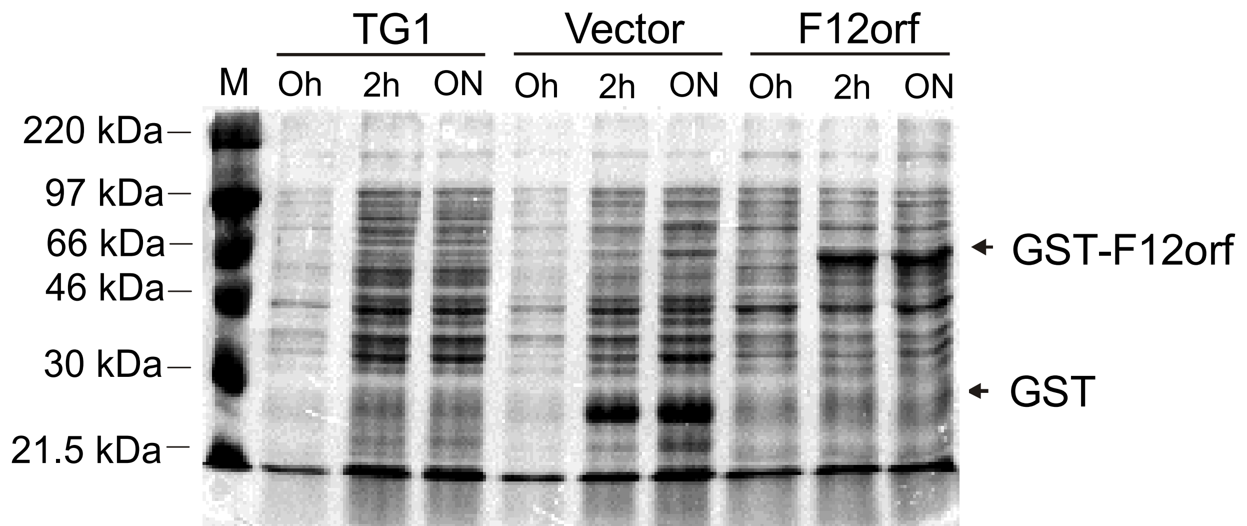


Figure 4.10. Coomassie blue stained polyacrylamide gel showing protein expressed from F12orf cloned into the vector pGEX2T. Controls include *E. coli* cells (TG1) and TG1 containing the vector. Samples were taken before (0h), two hours (2h) after IPTG induction and after overnight incubation (ON). The arrows show the positions of the IPTG-induced 24kDa GST protein and the GST-F12orf fusion protein at approximately 62 kDa.

The ORF in λH (H1orf) was similarly subjected to a BLAST search. This 190 amino acid polypeptide was predicted to be a bacterial inner membrane protein (PSORT). It had a very low level of similarity with a heavy-metal transporting P-type ATPase of *Proteus mirabilis* (829 amino acids) and an ATP-dependent proteinase of *Bacillus halodurans* (711 amino acids). Sequence comparisons with limited *Cowdria* sequences available showed that it was in fact part of a much larger protein called CPG1 (from *Cowdria* polymorphic gene; B. A. Allsopp, personal communication; Louw *et al.*, 2002). The encoding gene, *cpg*1, was identified by means of a 1.2 kb DNA fragment from the Gardel isolate generated by RAPD. It was found to be a polymorphic fragment which varied between *Cowdria* isolates (Perez *et al.*, 1997). The rabbit antiserum used to select H1 was raised against variable immunodominant proteins. Thus both the genetic (Perez *et al.*, 1997) and antigenic (phage display)

methods had resulted in the same gene being identified. Despite several attempts, this gene could not be expressed in *E. coli*, neither H12orf from the λ clone nor the complete *cgp*1. *In vitro* transcription yielded RNA, but translation did not produce detectable protein. It could also not be expressed in the baculovirus system (done by C. Potgieter, Biochemistry Division, Onderstepoort Veterinary Institute.)

4.3.2.5. *Cowdria* genome sequence database search.

As described in 4.3.1.4, the *Cowdria* genome sequence database was searched for sequences matching the peptides that had been identified with antibodies directed against the Gardel isolate proteins. This excluded those containing GYRAPP (Table 4.2). Again the results obtained by screening the λ library were confirmed (E3, F12, H1 and L24). In addition, new matches were found for D25, G4 and N49 for which only 'poly A' λ clones were found, and I8, J15, K18 and M37 for which no corresponding λ clones were found. Peptides D25,G4, I8, J15 K18 were all likely to represent peptides that mimicked *Cowdria* epitopes since they were not in the context of identifiable ORFs (Figure 4.11). Peptides M37 and N49, on the other hand, were in the context of potential proteins encoded by ORFs (Figure 4.11).

The region of the genome that matched the M37 DNA sequence contained a potential ORF (called M37orf) which coded for a protein of 230 amino acids with a predicted size of 26.5 kDa. Peptide M37 was present in this protein sequence. The M37orf-encoded protein was in the expected size range. It was predicted to be a bacterial outer membrane protein (PSORT) and contained a transmembrane region (positions 10 to 32; SMART).

Genome match M translated (M37orf):

YKINMLNIFMKKSIYFVISITFMLFSFDASSNSVYIKGNYNLGINYSDSFKSEYVNYERLSPDFA
VAIGYKLQNGLFCDLEMRYANIRPMINKLSNFEHIDLL**NILQRITNHNVSLSKA**ENILTINAITT
LINVGYSYVFNDKFRGYFTYGVGIGGLLNYQGFKSSLSTYYGISMQSEIGVCYIYNSKIDVCIGY
NYLKNYWKYETDRIPDEDGNTVMYNFQDFQLNSHTIFLGLNVLF*

Genome match N translated (N49orf):

TQVEIFMSCSNDKDIDSEYIGSSKYFQDALDWYCNKYLFCITERAWLSIVTLFLLYLILVLMFDI
YSYFPLKTDFSFVKYTDRYTDEFSVIRKLSTNNEDSEETLLSRYLVAQYVKRYESYSYDDLESQF
NFIENNSSRKIYLSFKEMKDPSNPSNSKYNSKALSISTVIDKVDLISQNFTSLNKAIVTFKTKIA
VNGVKSHTES**HKVLLSFSLSSIKMASSG**IMPLEFVVHDYKKLD*

Genome match D translated:

LLIDMINQYFHLYIIRTITTIATLNIITR**SLFVNISISPN**RECIILLFAEGKKDNNIIPKDMVTLV
TILNTVSIE*TLLIP*DSKNAKITEVHNIPADILTFSNIDKATPKRDA*EMLSPR*ERFFHIINEP
SEPVTRLTITMAIIPLIK*YSLIISNITTYNDIKAAQIFLPSYNKTLYVRQTAPVDICSRLKPNSK
K*FFISLGTKI*SLPVPTKTIFIVSSNNSMILFKLKILRFK*

Genome match G translated:

NFF*KYVKLDVILT*LFANIIIN*LGYK*VYISKYRFMLSNW*SINFLCFNILGTEPF*CFIFTT
NLIFGM**HIVNERFSRYGILLLDNVTI*YSWST*C*FSACYET**TAKKITIRCDSITC***RTTHKFS
T*TS*G*RNFGTLKSKCKSVX

Genome match I translated:

L*YSYYIIYNLALNIIYLI*GSMNIISN***KITDIFSNISLNYVL**KE*KLPLQI*FLTYINKI
KY*VLFIVISFL*HF*YTYILLVI*ANTLIILYYL*SI**HY**KKIFTYFITSYIN**L*LSQLTNLGN
*YCDTL*IK*C**INYKTIYKKIHILI*AMTYIISTTHN*LSSNIILGRX

Genome match J translated:

SIDVLLTFNYV**D*Y*SVYFLTILTGLYKKYSCILPI*LTVTIV*MLTIIKYG*CLIYW**H*N*
NIYSDYH*INSYRVANYWSYVFDIYFLVVYFLSGCVSFLVSILK*LNNFE*FGSLIIRC*YKKDVG
*LLKKML*FLYFNMLKMLITSQ*SVV**IISFTNSKFVYILKKRKS**SMSX

Genome match K translated:

LQHPYTVYY*TSLKH**FHIFHPLFYFQQILQHVPYS**RKHN*Q*KGNLSELIYS*KPSSDT*LITYTI
*YKN*NLLKHKDLSETSNFYAHPCIKRYTH*YY*QPQD*Y*NSQVHIN*RSMLMNVLVIGSGGREH
AMLHSIRKSTLLNKLFIAPGREGMENLAEIIDVDVNNTIEVIQICKKEKIDLVIIGPEIALMNGLS
DALTEEGILVFGPSKAAARLESSKGFTKELCIRYGIPTAKYGYFIDVGPACKFIDKQKLPLVVKAD
GLAQGKGTIICYTHEEAHNAVNAMLLGNKFGEAGHAVIIEEFLEGKEISFFTLVDGSNPIILGVAQ
DYKTIGDHNKGLNTGGMGSYS*PNIITQEM

Figure 4.11. Deduced amino acid sequences of the genome matches with M37, N49, D25, G4, I8, J15 and K18 (in bold). In M37orf and N49orf, the peptides were present in the context of ORFs, whereas in the rest of the sequences, this was not the case. '*' indicates the positions of the stop codons.

Indeed, PSI-BLAST showed amongst others, a match with the major outer membrane protein, P30-12 of *Ehrlichia canis* (score 39, E value 0.03), but the level of identity was only 23%. There were also similarities with a *Wolbachia* sp. surface protein, a *Neisseria gonorrhoeae* outer membrane protein and the major outer membrane proteins OMP-1V and P28-6 of *Ehrlichia chaffeensis*. PSI-BLAST is an iterative search program in which sequences found in one round of searching are used to build a score model for the next round, resulting in a 'more sensitive' search (Altschul *et al.*, 1997). Thus it is seems that the M37orf-encoded protein had features in common with membrane proteins. It is therefore probably exposed on the surface of the organism and hence can be 'seen' by the immune system.

Peptide N49 was part of the protein encoded by the ORF (called N49orf) contained in the genome region that matched the N49 phage displayed DNA sequence. N49orf codes for a potential protein of 232 amino acids, with a predicted size of 26.9 kDa. This is also within the size range of the proteins from which the selecting antibodies had been eluted. It was predicted to be a bacterial inner membrane protein (PSORT) and contained a transmembrane region (positions 40 to 62; SMART). It had a very high scoring match (score53, E value 3e-06) with the VirB proteins of *Rickettsia prowazekii* and *Rickettsia conorii*, even though the identities were only 23%. The VirB family of proteins is involved in membrane transport to and from the cells. VirB8 is an essential virulence protein of *Agrobacterium tumefaciens* (Dale *et al.*, 1993). It is an integral membrane protein (237 amino acids) and involved in DNA transport from bacteria to plant cells. (Kumar and Das, 2001). A 17 kDa protein of *Bartonella henslae,* a Proteobacterium, is related to the VirB family and found to be immunodominant (Padmalayam *et al.*, 2000).

4.3.3. Affinity selection with antibodies against potentially protective proteins of the Welgevonden isolate.

Protein fractions obtained from the Welgevonden isolate of *Cowdria* have been examined for their ability to stimulate lymphocyte proliferation *in vitro.* These studies identified a set of proteins in the size range of 11 to 23 kDa. They may thus play a role in protection against heartwater (section 1.3.10; van Kleef *et al.,* 2000). Rabbit antibodies were also raised against these proteins (Dr. M van Kleef, personal communication).

Rabbit antisera raised against pools of proteins of the Welgevonden isolate in the range of 11 to 23 kDa were supplied by Dr. M. van Kleef, OVI. Only two of the rabbits produced antibodies that could be used for panning. These were directed against proteins in the 23 to 24 kDa and 18 to 22 kDa ranges (Figure 4.12a and b). The antibodies (91 µg/ml) eluted from the 23 to 24 kDa proteins reacted only with proteins in the region from where they were eluted (Figure 4.12a). In the case of the second rabbit antiserum, antibodies were eluted from four protein bands in the18 to 22 kDa range (one to four, the concentrations were 140 µg/ml, 160 µg/ml, 160 µg/ml and 110 µg/ml respectively). Despite immunoaffinity purification, the eluted antibodies still reacted serologically with a range of *Cowdria* proteins (Figure 4.12b).

Four rounds of panning were performed with the affinity-purified antibodies from the antiserum raised against proteins in the 23 to 24 kDa molecular weight range. Since the antibody concentration was very low, a 1/20 dilution was used for panning. This corresponded to approximately 4.5 µg/ml of IgG. In round two, the concentration of antibody was increased to 18 µg/ml and panning was performed in microtitre wells. In the final round the concentration was decreased tenfold. The yields of phages

were relatively high in round one and there was no significant increase in output in subsequent rounds (Table 4.5). The sequences of selected phages showed there was a large number of background phages with no inserts. Of the sequenced fusion phages (Appendix D), three clones had the peptide sequence LSVCALKLPPILSKICAMS (type O4). There was one each of AVPGIDKVCESKTLSPHNAD (type P8) and ITKFIRSVLH (type Q23).

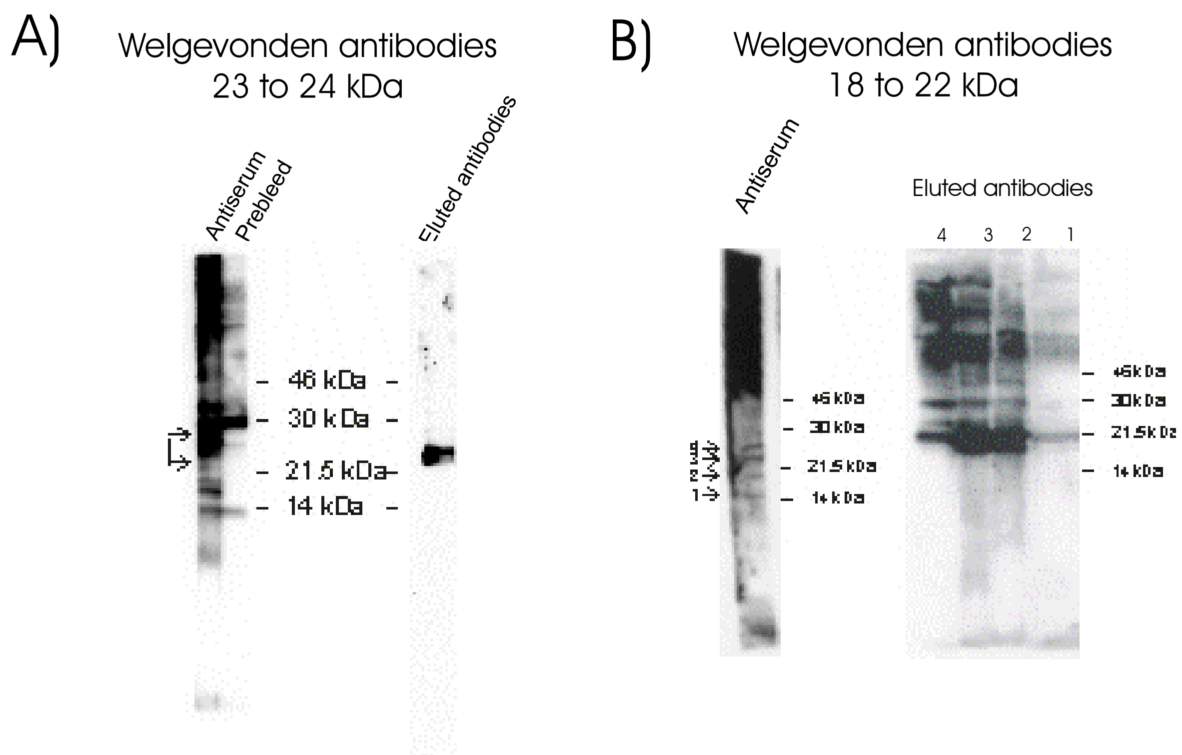


Figure 4.12. Western blots obtained with antisera directed against proteins of the Welgevonden isolate. a) prebleed, the antiserum against the 23 to 24kDa proteins (both 1/20 dilution) and the antibodies eluted from the region indicated by the arrow (1/5 dilution). b) the antiserum against the 18 to 22 kDa proteins (1/20 dilution), lanes 1 to 4, the antibodies eluted from the respective proteins (all diluted ½), indicated by the arrows. The positions of the molecular weight markers (kDa) are indicated. Super Signal chemiluminescent substrate was used and images were captured by a Lumi-Imager.

Four rounds of panning were also performed with rabbit antibodies eluted from the 18 to 22 kDa molecular weight range of proteins. In round one, the eluted antibodies from bands one to four were pooled, each at a final concentration of approximately 10 µg/ml. For the remaining three rounds, the antibodies were coated separately to microtitre plate wells at the same concentration. In round four the concentration of the antibodies was decreased tenfold. In general, there was an increase in output after every round, varying somewhat with the different antibodies. DNA sequencing revealed that once again there was a high level of background clones carrying no inserts. Nevertheless, seven clones had the sequence LSVCALKLPPILSKICAMS which was the same as O4 (above; Appendix D). These peptides were selected by antibodies released from both bands three and four. A single clone obtained with the antibodies eluted from band two had the sequence GAPAFVPFS (type R36). No specific selection was achieved with antibodies eluted from band one. This was not surprising since the antibodies did not react as strongly on the western blot as the other three antibody preparations. The antibodies released from bands two, three and four all reacted with similar proteins which could explain why they selected the same sequence. Bands three and four were larger than the 21.5 kDa marker, and were possibly the 23 to 24 kDa proteins used to elute antibodies. This might explain why the type O4 peptide was selected by antibody preparations derived from both the rabbit antisera.

As was found previously (Figure 4.7), the rabbit antisera cross-reacted with *E. coli* and phage proteins. In an ELISA there was no difference in signal between the selected phages and the fUSE2 phages. The eluted antibodies also did not react in a western blot (results not shown).

Table 4.5. The results of pannings with antibodies directed against the 23 to 24 kDa and 18 to 22 kDa Welgevonden isolate proteins. Outputs are expressed as the number of colony forming units (CFU) and the output increase ($round_n$/ $round_{n-1}$) after each round of panning are shown.

	Output (CFU)	Output increase
Antibodies against 23 to 24 kDa Welgevonden proteins		
Round 1	1.9×10^5	
Round 2	3.9×10^5	2x
Round 3	1.56×10^5	none
Round 4	1.2×10^4	none
Antibodies against 18 to 22 kDa Welgevonden proteins		
Round 1		
bands 1,2,3,4	6.27×10^4	-
Round 2		
band 1	8.2×10^3	none
band 2	3.2×10^4	none
band 3	4.5×10^5	7x
band 4	2.7×10^5	4.3x
Round 3		
band 1	1×10^5	12x
band 2	6.8×10^5	21x
band 3	3.9×10^5	none
band 4	6.8×10^5	2.5
Round 4		
band 1	3.5×10^5	3.5x
band 2	2.2×10^5	none
band 3	5.3×10^6	13x
band 4	1.5×10^5	none

4.3.3.1. *Cowdria* genome sequence database search.

The *Cowdria* genome sequence database was searched for DNA sequences matching those encoding peptides O4, P8, Q23 and R36. Once again, some of these were antigenic mimics (O4, P8 and R36),

whereas Q23 was part of a *Cowdria* ORF (Q23orf, Figure 4.13). This partial ORF coded for approximately 322 amino acids. It represented a potential bacterial inner membrane protein (PSORT) containing four transmembrane regions (SMART). The RPS-BLAST program, which compares a protein sequence against the Conserved Domain Database, found that it had structural domains in common with ABC transporter transmembrane regions. In addition, it had a high match with the alkaline protease secretion ATP-binding proteins (AprD) of *Rickettsia conorii* (score 206, E value 2e-52) and *Rickettsia prowazekii* (score 205, E value 6e-52). The Q23orf encoded protein had 37 % identity with these two proteins (Appendix H). These proteins are 584 and 583 amino acids long respectively. Thus in theory the Q23orf-encoded polypeptide is part of a protein larger than the 23 to 24 kDa proteins from which the antibodies that selected Q23 were eluted. An explanation could be that there were breakdown products derived from a larger protein in this area of the gel, or that there are cross-reactive epitopes on the different sized proteins.

4.3.4. Summary of ORFs identified.

The results obtained by screening the *Cowdria* λ library and searching the *Cowdria* sequence database with the DNA sequences identified by phage display are summarised in Table 4.6. Six potential protein encoding genes were identified whilst the remainder of the peptides probably mimicked *Cowdria* epitopes.

Analysing the peptides retrospectively with regards to GC content (Table 4.7; Appendix D) shows that there was a distinct difference between sequences derived from coding and non-coding regions. The sequences derived from coding regions (Table 4.7; type 3, F12, H1, M37, N49, Q23 and peptide 47) had an average GC content of 31.7% whereas those from non-coding regions (type 1, type 2, D25, I8

and J15) had an average of 24%. Those which were derived from a region encoding a protein, but in the reverse complement or another reading frame had an average GC content of 36% which as expected, was similar to those of the coding regions. There were exceptions; the DNA encoding M37 had a very low GC content (22%) but was part of an ORF and that coding for D25 was part of an intergenic region with a high GC content (40%).

Genome match O translated:

NMLPAW*CSCIYVCKFSLERLLTDSGVPKIDLPNT*LG*AASIYLSKMILVGESLV*YISCKTTPF
SLSISSCGNCEFNIISASTSMAILISRFNILTWYTVCS**LSVCALKLPPILSKICAMSS**AVRVLVPL
KAMCSRK*

Genome match P translated :

AVP**GIDKVCESKTLSPH**N

ISSIDSFDTG*GNLA*ISESGTSTKSLSCICL* **GIDKVCESKTLSPH**KIMSISIHLGDQYSLLNLP
IFFSMYLTLSNIRFGERFV*PDTTIFRKGVSQSLLAFSGNSAMLL*I*EYVLLYKGNIFIADLIFN
NILPQLLPNASTTIS*SGE*SMSII*VKLLQSI*IIE*IFGRFNF*VEKSITX

Genome match Q translated (Q23orf) :

*INCLEESVDGVNFNFRAMKELKKSVLYTSLEKCKSVFWFIFWFSSAINLLMLFLPLYTSQVLDRV
LTSGSVSTLVMLSAITIIAFACSAILEICRSLVMAKVGDWIDKVVTPDLIMKSISLTSIKSSTSSG
EVIRDLGVVKSFITGFGIFSLFDTPWAVLYLV0TIFMIHSVTDDWAQKNDKNRAMQIKAQNRSNLI
SG**ITKFIRSVL**QIAVIGIGAFLAVLGHKTAGGIIASSILMGRALAPFETSINTWKMLISARISYKR
LQMLLVASPKREQTMSLPIPQGKVVFDHKIIIGQVIDCSKLDSKDSPLVYYRGQYMV..

Genome match R translated:

IINI*LVYYCIEKEEIFISCSSWL*RYP*RLYYLIINANSKFNICISYIYY*HYFLYQLLCPIAW

G**APAF**VPF

LTVGKPSADNSSANSRLILFAILNP*YTIAVYSCISD**APAF**IFSYAWYPLLTPPQPISKSFCCVIL
YTSDKTLVDKSNIGFPLRPPCIFV*EFSSNL*YSDEVLLIIRPAYLYFRKHE

Figure 4.13. Deduced amino acid sequences of the *Cowdria* genome sequences matching the peptides O4, P8, Q23 and R36 (bold). Peptide Q23 was present in the context of Q23orf, whereas in the rest of the sequences, this was not the case. '*' indicates the positions of the stop codons in the DNA sequences.

Table 4.6. Summary of the genes identified. The λ clones identified by the cognate oligonucleotide probes derived from phage displayed peptides, selected by the respective antibodies as indicated and the matches with the *Cowdria* genome sequence are shown. 'Mimic' indicates that the peptide sequence probably mimics a *Cowdria* epitope and does not occur naturally in the *Cowdria* proteome. 'Poly A' indicates clones containing large stretches of adenosine residues. Peptides in bold could be aligned to proteins encoded by authentic *Cowdria* ORFs.

Peptide	λ clone	*Cowdria* genome sequence match
Antibodies against MAP1		
1 and 2	mimic	mimic
3	not done	membrane protein
4 and 5	mimic	mimic
Antibodies against 23 to 29 kDa Gardel isolate proteins		
D25	poly A	mimic
E3	mimic	mimic
F12	fumarase & poly A	fumarase
G4	poly A	mimic
H1	cpg1	cpg1
I8	-	mimic
J15	-	mimic
K18	-	mimic
L24	mimic, repeat	mimic
M37	-	permease, surface protein, OMP-1V (*Ehrlichia*)
N49	poly A	VirB (*Rickettsia*)
Antibodies against 23 to 24 kDa and 18 to 22 kDa Welgevonden isolate proteins		
O4	not done	mimic
P8	not done	mimic
Q23	not done	ABC transporter membrane region, alkaline protease secretion protein (*Rickettsia*)
R36	not done	mimic

Apart from the GC content, each organism exhibits its own pattern of codon usage. Accordingly, analysing the codons used for the different amino acids could also indicate whether it falls into the '*Cowdria*' pattern (Mathe *et al.*, 1999). This could conceivably help to distinguish whether the DNA fragment was cloned such that the authentic epitope (eg. F1, H1, M37, N49), or alternatively a peptide originating from a different reading frame or orientation was expressed (eg. E3, G4, K18, L24, O4,

P8). The codons used for each peptide were compared to the RSCU value for each codon determined for *Cowdria* (de Villiers, 2001; Table 4.7). The DNA inserts coding for peptides L24, M37 and N49 had the highest percentage of 'expected' codons (92 to 94%; Table 4.7), but only peptides M37 and N49 were authentic epitopes (Table 4.6). The other peptides which formed part of potential proteins, used 70 to 78% of the 'expected' codons. On the other hand, peptide I8 was encoded from an intergenic region and also used codons similar to *Cowdria* genes (78%).

The wealth of sequence data becoming available makes it easy to identify ORFs, but can antigenic regions be identified from these sequences? The positions of the experimentally identified phage displayed peptides (potential linear epitopes) were analysed within each protein. Two epitope prediction programs were used. One based on experimental data (Antigenic Site) and the other on the physio-chemical properties of the amino acid residues (BEPITOPE; see 4.2.6). It must be taken into account that the phage displayed peptides contains possible epitopes and the whole peptide is not considered to be the epitope. The Antigenic Site predictions were consistent with the experiment results where all six peptides formed part of the predicted epitopes, whereas only four of the six peptides were predicted by BEPITOPE to be antigenic (Table 4.7). Thus these regions might have been identified as antigenic if only the sequence was available, depending on the criteria used for prediction. They are also one of many possible antigenic sites on the protein. Both programs predicted a minimum of six antigenic sites per protein. This represented a six fold 'over-prediction' in the context of this study.

Table 4.7. The GC content of DNA sequences encoding peptides selected by phage display. The 'genome match' indicates whether the DNA fragment was derived from a protein encoding region

(ORF), an intergenic region (-) or overlapped with the end of an ORF (end). Those in bold indicated that the encoded peptides formed part of proteins encoded by the ORFs. The others are peptides derived from a region encoding a protein, but in the reverse complement or another reading frame. A measure of codon usage is given as a percentage of codons expected to be used by *Cowdria* (section 4.2.3). Epitope prediction indicates whether the peptide would have been predicted as an antigenic region by the Antigenic Site or BEPITOPE progams (4.2.6). '✔' and '✕' indicates whether or not the phage displayed peptide forms part of the predicted region.

Peptide	% GC	Genome match	% Expected codons	Epitope prediction	
				Antigenic site	BEPITOPE
Type 1	20	-	75	-	-
Type 2	24	-	60	-	-
Type 3	**31**	**ORF**	**66**	✔	✔
Type 4	40	end	61	-	-
Type 5	45	end	53	-	-
D25	42	-	72	-	-
E3	45	ORF	50	-	-
F12	**35**	**ORF**	**78**	✔	✔
G4	35	ORF	60	-	-
H1	**35**	**ORF**	**78**	✔	✔
I8	16	-	78	-	-
J15	18	-	77	-	-
K18	33	ORF	74	-	-
L24	27	ORF	92	-	-
M37	**22**	**ORF**	**93**	✔	✕
N49	**26**	**ORF**	**94**	✔	✕
O4	34	ORF	72	-	-
P8	43	ORF	72	-	-
Q23	**26**	**ORF**	**70**	✔	✔

4.4. DISCUSSION.

The phage display library constructed using fragments of the *Cowdria* genome was panned with a variety of different antibodies, the aim being to identify epitopes on antigenic *Cowdria* proteins. Antibodies directed against MAP1 selected five different peptides. Although multiple copies of four

of these were isolated, the peptides could be grouped into three categories. Peptides 1 and 2 overlapped with each other as did types 4 and 5. Peptide 3 was unique (Figure 4.2a). Their exact amino acid sequences were not represented in MAP1. This was in accordance with the way in which *in vivo* excision from the λ library would be expected to exclude *map*1 DNA. Parts of the overlaps between the different peptides could, however, be aligned to the same region on MAP1. Since the fusion phages reacted strongly with the MAP1 antiserum in ELISA and western blots, these peptides are therefore likely to represent mimics of antigenic sites on MAP1. Even though they did not completely match the authentic amino acid sequence, they nevertheless indicate the position of a putative epitope on MAP1. The unique peptide (type 3) could neither be aligned, nor was it recognised in either immunoassay. It may thus represent a less strongly binding mimic. Screening the λ library with oligonucleotide probes derived from their amino acid sequences showed that none of the overlapping peptides could be directly matched to any potential proteins encoded by *Cowdria* ORFs. This leads further credence to the notion that they in fact represented mimics of an authentic *Cowdria* epitope (Table 4.6).

When tracts of the *Cowdria* genome sequence became available, they were subjected to BLAST searches. The same sequences were identified as with the λ screenings. This means that had these sequences been available at the outset, it would not have been necessary to use DNA hybridisation to the original λ clones to identify them. Nevertheless, these findings confirm and validate the strategy of using oligonucleotides derived from phage displayed peptides as probes when a sequence is not available. In addition, using BLAST, a match for the type 3 peptide was found. This peptide was identical to a part of a 212 amino acid polypeptide. *Map*1 is thought be part of a multigene family. MAP1 and a member of this family (ORF2) have 43 to 47 % identity at the amino acid level, varying somewhat between the different *Cowdria* isolates (Sulsona *et al.*, 1999). In another multigene family

of polymorphic outer membrane proteins of *Ehrlichia chaffeensis*, (a close relative of *Cowdria*), the proteins are 45 to 83% similar (Ohashi *et al.,* 1998). Since the protein encoded by the type 3 ORF had no similarity with MAP1 (Bestfit), it is unlikely that this protein was part of a MAP1 family of proteins which possessed shared epitopes.

The antigenic mimics identified by the fragmented-genome approach result from the number of 'meaningless' clones in the library (Chapter 3). The level of complexity generated from a genome compared to a single gene greatly increases the probability of finding such peptides. None of the publications that describe fragmented-gene (as opposed to genome) phage display libraries has so far mentioned this phenomenon. For example, a display library derived from the BTV outer capsid protein (VP5) gene was panned with a monoclonal antibody and only specific binding was detected (Wang *et al.,* 1995). Similarly, no epitopes other than the 'real' ones were found after panning a library derived from a gene encoding a 100 kDa protein involved in autoimmunity in humans with affinity-purified antibodies (Blüthner *et al*., 1996). Petersen *et al.* (1995), working with libraries derived from the RNA polymerase, p53 and cytokeratin 19 genes, describe fragmented-gene phage display as a reliable and specific method. They detected no sequences derived from DNA inserts other than the actual epitope-containing region. The work on *Cowdria* described in this thesis has, however, shown that this is not always true for fragmented-genome libraries and that the peptides identified may not necessarily be authentic epitopes located on proteins of the organism. Nonetheless, despite this added complexity, a gene encoding an unknown protein was identified. In addition, an antigenic region on MAP1 could be located.

By panning the *Cowdria* (Welgevonden isolate) phage display library with an antiserum that recognised a range of immunodominant polymorphic proteins of the Gardel isolate, it was envisaged

that epitopes common to both isolates would be identified. IgG isolated from the antiserum and antibodies eluted from two immunoblotted proteins (22.5 and 26 kDa) yielded a total of 12 different peptides with multiple clones of three of these (E3, G4 and H1; Table 4.2) being isolated. Only one copy of each of the remaining nine peptides was selected. In contrast with the MAP1-pannings, no overlapping clones and no consensus sequences were found. The peptides could, however, be grouped into three categories, each of which was rich in certain amino acids. By screening the *Cowdria* λ library and searching the *Cowdria* sequence database, four potential *Cowdria* protein-coding sequences were identified (Table 4.6). One (M37orf) encoded a protein with a predicted size of 26.5 kDa with features in common with membrane proteins. Another (N49orf) coded for a protein with a predicted size of 26.9 kDa. This putative protein had a very high scoring match (BLAST) with the rickettsial VirB8 proteins. These are integral membrane proteins (Dale *et al.*, 1993; Kumar and Das, 2001) and a VirB8-like protein of *Bartonella henslae* was shown to be immunodominant (Padmalayam *et al.*, 2000). Both ORFs encoded potential proteins in the expected size range. They appeared to be membrane associated and were common between at least the Welgevonden and Gardel isolates. Accordingly, it may be useful to investigate them as diagnostic or vaccine reagents. Another polypeptide of much interest was coded for by H1orf. It was found to be part of a larger protein, the (estimated) 84 kDa CPG1, which has no significant similarities to known proteins. The gene is currently being tested in a DNA vaccine format (Louw *et al.*, 2002). This protein is particularly interesting since it was identified both by cloning a polymorphic DNA fragment generated by RAPD and, independently in this study with an antiserum directed against polymorphic proteins. An ORF that codes for a protein likely to be a fumarase was also identified (F12orf). As with CPG1, this protein was unexpectedly larger than the size range of the proteins from which the antibodies had been eluted.

In the final series of panning experiments, rabbit antisera raised against Welgevonden isolate proteins with the potential to induce protective immune responses were used. In this series of experiments, four different peptides were identified, one of which was selected by two different antibody preparations. Since the proteins were in a relatively small size range (18 to 24 kDa), the separation could perhaps have been inadequate with the result that a mixture of proteins was used either to produce the antiserum, or to affinity-purify the antibodies. Sequence analysis showed that three of the four peptides could not be mapped to ORFs and again probably represented mimics of *Cowdria* epitopes. One (Q23orf) coded for part of a protein similar to a rickettsial alkaline protease secretion ATP-binding protein (Table 4.6). It too, was larger than the proteins against which the antiserum was raised. Before affinity-purification on electroblotted proteins, however, the antibodies apparently recognised a range of proteins of different sizes. The protein encoded by this ORF may therefore have an epitope in common with the smaller proteins.

As described above, some of the ORFs identified in this study coded for putative proteins that were larger than those from which the panning antibodies had been affinity-purified. This might have been due to the intrinsic properties of the antibodies themselves; the presence of cross-reactive antibodies seems to be a common feature of antisera directed against *Cowdria* proteins. This was evident on western blots (Figures 4.1a, 4.5a, 4.12a and b), both with antisera and affinity-purified antibodies. ‘Specific’ antibodies cross-reacted with *Cowdria* proteins other than those against which the antisera had been raised or eluted. This apparent cross-reactivity has been described previously by van Kleef *et al.* (1993). They found that the 27 kDa protein and MAP1 apparently share common antigenic determinants. Indeed, heteroclitic antibodies which reacted more strongly with MAP1 than with the 27 kDa protein itself were elicited by the 27 kDa protein. Two different proteins of *Moraxella catarrhalis* contain the same epitope (Aebi *et al.,* 1997). Proteins of *Theileria parva* also share

epitopes (Skilton *et al.*, 1998). Thus the same epitope, or two different epitopes sufficiently similar to each other, could conceivably occur on different proteins of *Cowdria*, resulting in the identification of the ORFs encoding these larger proteins. The presence of breakdown products of the large proteins in the smaller size ranges would have a similar result. Alternatively, the larger proteins could have been post-translationally processed into smaller proteins. This is known to occur with some rickettsial proteins (Gilmore *et al.*, 1991; Hackstadt *et al.,* 1992).

In summary, the peptides selected from a fragmented-genome phage display library helped to identify six genes encoding potentially antigenic proteins of *Cowdria*. In addition, several antigenic mimics were identified. Results obtained with fragmented-genome phage libraries should therefore always be analysed with caution. This aspect of fragmented-genome phage display has not been reported on before. Previously, IgG-binding phages, selected from a fragmented-genome library derived from *Staphylococcus aureus,* were found to be in the 'wrong' reading frame (Jacobsson and Frykberg, 1995; Jacobsson and Frykberg, 1996). The DNA sequences contained in the phages were the known binding sequences, but in the wrong reading frame. It was postulated that 'ribosome slippage' caused the pIII fused to the peptide to be expressed in the 'right' reading frame. However, the *Cowdria* epitope mimics presented in this thesis were irrelevant sequences (Figure 4.4) in that the region often did not code for a potential protein in any of the three reading frames on either DNA strand. Thus the *Cowdria* phage display library behaved in many ways like a random peptide library. Despite having been generated from a targeted genome this library was thus an unexpected source of molecular diversity. This finding suggested that it could possibly be used to identify epitopes on proteins other than those of the *Cowdria* proteome (see Chapter 5).

CHAPTER 5

ANTIGENIC MIMICS IN THE FRAGMENTED-GENOME LIBRARY.

5.1. INTRODUCTION.

Identifying antigenic regions on proteins provides useful information for the design of recombinant vaccines and diagnostic reagents (section 1.1.4). Epitopes have been mapped with the help of mimotopes, i.e. amino acid sequences which mimic antigenic regions on proteins. Chemically synthesised peptides (Geysen *et al.*, 1987) and random peptides expressed on the surface of phages (Scott and Smith, 1990; Devlin *et al.*, 1990) have both yielded useful mimotopes (sections 1.1.4; 1.2.3.4 and 5).

The results presented in Chapter Four showed that, in addition to finding authentic epitopes and allowing genes encoding potentially interesting proteins to be identified, antigenic mimics could be identified in a *Cowdria* genome-targeted phage display library. The manner in which the DNA fragments were generated and cloned resulted in a diverse population of displayed peptides. Five out of every six of the phage clones in the library were 'meaningless' with regard to the *Cowdria* proteome, with the antigenic mimics therefore probably originating from these clones. This finding led to the notion that this source of diversity could be exploited in a manner similar to random peptide libraries and that epitopes on proteins other than those of the target proteome could be mapped. To test this idea, two well characterised monoclonal antibodies directed against VP7 of BTV (Lunt *et al*., 1988; du Plessis *et al.*, 1994) were used to pan the *Cowdria* display library. In addition, the library was panned with a monoclonal antibody directed against malignant catarrhal fever virus (MCFV). This was done after no antigenic mimics were found in two different random

peptide libraries (personal communication, Dr. C. W. Bremer). The aim was to find peptide mimics which could be used as diagnostic reagents. The MCFV antibody (15-A) recognises a conserved epitope on a glycoprotein complex of MCFV. The complex consists of five proteins (115, 110, 105, 78, 45 kDa), held together by disulphide bonds and was identified by immunoprecipitation. Only the 45 kDa protein was identified by western blotting. Antibody binding to the epitope was glycosylation-dependent (Li *et al.*, 1995). Lastly, the peptides affinity-selected earlier with antibodies directed against *Cowdria* proteins and which were found not to be part of proteins (Chapter 4) were analysed further. Although they were not authentic *Cowdria* peptides, they could conceivably have mimicked epitopes on the potential proteins already identified. Alternatively, they could have mimicked regions on other proteins.

5.2. MATERIALS AND METHODS.

5.2.1. Antibodies.

Monoclonal antibodies D11 and F10 directed against VP7 of BTV were provided by Dr. D. H. du Plessis, have been described (Lunt *et al.*, 1988) and their epitopes mapped (du Plessis *et al.*, 1994). Monoclonal antibody 15-A raised against MCFV was obtained from VMRD, Inc (Pullman, USA; Li *et al.*, 1995). IgG was purified from mouse ascites fluid by ammonium sulfate precipitation (section 4.2.2), desalted with a PD-10 column (Pharmacia Biotech) and further purified by a DEAE-Sepharose CL-6B column as instructed by the manufacturer.

5.2.2. Affinity selection.

The *Cowdria* fragmented-genome phage display library was panned essentially as in 4.2.3, the phages PEG precipitated (section 3.2.4) and their DNA sequences analysed (sections 3.2.7 and 4.2.6). Pannings with the anti-MCFV monoclonal antibody were performed by Dr. C. W. Bremer with a few modifications. In the first two rounds, purified IgG at a concentration of 10 µg/ml was used to coat a Petri dish. To ensure release of all binders, phages were eluted with two different buffers. Firstly, 1 ml 0.1 M triethylamine (pH 12) was added and the dish was incubated at room temperature for 10 minutes. The supernatant fluid was transferred to a tube containing 0.5 ml 1 M Tris pH 7.4. The dish was washed with sterile distilled water and eluted as before at pH 2.2 (section 4.2.3). In the third round panning, two microtitre wells were coated with 10 µg/ml IgG. In one of the wells the phages were eluted at pH 12 (180 µl 0.1 M triethylamine was neutralised with 90 µl 1 M Tris pH 9) and the second well eluted at pH 2.2.

5.2.3. Antigenic reactivity.

Phages selected with the anti-MCFV monoclonal antibody were tested for immunoreactivity. Spotblots were made and treated essentially as western blots (section 4.2.2). Phages were grown in 1.5 ml LB broth containing 40 µg/ml tetracycline overnight and PEG precipitated once (section 3.2.4). The pellet was resuspended in 5 µl TBS and 1 µl was spotted on a membrane (Hybond C, Amersham). The blots were rinsed in PBS and blocked in PBS2%MP. Anti-MCFV IgG was diluted in PBST1%MP to 10 µg/ml and incubated at room temperature for 1h. Anti-mouse IgG peroxidase conjugate was used with the SuperSignal chemiluminescent substrate (Pierce) and the image captured with a Lumi-Imager (Roche).

5.2.4. Sequence analysis.

The DNA sequences encoding peptides which were not in the context of ORFs (sections 4.3.2 and 4.3.3) were subjected to a BLASTN search of the *Cowdria* genome database. The partial sequence matches were analysed to identify ORFs containing sequences encoding amino acids similar to the peptides. Further analysis was done by eye to look for partial matches in the already identified proteins for sequences which could possibly be mimicked by the selected peptides.

5.3. RESULTS.

5.3.1. Affinity selection with monoclonal antibodies against VP7 of BTV.

The *Cowdria*-derived phage display library was panned with two monoclonal antibodies (D11 and F10) both of which were directed against VP7 of BTV and bind to the same epitope (du Plessis *et al.*, 1994). These antibodies were used in panning at a concentration of 10 μg/ml which was decreased to 1 μg/ml in the third round. After round three, ten randomly picked clones from each panning were sequenced. All the clones selected by both antibodies contained a DNA insert encoding EMMQYPTLLSLH. A tract of six amino acids (QYPTLL) was strikingly similar to the sequences identified earlier by screening a random hexapeptide library and by scanning overlapping peptides (Table 5.1). Six of the total of eight peptides selected either from the *Cowdria*-derived phage display library or the random hexapeptide library (Scott and Smith, 1990) contained the QxPxLL motif. In addition, tyrosine (Y) in the second position was present in five of the peptides. The authentic epitope on VP7 of BTV is QYPALT. It thus had four amino acids in common with the peptide selected from the fragmented-genome derived phage library. This experiment provided

the first evidence that the *Cowdria* fragmented-genome phage display library could be used to identify epitopes on antigens not derived from the *Cowdria*-genome.

Table 5.1. Peptides selected with antibodies against VP7 of BTV.

Library	**Peptides selected by monoclonal antibody:**	
	D11	**F10**
Cowdria fragmented-genome	EMM**QYPTLL**SLH	EMM**QYPTLL**SLH
Random hexapeptide (du Plessis *et al.*, 1994)	QFPALL QYPSLL QYPVLV	QFPALL QYPSLL QWPAVL

5.3.2. Affinity selection with a monoclonal antibody against MCFV.

With the aim of finding peptides to replace cell-cultured MCFV antigen in a serodiagnostic test, two random peptide phage display libraries were screened with the monoclonal antibody (15-A) directed against a conserved epitope on a 45kDa protein in a glycoprotein complex of MCFV (Li *et al.*, 1995). These yielded no convincing consensus sequences or good binders. When three rounds of panning were performed with the *Cowdria*-derived phage display library, all with the monoclonal antibody at a concentration of 10 µg/ml, an increase in output was seen after round three (Table 5.2). Not all phages were released at pH12, however, since in rounds one and two the pH2.2 elution followed the pH12 elution and additional phages were released (Table 5.2). In round three, two

separate pannings were done and the phages in each eluted at a different pH. Single clones from round three were analysed to determine if the clones differed, depending on the two elution procedures. Since the aim was to identify peptides that bound strongly to the monoclonal antibody, they were tested for immunoreactivity by spotting PEG-precipitated phages onto a membrane. Clones released with both pH2.2 and pH12 reacted strongly with the 15-A monoclonal antibody (Figure 5.1a). The sequence of the inserts of 26 of the strong binding clones was determined (Figure 5.1b). All the phages released at pH12 contained the sequence LDLIEDIYKKYQKDNNSVSLGWRKFFSSEL (L30L). Three peptides were identified among the pH2.2 released phages, one of which was the same as above. Seven clones expressed the peptide HSWVLENYVAY (H11Y) and four clones ILRYFYELKTELVPYHI (I17I). Thus it seems that with monoclonal antibody 15-A, eluting the phages at low pH would have been sufficient since all three sequences were released under these conditions. In contrast only one sequence was identified using a pH 12 elution buffer. There was no common motif between the three peptides and they could not be aligned to the known sequences of MCFV. These peptides thus conceivably represent a carbohydrate epitope since the 15-A monoclonal antibody binding to the glycoprotein complex was found to be glycosylation-dependent (Li *et al.*, 1995). The three peptides had similarities with previously identified carbohydrate mimics in which aromatic amino acids seem to play a prominent part. For instance peptide L30L had the VxLxWR motif in common with the meningococcal group A polysaccharide epitope mimic, VRLSWR (Grothaus *et al.*, 2000). Peptide I17I, on the other hand, contained the motif YxY which bound to anti-carbohydrate antibodies directed against *Cryptococcus neoformans* (Valadon *et al.*, 1996). Also, H11Y and a mimic of the group B streptococcal type III capsular polysaccharide (Pincus *et al.*, 1998) had D/ExxVxY in common, where aspartic acid (D) and glutamic acid (E) are both acidic amino acids. Thus these peptides probably mimic a carbohydrate moiety on the MCFV glycoprotein complex. Since they react strongly with the monoclonal antibody (Figure 5.1), they can conceivably be used to develop new diagnostic reagents for MCFV (not the scope of this thesis). This experiment consequently provides

more evidence that the molecular diversity inherent in the *Cowdria* fragmented-genome phage display library can be exploited to identify epitopes on antigens other than those of *Cowdria*.

Table 5.2. Results of pannings with monoclonal antibody 15-A directed against MCFV. Outputs expressed as the number of colony forming units (CFU) and output increase ($round_n/rRound_{n-1}$) after each round of panning are shown. In round one and two the pH12 and pH2.2 elution was consecutive and in round three two separate pannings were done.

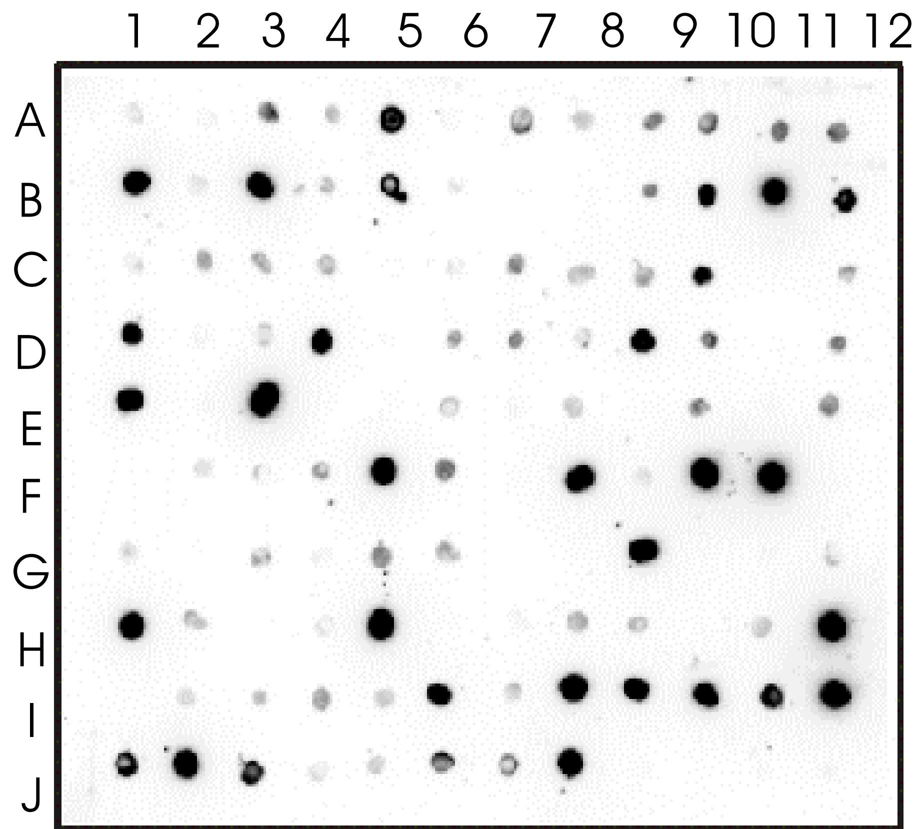
	pH 12		**pH 2.2**	
	Output (CFU)	**Output increase**	**Output (CFU)**	**Output increase**
Round 1	7×10^5	-	9.1×10^4	-
Round 2	2.35×10^4	none	4.81×10^3	none
Round 3	2.01×10^6	85x	2.21×10^6	459x

5.3.3. Potential *Cowdria* antigenic mimics.

As discussed (section 4.3.1), some peptides selected with the MAP1-antibodies reacted strongly with the same antibodies in immunoassays. These peptides were regarded as mimics of MAP1 epitopes since they were not identical, but could be aligned to a region on MAP1 (Figure 4.2b). Thus an antigenic region on MAP1 could be identified with the aid of these peptides.

Antibodies directed against *Cowdria* proteins other than MAP1 were also used to select peptides

a)



b)

Peptide sequence	pH 2.2 released	pH 12 released
HSWVLENYVAY	A3 A5 B5 B12 D1 D9 J6	-
ILRYFYELKTELVPYHI	B3 B10 D4 J3	-
LDLIEDIYKKYQKDNNSVSLGWRKFFSSEL	J2 J8	E1 E3 F5 F8 F11 G9 H1 H5 H12 I8 I9 I10

from the fragmented-genome library with the aim of developing a method to identify new genes. Figure 5.1. Phage clones selected with a monoclonal antibody (15-A) directed against MCFV. a) PEG precipitated phages spotted on a membrane and exposed to 10 mg/ml 15-A IgG. Rows A to D, I11-12 and J1-8; phages released by pH2.2. Rows E to H, I7-10; phages released by pH12. I1; fUSE2 negative control, I2-3; uninfected cell lysate, I4-5, 1,5 ml and I6, 3 ml MCFV infected cell lysate. Super Signal chemiluminescent substrate was used and the image captured by a Lumi-Imager. b) Sequence of peptides. Clone names refer to spot blot.

In contrast to the MAP1 studies, amino acid sequences were not available for direct comparison with the peptides. Twelve peptides were selected by the antiserum directed against variable immunodominant proteins of the Gardel isolate. Subsequently four genes were identified, with the selected peptides forming part of the potential proteins encoded by these genes (sections 4.3.2.4 and 5). The question arose as to whether the remaining eight peptides that were also selected in fact represented mimics of *Cowdria* antigenic determinants and whether they could be used to identify additional epitopes and/or protein encoding genes. Since there were similarities between these peptides (Figure 4.6), the possibility existed that some could have been mimicking each other, epitopes on the identified proteins, or epitopes on as yet unidentified *Cowdria* proteins. These possibilities were investigated *in silico* in two ways; the partial DNA matches found by searching the existing (incomplete) *Cowdria* database with BLAST were analysed (sections 4.3.2.5. and 4.3.3.1) and the amino acids sequences encoded in the identified ORFs were scanned for regions matching the peptides by eye.

Regions on the *Cowdria* genome were indeed identified which contained ORFs with sequences similar to those encoding the peptides. For example, peptides J15 and M37 could be aligned to the sequences encoded by J15orf-2 and M37orf-2 (Figure 5.2). The protein coded for by J15orf-2 and the peptide J15 had the sequence NSKFVxIxxK in common. This is a 287 amino acid protein with a predicted size of 31.6 kDa. Its sequence had no significant similarities with known proteins. Similarly the M37 peptide sequence could be aligned to a region on the M37orf-2-encoded protein, having five out of eight identical amino acids (TNHNxxxS; Figure 5.2). M37orf-2 encodes a partial protein of 141 amino acids. This predicted protein had 40% identity with an unknown *R. conorii* protein of 1033 amino acids (score 67, E value 4e-11) and 43% identity with a hypothetical protein of *R. prowazekii* of 295 amino acids (score 65, E value 1e-20). The smaller *R. prowazekii* protein seems to be identical to a part of the larger protein. Accordingly, while it is plausible that peptides

J15 and M37 contain mimics of epitopes on the proteins encoded by the above ORFs, this needs to be confirmed by additional studies. Since the immunoreactivity of the peptides displayed on the phages in assays other than the panning could not be shown conclusively, the ORFs will need to cloned, expressed and tested in immunoassays. Synthetic peptides can also be used to confirm the serological reactivity of the epitopes.

Genome match J-2 translated (J15orf-2)

```
J15orf-2    K Q V E R Q Q N M N N N K D Q N N A C Y Q L D Y F E F D G K N Y P S L T K K N R D F
J15orf-2    I E A I L S I D S N Y R Q Y K N N T I H N H I Y E I T A D S L S E I F K R L N A E N

J15                                                            I I S F T N S K F V Y I L K K R
J15orf-2    E D F E T T L L K I I V L I D S S N S T H L S T S F Y T G D D N S K F V N I T N K N

J15orf-2    Y N V L N G L K E M R D R I P K K I N N A N E L I E A I K I P F N P Y D K N H I F N
J15orf-2    I M N E P T I K R E K N K K Y S E V N E A Y I Q K N K K F N T S F V S K F L S Y C Y
J15orf-2    S Y L I G N N N I Y S K Y D N V L V K Y L P I Y Y A H Y K N S E N K I D K N D K N K
J15orf-2    Y I L S N L N K E Q L R K Q G I E N L K I G F A K L Y K K Y N D D I Q *
```

Genome match M-2 translated (M37orf-2)

```
M37ORF-2    N M I N F E L T T Q N L Y T R E K L L E L D N I F L S Y L Q S H D K S L Y Q T L I

M37         N I L G R I T N H N V S L S K A
M37ORF-2    L A R Q G N T N H N N T S S I I E L S Y V V E D F I S Q L F N I E D E V I L Q K N

M37ORF-2    I H K E F I E I Y K C K R L F V Q R Y A L K K Y P N I N N L N I E E I T N K I S Q
M37ORF-2    I F S L P N S R K R I L *
```

Figure 5.2. Alignment of peptides with potential *Cowdria* proteins. Similar colours indicate similar amino acids. '*" indicates a stop codon.

The translated sequences of all the potential ORFs identified in this study were scanned for the presence of regions matching the peptides selected by phage display (Figure 5.3). Together with F12, peptides K18, 47 and M37 could all be aligned with the putative protein encoded by F12orf. K18 had the sequence YxxQILQ in common with the protein. In addition there were similar amino acids following this sequence. Peptides 47 and M37 each had three identical, and three similar amino acids in this region (Figure 5.3). Apart from M37, all the peptides were initially part of group

C (Figure 4.6). M37 also aligns to a slightly different region of the F12orf-encoded protein, possibly indicating that it was recognised by antibodies in the population with a different specificity. Also, K18 and F12 could be aligned to M37orf-encoded protein (Figure 5.3). The critical amino acids were N/QILQxI/V, with x being basic amino acids in two of the three sequences. In addition, peptide 47 could be aligned to another region on the protein. Five of the eight amino acids were identical and three were very similar.

F12orf

```
F12orf  R L S K V M R K E K D S L G E V N V P A C H Y W G A Q T Q R S I D N F K I G
F12orf  S E K M P K P L I R A M G I V K L A A A R V N M K N G D I N E V I G N A I C
F12orf  N A A A E V I D G K L D N E F P L V V W Q T G S G T Q T N M D V N E G I A N
F12orf  R A I E I L G G E K G S K V P V H P N D H V N Y S Q S S N D T F P T A M H I
F12orf  A T V S E T E N Y L L P S L K N L Y D A L H S K S I A F Q N I V K V G R T H

M37                                        N I L G R I T N H ..
F12orf  L Q D A T P L T L G Q E F S G Y A Y Q I L Q G I G R I K S A L S N L L E L A
F12                                  G Y A Y Q I L Q G I G R I K
K18                         .. P L F Y F Q Q I L Q H V P Y S S
47                          .. N V V G P Y G L L Q P

F12orf  Q G G T A V G T G I N S R K Q F D V H I A N E I K K I T G F N F V S S V N K
F12orf  F E A L A T H D A L V E F S G A L N V L A V S L M K I A N D I R L L S S G P
F12orf  R C G I G E I I L P A N E P G ..
```

M37orf

```
M37orf  Y K I N M L N I F M K K S I Y F V I S I T F M L F S F D A S S N S V Y I K G N Y N L G
M37orf  I N Y S D S F K S E Y V N Y E R L S P D F A V A I G Y K L Q N G L F C D L E M R Y A N

F12                             G Y A Y Q I L Q G I G R I K
M37orf  I R P M I N K L S N F E H I D L L N I L Q R I T N H N V S L S K A E N I L T I N A I T
M37                                    N I L Q R I T N H N V S L S K A
F12                                 G Y A Y Q I L Q G I ..
K18                                   .. F Q Q I L Q H V P Y S

M37orf  T L I N V G Y S Y V F N D K F R G Y F T Y G V G I G G L L N Y Q G F K S S L S T Y Y G
pep 47                                .. G P N V V G P Y G L L Q P

M37orf  I S M Q S E I G V C Y I Y N S K I D V C I G Y N Y L K N Y W K Y E T D R I P D E D G N
M37orf  T V M Y N F Q D F Q L N S H T I F L G L N V L F *
```

Figure 5.3. Alignment of peptides with proteins encoded by F12orf and M37orf. Similar colours indicate similar amino acids. '..' indicates a partial sequence and '*' indicates a stop codon.

5.4. DISCUSSION.

The overall aim of this study was to determine whether a fragmented-genome phage display library could identify epitopes and/or protein-encoding genes of *Cowdria.* The random fragmentation of the DNA used to construct the library introduced new complications brought about by added diversity. This, together with the difficulties in getting antibodies that recognise individual proteins, made data interpretation somewhat difficult. Nevertheless, the *Cowdria*-derived phage display library turned out to be a good source of molecular diversity and the identification of MAP1-mimics allowed an antigenic region on this protein to be mapped (section 4.3.1). Peptide sequences which bound to monoclonal antibodies not related to *Cowdria* were also identified. Thus, apart from identifying *Cowdria* epitopes, the fragmented-genome library acted as a 'universal' source of peptides for identifying antigenic mimics.

Bioinformatic studies on the peptides selected with antibodies directed against unknown variable immunodominant *Cowdria* proteins showed that apart from four genes (Chapter 4), two additional potential ORFs could be identified (Figure 5.3). These putative proteins contained sequences that could have been mimicked by the peptides. Thus an epitope on, for example, the protein encoded by J15orf-2 could be responsible for eliciting an antibody that recognised the very similar J15 phage-displayed peptide. Since two ORFs were identified by peptide M37, one encoding the complete peptide sequence (M37orf) and the other a very similar sequence (M37orf-2), the proteins encoded by these two ORFs could conceivably share antigenic properties. Additional peptides could be aligned to the positions of the F12 and M37 peptides within the proteins encoded by F12orf and M37orf. These regions could therefore represent antigenic sites on the proteins. Since the same peptides can be aligned to both proteins, it seems that the 26 kDa M37orf-encoded protein and that of F12orf share epitopes. Such shared epitopes could conceivably explain the confusing serological

cross-reactions between *Cowdria* proteins, a well established phenomenon that has been observed on western blots (van Kleef *et al.*, 1993) and in this study.

Overall, the results confirm that, where possible, a panel of peptide libraries should be screened to find optimal binders and that no one library performs 'best' (Bonnycastle *et al.*, 1996). The random peptide libraries used in initial pannings with an MCFV antibody (personal communication, Dr. C.W. Bremer) either contained 2 x 10^8 clones expressing peptides of six amino acids (Scott and Smith, 1990) or 1 x 10^9 clones expressing peptides of 17 amino acids (XCX_{15}, Bonnycastle *et al.*, 1996). Thus both contained more clones than the *Cowdria* library, even if the 'sliding window' of hexamers (approximate size of an epitope) contained in each clone is taken into account (Kay *et al.*, 1993). The XCX_{15} library theoretically contains 12 different hexamers, thus increasing the diversity of the library to 1.2 x 10^{10} different peptides. At a conservative estimate, the *Cowdria* library had an average peptide size of 15 amino acids (45bp, section 3.3.2), i.e. ten hexamers per clone and 5.5 x 10^7 six amino acid combinations (5.5 x 10^6 meaningful and meaningless clones x 10). There are 6.4 x 10^7 possible hexamer combinations of the 20 natural amino acids. Theoretically therefore the *Cowdria* library could not contain all possible peptides. It has been proposed, though, that longer inserts are a richer source of diversity due to the sliding window of short peptides (Kay *et al.*, 1993). The context of each hexamer is also different within a larger peptide which can potentially assume a secondary structure independent of pIII (or other vector-derived fusion). Longer peptides can also have multiple contact sites and so act as discontinuous epitopes. In the XCX_{15} library, a cysteine (C) was incorporated to impose some structural constraints on the peptides. For this particular anti-MCFV antibody, binding partners were found in the genome-targeted library whereas those sequences were either simply not present or in the right context in the other libraries.

The identification of these peptides therefore shows that the *Cowdria*-derived phage display library has 'added value' since it could serve as a universal epitope or antigenic mimic identification reagent. These could be used for epitope mapping and to identify surrogate antigens in diagnostic assays. With regards to *Cowdria,* additional potential antigenic regions and genes were identified. Fragmented-genome phage display libraries are relatively easy to make and the diversity can be enlarged by using DNA from different sources. Random peptide phage display libraries on the other hand are generated by cloning random oligonucleotides. This is a totally synthetic system and an advantage is that the insert size can be controlled and structural constraints incorporated to control the context in which peptides are expressed. In the quest for optimal binding peptides, it is probably advisable to exploit both these avenues.

CONCLUDING DISCUSSION.

Phage display technology has contributed greatly to epitope mapping, mimotope identification, research into protein-protein interactions and the isolation of recombinant antibodies. This thesis describes the use of phage display to locate genes that encode antigenically relevant proteins of the rickettsial parasite, *Cowdria ruminantium*. Despite display technology having been widely used, there are no reports of it being applied to intracellular bacteria. It has, however, been used with other parasites. Indeed, one of the first applications was to express known *Plasmodium falciparum* repeat sequences (de la Cruz *et al.*, 1988) while more recently, a random peptide library was used to identify epitopes of another intracellular parasite, the eukaryotic *Theileria annulata* (Prickett and Hall, 2000). Rather than using random peptide phage display libraries, a targeted approach was followed with *Cowdria*. Epitopes have been mapped using fragmented-gene libraries derived from single genes (du Plessis *et al.*, 1995; Wang *et al.*, 1995; Petersen *et al.,* 1995; Blüthner *et al.*, 1996) and this study examined the extension of the method to a genome, as opposed to a single gene. An important goal was to ascertain whether genes encoding antigenically important proteins of *Cowdria* could be identified. Owing to the size of the *Cowdria* genome it was anticipated that the complexity of the task would be greater than e.g. identifying a viral epitope, yet still be within the realms of the technology.

Cowdria is an intracellular bacterium that is not easy to cultivate in cell culture. Owing to its intimate association within the cell, it is extremely difficult to isolate DNA without co-isolating host cell (bovine) DNA. Small amounts of ‘clean’ DNA could, however, be obtained after immunoaffinity purifying the organisms (Brayton *et al.*, 1997). As a first attempt at identifying genes encoding antigenic proteins, a λZAPII genomic library which represented the genome was constructed. A cDNA expression library could not be constructed owing to the fact that prokaryotic *Cowdria* mRNA is not polyadenylated. Despite several attempts, screening with antibodies raised against MAP1 and

against sonicated *Cowdria* yielded no antigenically reactive proteins. A possible reason was that the library comprised random DNA fragments obtained from a prokaryote, ie. it was not a cDNA library derived from mRNA. Screening depends largely on high level protein expression. Clones containing random DNA fragments may contain promoter sequences upstream of the genes, which then drive expression, but probably not to the levels attainable under control of the lacZ promoter. For example, a clone containing an insert coding for MAP1 was isolated from this library by using a DNA probe. It was shown to express the expected protein. Nevertheless, plaque lifts of the library screened with MAP1 antiserum revealed no clones that detectably expressed MAP1. Notwithstanding this 'traditional' library's limitations, it proved to be a useful tool in *Cowdria* research. It is currently used as a source of DNA for the *Cowdria* genome sequencing project at Onderstepoort and is the basis for the filamentous phage display library described in this thesis.

With bacteriophage λ expression libraries, the quality (affinity, avidity or specificity) of the antibodies used to probe for a target protein affects the efficiency with which positive clones are detected. Since it is necessary to use large volumes of antibodies to screen λ libraries on membranes, it was not feasible to affinity-purify sufficient antibodies from available sources for screening. In contrast, phage display libraries allow a large number of clones to be screened with a small amount of antibodies. This made it feasible to use immunoglobulins released from electroblotted *Cowdria* antigens. A possible drawback of phage display libraries derived from a fragmented genome is that the antibodies need to recognise linear epitopes, ie. stretches of four to five contiguous residues (Fieser *et al.*, 1987) to allow specific panning. Most epitopes of globular proteins are thought to be discontinuous (Atassi, 1975; Atassi *et al.*, 1976; Benjamin *et al.*, 1984). Thus in theory, the immunopurification of antibodies on denatured proteins should enrich the pool of antibodies for those that recognise the required linear epitopes (du Plessis *et al.*, 1995; Blüthner *et al.*,1996). With random epitope libraries, it is highly likely that antigenic mimics will be identified (Scott and Smith, 1990).

These can also mimic discontinuous determinants, which can be difficult to 'align' to the target sequence. In any event, panning random peptide libraries with antibodies directed against *Cowdria* proteins would probably not have been useful due to the limited DNA sequence information that was available at the onset on this study. Mimotopes selected from random peptide libraries have proved to have potential as diagnostic reagents and for vaccine development (1.2.3.4, 5 and 9), but not for the gene identification which was the aim of this study. For these reasons a genome-targeted approach was followed.

The *Cowdria* fragmented-genome phage display library contained 1.67 x 10^7 clones, each expressing a peptide of at least 15 amino acids. It was thus theoretically large enough to represent the entire *Cowdria* genome. Since it was made using *Pvu* II-digested plasmids *in vivo* excised from λ clones, genes containing a *Pvu* II site were unlikely to be included. Accordingly, it lacked phages displaying peptides derived from the *map*1 gene. This state of affairs was not regarded as a drawback since it had the potential to facilitate a search for proteins other than the immunodominant MAP1, an already well-characterised protein (1.3.8 and 1.3.9). Panning with an antiserum directed against MAP1 did, however, yield a number of binding peptides. Despite their not having sequences identical to the target protein, four of the five sequences were sufficiently similar to be aligned with those of MAP1. None of these peptide sequences could be assigned to an ORF. They were therefore most likely to have resulted from antigenic mimicry arising from the wide diversity of peptides expressed by randomly digested DNA fragments which were cloned in different orientations and reading frames. Notwithstanding their origins, the alignments allowed an antigenic region (Figure 4.2) to be identified on MAP1. Thus, a fragmented-genome phage display library containing small inserts can be complex enough to contain potential antigenic mimics in addition to the authentic continuous epitopes. This possibility therefore needs to be taken into account when analysing peptide sequences isolated from such a library.

The fragmented-genome phage display library allowed several antibody-reactive peptides to be isolated. Taking the potential of finding antigenic mimics into account together with the absence of adequate sequence information, an additional step was introduced to place these sequences in the context of a protein. This was done by screening the λ library for clones containing peptide-encoding sequences and later, when more *Cowdria* DNA sequence information became available, by *in silico* searches. Twelve peptide sequences were identified by panning with antibodies directed against variable and immunodominant proteins of the Gardel isolate. With one exception, they had no similarity to any known proteins. Placed in context of the *Cowdria* genome, four genes were identified which coded for potential ORFs. Some had similarities with known proteins (Table 4.6). The remainder of the regions that contained the peptide-encoding DNA sequences did not encode ORFs of which the peptide sequences formed a part. These phage-displayed peptides were therefore probably mimics of *Cowdria* epitopes. Of the four genes identified *in silico*, two (M37orf and N49orf) coded for potential proteins in the size range against which the antiserum was raised. These proteins had similarities with known ehrlichial and rickettsial proteins. They are also candidates for inclusion in vaccine studies since they are immunoreactive and common to at least two *Cowdria* isolates. The other two genes (F12orf and H1orf), on the other hand, coded for proteins much larger than the expected size range. This added another level of complexity, probably brought about by the *Cowdria* proteins which share, or have cross-reactive epitopes (van Kleef *et al.*, 1993). These larger proteins could have had epitopes similar to those on the proteins against which the antiserum was raised. Another possibility is simply the presence of breakdown products in the protein blots used to elute the antibodies or in the protein preparation used to produce the antiserum. The *in silico* results nonetheless support the notion of cross-reactive epitopes. For example, the same peptides (F12, K18 and M37) could be aligned to both M37orf and F12orf (Figure 5.3).

The question arises whether it is possible to ascertain whether phage displayed peptides are authentic

epitopes or antigenic mimics when either no, or only very limited genomic DNA sequence is available. The GC content of the encoding DNA inserts can distinguish between coding and non-coding regions (Table 4.7). In general, if a sequence has a GC content of 24% or lower, it is most probably from an intergenic region. This corresponds to what de Villiers (2001) recently found from analysis of a 14 848 bp *Cowdria* genome fragment where intergenic regions had an average of 23% and coding regions an average of 31% GC content. The phage displayed sequences encoding authentic epitopes had a GC content with an average of 31.7%. The complication is that an antigenic mimic could also be derived from a region encoding a protein. These sequences also have a high GC content, but can be encoded from a different reading frame, from the reverse strand or from a combination of these two situations. Thus little can be predicted about a selected peptide purely from its GC content. There may of course be exceptions, for example M37, a *Cowdria* epitope encoded by DNA with a GC content of 22% (Table 4.7). The DNA fragments encoding the peptides are short and this could result in the sometimes 'meaningless' statistics with regards to GC content.

This implies that clones derived from coding regions could still represent peptides mimicking *Cowdria* epitopes. Codon usage might also be an aid in determining whether DNA encoding the peptides followed the *Cowdria* codon usage pattern. Codon usage values determined from analysis of 36 *Cowdria* genes (de Villiers, 2001) were used to compare the codons used in the phage displayed peptides. In general the authentic peptides were encoded by a high percentage of 'expected' codons, but once again there were several exceptions (L24, Type 3, I8). The problem was that codon usage values were determined from 36 genes, some of which were multiple copies of the same gene, consequently, the results were skewed towards those. It is known the codon usage differs between genes that are expressed at high and low levels (Mathe *et al.*, 1999). When the full genome is analysed these values will be more meaningful, but then direct genome comparisons will quickly reveal the origin of the peptide encoding sequence. Another complicating factor is that the peptides

are relatively short and a statistically relevant picture cannot be obtained. The method of analysis was perhaps flawed. If a rare codon was used it got a negative score and the sequence would get a low score. This does not mean, however, that *Cowdria* never uses these codons. As a consequence, this kind of analysis turned out to be less than meaningful. It seems therefore that it would not be prudent to decide purely on the basis of sequence data whether peptides are authentic epitopes or epitope mimics. Finding the encoding gene is therefore still an essential step, especially for genomes with no available sequence information. There are always clones that do not follow the general pattern and which would be disregarded if the above criteria were used.

Would the regions identified by phage display have been predicted as being antigenically important by using sequence analysis computer programs? The positions of the potential linear epitopes identified within each putative protein were analysed using two programs designed to predict antigenicity (section 4.3.4). Four of the six regions would have been predicted as being antigenic by both programs with the Antigenic Site (based on experimental data) being the more accurate. Indeed, it would have identified all six regions (Table 4.7). These regions were, however, not all located in major peaks of the prediction graphs. It should be remembered that the antibodies used in this study were directed against denatured, PAGE-purified proteins. It is therefore perhaps not to be expected that these antibodies and their corresponding epitopes follow the general rules for proteins.

Since antigenic mimics were selected from the *Cowdria* phage display library, it became apparent that the molecular diversity in the targeted library could perhaps be exploited as a ‘universal’ tool in the same manner as a random peptide library. Indeed, antigenic mimics were identified with monoclonal antibodies directed against proteins other than those of *Cowdria*. A known epitope was identified with antibodies directed against VP7 of BTV. In addition, peptides which probably mimic a carbohydrate moiety on the MCFV glycoprotein complex could be found.

Several factors such as the type of library, the vector, the antibodies used for panning and how the panning was performed all have an influence on the outcome of the results obtained with phage display. In this study a targeted approach was followed. From the many different vectors available, the phage vector, fUSE2 that results in a pIII fusion protein was chosen. Gene III vectors allow larger DNA inserts to be cloned compared to gene VIII fusions. More efficient cloning is achieved with phagemid vectors but phage vectors result in multivalent expression since all pIII copies are fused to the foreign peptide. This theoretically allows a higher diversity of peptides to be selected since the phage can attach multivalently. Increased avidity then permits weak, but relevant, binders to be isolated (Sparks *et al.,* 1996; Menendez *et al.*, 2001). The fUSE2 vector was successfully used with gene-targeted libraries and was thus chosen to extend the technology to the much larger genome-targeted library described in this study. One disadvantage was that only 1/18 clones expressed the authentic peptide while 6/18 of the original clones produced functional phages. In systems such as pJuFo, where only one end of the insert needs to be in-frame, a higher percentage of the clones are functional (Palzkill *et al.*, 1998). This is especially useful for expressing cDNA clones that contains stop codons at the end of each reading frame. Nevertheless, this 'disadvantage' has proven to be an advantage in some applications due to the diversity that resulted from all the 'non-authentic' clones (Chapter 5).

Panning procedures can also dictate results. In this study the antibody was immobilised directly onto a plastic surface. This has worked well with gene-targeted libraries (Wang *et al.,* 1995; du Plessis *et al.*, 1995), but there are many other ways of panning. For instance, Parmley and Smith (1988) used biotinylated antibodies and immobilised them by trapping on streptavidin coated plates. Antibodies have also been trapped with anti-IgG antibodies, protein A and G. This approach allows phages and antibodies to react in solution. The 'conservative' approach is solid-phase panning followed by

optimisation (Menendez *et al.*, 2001). Since strong binding peptides were isolated in the initial pannings with the MAP1 antiserum, this protocol was followed for the remainder of this study.

In summary, the fragmented-genome phage display library was better suited to identifying gene products that react with antibodies than the genomic bacteriophage λ library. Phage display actually enabled several potential ORFs to be identified. There were, however, limitations. These was unavoidable due to the diversity created by random fragmentation of the *Cowdria* genome. Even with added advantages such as high selectivity and the linked genotype and phenotype, phage display techniques did not always give clear-cut results. Indeed, the added diversity made the results even more complex. Another problem is that appropriate immunological tools to identify antigenic proteins of *Cowdria* are not available. Indeed, this study shows that antibodies used for panning play an important role in the outcome of any experiment. For example, the high titre MAP1 antibodies gave clear-cut results compared to those obtained with the other antibody preparations. In addition, affinity purifying the antibodies eliminated confusing results resulting from background clones (section 4.3.2). The monoclonal antibodies raised against *Cowdria* so far are all directed against either the well characterised MAP1, or two other proteins of 40 and 43 kDa (Jongejan *et al.*, 1991; Shompole *et al.*, 2000). Most available polyclonal antibodies cross-react with many of the other *Cowdria* proteins. Despite several attempts, it was not possible to raise high affinity antibodies to electrophoretically separated protein fractions (M. van Kleef, personal communication). A possible way of addressing the paucity of useful *Cowdria* antibodies is to exploit phage display in yet another way, ie. by selecting recombinant antibody fragments from display libraries expressing single chain antibody fragments (1.2.3.10). In this manner antibodies can be obtained without immunological biases imposed by the immunised animal, especially if synthetic random peptides are included in the CDRs (Hoogenboom and Winter, 1992; Nissim *et al.*, 1994). Initial experiments with a semisynthetic human antibody library (Nissim *et al.*, 1994) yielded clones with irreproducible binding

characteristics (Fehrsen, unpublished results). Improvements in the display of the antibody fragments (Rondot *et al.*, 2001) or using antibody libraries derived from hosts other than human (Davies *et al.,* 1995) may increase the chances of finding novel antibodies which recognise *Cowdria* proteins. This may still not solve the cross-reactivity problem which seems to be an inherent property of *Cowdria* proteins, and probably many others as well (van Regenmortel, 1998).

Notwithstanding the difficulties experienced and the sometimes unexpected results obtained, several putative protein-encoding genes were identified and still need to be characterised. Their potential role in diagnostic and vaccine development remains to be established. *Cowdria* genes are not easily expressed (4.3.2.4) and additional expression systems need to be tested. Difficulty in expressing *Cowdria* proteins could have been due to the instability of the cloned DNA, different codon usage or the toxicity of the protein product. Moreover, some recombinant proteins might not adopt the same conformation as the native protein, resulting in different characteristics (eg. immune reactivity). Proteomics is an up-and-coming field and new technologies for protein expression are likely to be developed to overcome these problems. The sequencing of entire genomes identifies many ORFs. Immunoreactive peptides identified by phage display can give 'meaning' to some of these ORFs and can help focus vaccine and diagnostic studies on immunologically relevant genes and their proteins.

SOLUTIONS AND BUFFERS

CTAB/NaCl solution:

10% CTAB in 0.7 M NaCl. Dissolve 4.1 g NaCl in 80 ml H_2O and slowly add 10 g CTAB (hexadecyltrimethyl ammonium bromide) while heating and stirring. Adjust volume to 100 ml.

DE52 (Diethylaminoethyl Cellulose; Whatman):

Add 1.5 g DE52 to 20 ml PBS, mix and let the DE52 settle. Remove the PBS and replace with 10 x PBS. Repeat this procedure until the PBS is pH 7.4. Equilibrate the slurry with ½ x PBS.

Horse radish peroxidase substrate solution:

Dissolve 60 mg 4-chloro-1-napthol (Sigma) in 10 ml ice cold methanol and add 60 µl hydrogen peroxide to 100 ml cold PBS. Mix the two solutions just before use. The membranes are incubated in the substrate for 10 to 60 minutes for the colour to develop, rinsed in water to enhance the colour, dried and stored in a dark place.

Klenow fragment buffer:

1mM each dNTP, 6mM Tris (pH 7.5), 6mM sodium chloride, 6mM magnesium chloride, 0.05% gelatin, 1mM DDT.

LB:

Dissolve10g bacto-tryptone, 5g yeast extract, 10g NaCl in 1l H_2O, autoclave.

LB plates:

Dissolve10g bacto-tryptone, 5g yeast extract, 10g NaCl in 1l H_2O, add 15g agar, autoclave.

PBS:

8g NaCl, 0.2g KCl, 1.44g NA_2HPO_4, 0.24g KH_2PO_4 in 1l, pH 7.4 with HCl

PEG/NaCl:

Dissolve 100g PEG 8000(16.7%), 116.9g NaCl (3.3M) in 475ml H_2O, autoclave, store at 4°C.

SSC:

175.3g NaCl (3M), 88.2g sodium citrate(0.3M) in 1l, pH 7 with 10N NaOH.

TBS:

6.05g Tris (50mM), 8.76g NaCl (150mM) in 1l.

TE:

10mM Tris, pH 7.4 and1mM EDTA, pH 8

Top agar:

Dissolve 1g bacto-tryptone, 0.5g yeast extract, 0.5g NaCl in 100ml H_2O, add 0.8g agar, autoclave.

SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE; Laemmli, 1970)

Running buffer:
25mM Tris /192mM glycine /0.1% SDS

PAGE sample buffer:
0.5 M Tris, 2% (w/v) SDS, 10% (v/v) glycerol, 5% β-mercaptoethanol, 0.001% bromophenol blue

Stacking gel	**5%**
30% Acrylamide	3.26 ml
1% N,N'-Methylene-bis-acrylamide	2.50 ml
1M Tris pH6.8	3.75 ml
H_2O	10.30 ml
10% APS	0.2 ml
TEMED	13.4 µl
Final volume	20 ml

Separating gel	**15%**
40% Acrylamide	10.96 ml
2% N,N'-Methylene-bis-acrylamide	5.62 ml
1.5M Tris p H 8.8	7.5 ml
H_2O	5.6 ml
10% APS	0.3 ml
TEMED	20 µl
Final volume	30 ml

Separating gel	**10%**	**12%**
30% Acrylamide	9.76 ml	11.7 ml
1% N,N'-Methylene-bis-acrylamide	7.5 ml	9 ml
1.5M Tris p H 8.8	7.5 ml	7.5 ml
H_2O	4.95 ml	1.48 ml
10% APS	0.3 ml	0.3 ml
TEMED	20 µl	20 µl
Final volume	30 ml	30 ml

WESTERN BLOTTING

Transfer buffer:
40 mM Tris, 40 mM glycine, 0.4% SDS and 20% (v/v) methanol

Towbin buffer:
25mM Tris, 192mM glycin

Appendix C

PREPARATION OF BACTERIA

***E. coli* XL1-Blue MRF' plating cells (Stratagene):**

Cells were streaked out weekly from glycerol stocks for single colonies on LB plates containing 12.5 µg/ml tetracycline. Cells were grown in LB containing 0.2% maltose and 10 mM $MgSO_4$ until OD_{600} = 0.5 at 37°C or overnight at 30°C. Cells were centrifuged at 1000 x g for 10 min and resuspend in 0.5 vol 10mM $MgSO_4$.

***E. coli* K91, starved cells (Smith, 1992):**

A single colony of K91 cells was grown in 20ml LB to mid log phase (an absorbance at 600nm of 0.45) at 37°C. The cells were incubated for a further 5 minutes with gentle shaking. They were then collected by centrifugation for 10 minute at 2000rpm and resuspended in 20ml 80mM NaCl. The cell suspension was incubated for 45 minutes with gentle shaking and the cells collected as before. The final cell pellet was resuspended in 1ml cold NAP buffer (80 mM NaCl, 50mM $NH_4H_2PO_4$ pH 7 with NH_4OH) and stored at 4°C. These cells were be used for up to 5 days.

***E. coli* K91, mid log cells (Smith, 1992):**

K91 cells were grown in LB broth overnight at 37°C. One hundred microlitres was inoculated into 10 ml LB broth and grown to an absorbance at 600nm of 1.25 to 2.5 when the shaking was slowed down for 5 minutes to allow the F-pili to regenerate. These cells were used within 1h.

***E. coli* TG1, TSS competent cells (Chung *et al.,* 1989):**

A single colony of TG1 cells was grown in LB overnight at 37°C. Two ml of this culture was inoculated into 50ml LB and grown to an absorbance at 600nm of 0.3 to 0.4. The cells were collected by centrifugation at 1000g for 10 minutes. The pellet was resuspended in 5ml ice cold TSS (LB containing 10% PEG, 5% DMSO, 50mM $MgCl_2$, pH6.5). Cells were transformed by adding ligation mixture and incubating on ice for at least 30 minutes. One millilitre of LB was added and grown at 37°C for 1h before plating 300µl aliquots.

***E. coli* MC1061, electrocompetent cells (Smith, 1992):**

A single colony of MC1061 cells were grown in 5ml LB containing 100µg/ml streptomycin overnight at 37°C. The next day 2ml of the overnight culture was inoculated into 500ml LB and grown to an absorbance at 600nm of 0.3 to 0.4. The cells were chilled in ice-water for 15 minutes. The remaining steps were all done with cold solutions, centrifuge bottles and rotors or on ice. The cells were pelleted at 3000g for 15 minutes and resuspended in 500ml 1mM HEPES. The cells were centrifuged again and resuspended in 250ml 1mM HEPES and collected again. The cell pellet was resuspended in 10ml 10% glycerol and after a final centrifugation resuspended in 600µl 10% glycerol. Cells were used immediately or 50µl aliquots were stored at -70°C.

***E. coli* SOLR (Stratgene):**

A single colony of SOLR cells were grown in 5ml LB containing 50µg/ml kanamycin overnight at 37°C. The next day l00µl of the overnight culture was inoculated into 10ml LB and grown to an absorbance at 600nm of 1.

DNA and deduced amino acid sequences of peptides selected from the *Cowdria*-derived phage display library. The GC content of each DNA insert is given as a percentage.

Type 1 GC = 20%
CTTAAATATTTATCAGATAAGTATAAAATTAAGGAC
L--K--Y--L--S--D--K--Y--K--I--K--D--

Type 2 GC = 24%
TATTTATCAGATAAGTATAAAATTAAGGAGAATCTCACGTATTTC
Y--L--S--D--K--Y--K--I--K--E--N--L--T--Y--F--

Type 3 GC = 31%
TCTAAACAGTGTGATGAAATACGAGAAAAGATTAAAAAGTGTAATCTAAGGCAAGGAAAGAAGAAAAGTGCATTGTCG
S--K--Q--C--D--E--I--R--E--K--I--K--K--C--N--L--R--Q--G--K--K--K--S--A--L--S--
AAATTTACAGATCATTTC
K--F--T--D--H--F--

Type 4 GC = 40%
AAATCCAGATTAACAGAAGTAAATCTTTATATCAGCCAGAAATTTATGCCCCAACACAATCCGCTGCAAGACGCGGAT
K--S--R--L--T--E--V--N--L--Y--I--S--Q--K--F--M--P--Q--H--N--P--L--Q--D--A--D--

Type 5 GC = 45%
ATCAGCCAGAAATTTATGCCCCAACACAATCCGCTGCAAGATGTCGCATTA
I--S--Q--K--F--M--P--Q--H--N--P--L--Q--D--V--A--L--

D25 (without pBluescript) GC = 40%
GATCCCTATTCGTTAACATAAGCATATCTCCAAACCCTG
R--S--L--F--V--N--I--S--I--S--P--N--P--

E3 GC = 45%
GTGTTATTGAGAATAATGAATATGCAATGGGTACATCAGTATCTAGATCGAGGGGGGGCCCGGTAC
V--L--L--R--I--M--N--M--Q--W--V--H--Q--Y--L--D--R--G--G--A--R--Y--

F12 GC = 35%
GGTTATGCATATCAAATTTTACAAGGCATTGGCAGAATCAAA
G--Y--A--Y--Q--I--L--Q--G--I--G--R--I--K--

G4 GC = 35%
ACAGCTAAGAAAATTACTATCAGATGCGATTCAATCACCTGTTAC
T--A--K--K--I--T--I--R--C--D--S--I--T--C--Y--

H1 GC = 35%
CGAGAAGGTAATCATCAGTGTAAGGTAGTAAAATTTACAGACAAATGCAAGCATAAACCC AAAAAACACAG
R--E--G--N--H--Q--C--K--V--V--K--F--T--D--K--C--K--H--K--P--K--K--H--R--

I8 GC = 16%
TACAAAAAAATTTTCACTTATTTCATAACATCATATATCAAT
Y--K--K--I--F--T--Y--F--I--T--S--Y--I--N--

J15 GC = 18%

ATAATCTCATTCACTAATAGTAAATTTGTGTATATTCTAAAAAAANNAAAATCATTC

I--I--S--F--T--N--S--K--F--V--Y--I--L--K--K--X--X--S--F--

K18 GC = 33%

TTTCATATATTCCATCCGCTGTTCTATTTTCAACAGATTTTACAACACGTTCCTTACTC

F--H--I--F--H--P--L--F--Y--F--Q--Q--I--L--Q--H--V--P--Y--S

L24 GC = 27%

TTTTTATCTAATTCAGAAGATTTACTTTCACTTGATACCTTAGCTTTATCTTCTTTAATTGCAGGAAGCTTACTAACA

F--L--S--N--S--E--D--L--L--S--L--D--T--L--A--L--S--S--L--I--A--G--S--L--L--T--

M37 GC = 22%

AATATTTTGCAGAGGATTACTAATCATAATGTATCTTTATCTAAA

N--I--L--Q--R--I--T--N--H--N--V--S--L--S--K-

N49 GC = 26%

CGTCACAAAGTATTATTAAGTTTTTCTTTATCTAGTATTAAAATGGCTAGTTCA

R--H--K--V--L--L--S--F--S--L--S--S--I--K--M--A--S--S--

O4 GC = 34%

CTATCTGTATGTGCATTAAAATTGCCTCCGATATTGTCAAAAATCTGTGCTATG

L--S--V--C--A--L--K--L--P--P--I--L--S--K--I--C--A--M--

P8 GC = 43%

GCGGTCCCGGGAATAGATAAAGTTTGTGAATCTAAGACTTTATCTCCCCATAAC

A--V--P--G--I--D--K--V--C--E--S--K--T--L--S--P--H--N--

Q23 GC = 26%

ATAACTAAGTTCATTAGATCAGTGTTA

I--T--K--F--I--R--S--V--L--

#47 GC = 45%

GGCTATATACAACTCCGAAAAGGTCCAAATGTCGTAGGTCCATATGGCCTACTTCAACC

G--Y--I--Q--L--R--K--G--P--N--V--V--G--P--Y--G--L--L--Q--P-

Bestfit alignment (GCG) of the NADH dehydrogenase protein sequences of *Branchiostoma floridae* (top) and *Bos taurus* (bottom) to show overall similarity between the proteins. Genbank accession numbers AF035175 and J01394. '|' indicates identical and ':' and '.' indicates similar amino acids. Position of peptide 47 is shown in bold letters. The overall similarity is 72.4% and identity is 63.6%.

```
             .         .         .         .         .
  5 VIHIFLYFVPVLLAVAFLTLTERKVIGYVQLRKGPNVVGPYGLLQPIADG 54
    .|.| :  :|:||||||||| ||||:||.||||||||||||||||||||
  3 MINILMLIIPILLAVAFLTLVERKVLGYMQLRKGPNVVGPYGLLQPIADA 52

             .         .         .         .         .
 55 VKLFIKEPIKPSSSNPSVFFFAPMLALILALLLWMPVPLMDAFIELNFAV 104
    :|||||||::|..|. |.|  ||.:|| ||| :|.|.|:    | :|  |
 53 IKLFIKEPLRPATSSASMFILAPIMALGLALTMWIPLPMPYPLINMNLGV 102

             .         .         .         .         .
105 LFVLAISSLSVYSIMASGWSSNSKYALLGALRAVAQMVSYEVSLGLIILS 154
    ||.||.|||.||||: |||.|||||||:|||||||| :||||.| :|:||
103 LFMLAMSSLAVYSILWSGWASNSKYALIGALRAVAQTISYEVTLAIILLS 152

             .         .         .         .         .
155 LICLVGGFNLAQFFSAQEEVMLMLSCWPLGIMWFISTVAETNRSPFDLTE 204
    .: : | | |.   . ||:. |.|  ||| .||||||.|||||.||||||
153 VLLMSGSFTLSTLITTQEQMWLILPAWPLAMMWFISTLAETNRAPFDLTE 202

             .         .         .         .         .
205 GESELVSGFNVEYSGGPFALFFLAEYANILFMNVLSALLFL.AAH..... 248
    |||||||||||||. |||||||:||||||: ||: .|:|||  .|
203 GESELVSGFNVEYAAGPFALFFMAEYANIIMMNIFTAILFLGTSHNPHMP 252

             .         .         .         .         .
249 .FSLLGVAVKVGLLAGLYLWFRASYPRFRYDQLMHLAWKSFLPLSLGLLM 297
        :   :|  ||   :|| ||||||||||||||| ||.||||.| | |
253 ELYTINFTIKSLLLTMSFLWIRASYPRFRYDQLMHLLWKNFLPLTLALCM 302

             .
298 LNFSLPLTFSGI 309
     . |||:  |||
303 WHVSLPILTSGI 314
```

Amino acid sequence alignment of protein encoded by F12orf and the fumarases of *Rickettsia conorii* (*R. con*) and *Rickettsia prowazekii* (*R. prow).* Genbank accession numbers AE008652 and AJ235272. The consensus is in the bottom row. The F12 peptide sequence is in bold.

```
F12orf     SDSLGEVNVPACHYWGAQTQRSIDNFKIGSEKMPKPLIRAMGIVKLAAARVNMKNGDINE
R. con     SDSFGEIQIEEKFYWGAQTQRSLENFKIGKQKMPEILIRALAILKKCAAQVNHEFGDLEA
R. prow    SDSFGEIQIEEKFYWGAQTQRSLNNFKISKQKMPKILIRALAILKKCAAQVNYEFGDLEY
consensus  SDS GE+ +    YWGAQTQRS++NFKI  +KMP+ LIRA+ I+K  AA+VN + GD+

F12orf     VIGNAICNAAAEVIDGKLDNEFPLVVWQTGSGTQTNMDVNEGIANRAIEILGGEKGSKVP
R. con     KIAISIDKATDRILEGEFEDNFPLVVWQTGSGTQTNMNMNEVIASIANEELTGKKGGKSP
R. prow    KIATSIDKAIDRILAGEFEDNFPLVVWQTGSGTQTNMNMNEVIASIANEELTGKKGGKFP
            I  +I  A   ++ G+ ++ FPLVVWQTGSGTQTNM++NE IA+ A E L G+KG K P

F12orf     VHPNDHVNYSQSSNDTFPTAMHIATVSETENYLLPSLKNLYDALHSKSIAFQNIVKVGRT
R. con     VHPNDHVNKGQSSNDSFPTAMHIATVLATKQQLIPALNNILTSLQDKSKDWDKIIKIGRT
R. prow    VHPNDHVNKGQSSNDSFPTAMHIATVLATKQQLIPALNNLLTYLQDKSKDWDKIIKIGRT
           VHPNDHVN  QSSND+FPTAMHIATV  T+  L+P+L N+   L  KS   +  I+K+GRT

F12orf     HLQDATPLTLGQEFSGYAYQILQGIGRIKSALSNLLELAQGGTAVGTGINSRKQFDVHIA
R. con     HLQDATPLTLKQEFSGYITQIEYALERIENALQKVYLLAQGGTAVGTGINSKIGFDIKFA
R. prow    HLQDATPLTLKQEFSGYITQIEYALERIEDALKKVYLLAQGGTAVGTGINSKIGFDIKFA
           HLQDATPLTL QEFSGY  QI   + RI+ AL  +  LAQGGTAVGTGINS+  FD+  A

F12orf     NEIKKITGFNFVSSVNKFEALATHDALVEFSGALNVLAVSLMKIANDIRLLSSGPRCGIG
R. con     EKVAEFTKQPFKTAPNKFESLAAHDALVEFSGTLNTIAVSLMKIANDIRLLGSGPRCGLG
R. prow    QKVAEFTQQPFKTAPNKFESLAAHDALVEFSGTLNTIAVSLMKIANDIRLLGSGPRCGLG
            ++ + T   F ++ NKFE+LA HDALVEFSG LN +AVSLMKIANDIRLL SGPRCG+G

F12orf     EIILPANEPG
R. con     ELHLPENEPG
R. prow    ELHLPENEPG
           E+ LP NEPG
```

DNA sequence of insert in λF with the translation of the potential ORF (F12orf). The start codon for F12orf (ATG, bold) is 196 bases from the β-galactosidase sequences of the λZAPII vector.

```
vector-TTATTGTAGTATAGAAATTGCAGAATATATAGTATGAAATTAAAATTATAGGTGAGAATA
1        Y  C  S  I  E  I  A  E  Y  I  V  *  N  *  N  Y  R  *  E  *
61     GGGTATAGATGATTATTTTATATTTATGGTATACCTATTTTTTTAATCTGTTTATTCAGA
21       G  I  D  D  Y  F  I  F  M  V  Y  L  F  F  *  S  V  Y  S  D
121    TCAGCTATATAAGTAGACATGTTTGAGAAACTGAATTATAAACTTTTATGAATAGGAACT
41       Q  L  Y  K  *  T  C  L  R  N  *  I  I  N  F  Y  E  *  E  L
181    AAGATTAAGTAAAGTTATGAGAAAAGAAAAAGATAGTTTAGGAGAGGTGAATGTACCAGC
61       R  L  S  K  V  M  R  K  E  K  D  S  L  G  E  V  N  V  P  A
241    ATGTCACTATTGGGGAGCACAAACACAACGTTCTATTGATAATTTTAAAATTGGCTCAGA
81       C  H  Y  W  G  A  Q  T  Q  R  S  I  D  N  F  K  I  G  S  E
301    GAAAATGCCTAAACCTTTGATAAGAGCTATGGGGATTGTAAAATTAGCTGCAGCACGTGT
101      K  M  P  K  P  L  I  R  A  M  G  I  V  K  L  A  A  A  R  V
361    TAATATGAAAAATGGTGATATAAATGAGGTGATAGGTAATGCAATTTGTAATGCTGCTGC
121      N  M  K  N  G  D  I  N  E  V  I  G  N  A  I  C  N  A  A  A
421    TGAGGTTATAGATGGTAAGTTAGATAATGAATTTCCCCTGGTAGTATGGCAGACTGGTTC
141      E  V  I  D  G  K  L  D  N  E  F  P  L  V  V  W  Q  T  G  S
481    TGGTACTCAAACTAATATGGATGTTAATGAGGGTATAGCAAACCGTGCGATTGAAATTTT
161      G  T  Q  T  N  M  D  V  N  E  G  I  A  N  R  A  I  E  I  L
541    AGGTGGTGAAAAGGGTAGTAAAGTTCCAGTTCATCCAAATGATCATGTGAATTATTCTCA
181      G  G  E  K  G  S  K  V  P  V  H  P  N  D  H  V  N  Y  S  Q
601    GTCTTCAAATGATACTTTTCCGACAGCAATGCATATTGCTACTGTTAGTGAAACTGAAAA
201      S  S  N  D  T  F  P  T  A  M  H  I  A  T  V  S  E  T  E  N
661    CTATTTATTACCGAGTCTTAAAAATTTATATGATGCATTACATAGTAAATCTATAGCATT
221      Y  L  L  P  S  L  K  N  L  Y  D  A  L  H  S  K  S  I  A  F
721    TCAAAATATTGTCAAAGTGGGACGTACTCATCTTCAAGATGCAACTCCGCTAACTTTAGG
241      Q  N  I  V  K  V  G  R  T  H  L  Q  D  A  T  P  L  T  L  G
781    GCAAGAGTTTTCTGGTTATGCATATCAAATTTTACAAGGCATTGGCAGAATCAAATCAGC
261      Q  E  F  S  G  Y  A  Y  Q  I  L  Q  G  I  G  R  I  K  S  A
841    ATTAAGTAATTTGCTTGAGTTAGCACAGGGTGGTACTGCAGTAGGTACTGGTATTAATTC
281      L  S  N  L  L  E  L  A  Q  G  G  T  A  V  G  T  G  I  N  S
901    TAGGAAGCAGTTTGATGTTCATATTGCCAATGAAATAAAAAAAATTACAGGTTTTAATTT
301      R  K  Q  F  D  V  H  I  A  N  E  I  K  K  I  T  G  F  N  F
961    TGTTTCTTCTGTAAATAAGTTTGAAGCTTTGGCTACACATGATGCACTAGTTGAGTTTAG
321      V  S  S  V  N  K  F  E  A  L  A  T  H  D  A  L  V  E  F  S
1021   TGGTGCATTAAATGTACTTGCTGTTAGTTTGATGAAAATTGCTAATGATATTAGATTACT
341      G  A  L  N  V  L  A  V  S  L  M  K  I  A  N  D  I  R  L  L
1081   GAGCTCTGGTCCAAGGTGTGGTATAGGGGAAATTATATTACCTGCAAATGAGCCAGGA-vector
361      S  S  G  P  R  C  G  I  G  E  I  I  L  P  A  N  E  P  G
```

Amino acid sequence alignment of the protein encoded by Q23orf and the alkaline proteinase secretion ATP-binding proteins of *Rickettsia prowazekii* (*R. prow)* and *Rickettsia conorii* (*R. con*). Genbank accession numbers NP 220698.1 and NP360066.1. The consensus is in the bottom row. The Q23 peptide sequence is in bold.

```
Q23orf    QVLDRVLTSGSVSTLVMLSAITIIAFACSAILEICRSLVMAKVGDWIDKVVTPDLIMXXX
R. prow   QVLDRVLGSGNLQTLLFLSIIIAYIYFVYGLLQIARSFTLIKVGEWLDKTVAPVIFASSI
R. con    QVLDRVLGSGNLQTLLFLSIIIAYIYFVYGLLQIARSFTLIKVGEWLDRTVAPVIFASSI
consensus QVLDRVL SG++ TL+ LS I    +    +L+I RS  + KVG+W+DK V P +


Q23orf    XXXXXXXXXXXGEVIRDLGVVKSFITGFGIFSLFDTPWAVLYLVTIFMIHSVTGYIAIVG
R. prow   SATATRVNIGSSQLLRDFQAVKTFLTSTGINTLFDAPWSIIYIAVIFSIHPYIGFITVFG
R. con    SAAATRANMGSSQLLRDFQAVKTFLTSTGINTLFDAPWSLIYIAVIFSIHPYIGLITVFG
consensus             +++RD   VK+F+T  GI +LFD PW+++Y+  IF IH   G+I + G


Q23orf    IVLLTAMGVWNELATKKILQEASEEGIRNINSIDVATRNAEVVEAMGMVNYIVDDWAQKN
R. prow   AIIIVSTAFFNAAATNKTLGEATEFSIKGMTQADIANRNAESIEAMGMMKNVTKNWHKFN
R. con    AIIIVSTAFFNAAATNKTLGEATEFSIKGMTQADIANRNAEAIEAMGMMKNVTKNWHKFN
consensus  +++ +   +N  AT K L EA+E  I+ +   D+A RNAE +EAMGM+  +  +W + N


Q23orf    DKNRAMQIKAQNRSNLISGITKFIRSVLQIAVIGIGAFLAVLGHK---TAGGIIASSILM
R. prow   ILALDKQSVASYRNGVISNFSRFVRNIMQMAVTGVGAYIVVKSHSADMTPGNMIMSSIIV
R. con    ILALDKQSVASYRNGVISNFSRFIRNIMQMAVTGVGAYIVVKSHSADMTPGNMIMSSIIV
consensus       Q  A  R+ +IS  ++FIR+++Q+AV G+GA++ V  H    T G +I SSI++


Q23orf    GRALAPFETSINTWKMLISARISYKRLQMLLVASPKREQTMSLPIPQGKVVFDH
R. prow   GRALAPFDNAIELWKSMSGAIKSYKNINNLFNNYSSRDEAMPIPNVNGHLTVEN
R. con    GRALAPFDNAIDLWKSMSGAIKSYKNINNLFNTYSSRDETMPIPNVDGHLTVEN
consensus GRALAPF+ +I+ WK +  A  SYK +  L      R++ M +P   G +  ++
```

REFERENCES:

Acuña, G., Ebeling, S. and Hennecke, H., 1991. Cloning, sequencing and mutational analysis of the *Bradyrhizobium japonicum fumC*-like gene: evidence for the existence of two different fumarases. *J. Gen. Microbiology*. 137, 991-1000.

Aebi, C., Maciver, I., Latimer, J. L., Cope, L. D., Stevens, M. K., Thomas, S. E., McCracken, G., H. and Hansen, E. J., 1997. A protective epitope of *Moraxella catarrhalis* is encoded by two different genes. *Infect. Immun*. 65, 4367-77.

Allsopp, M. T. E. P., Hattingh C. M., Vogel, S. W. and Allsopp, B. A., 1999. Evaluation of 16S, *map*1 and pCS20 probes for the detection of *Cowdria* and *Ehrlichia* species. *Epidemiol. Infect*. 122, 323-8.

Allsopp, M. T. E. P., Visser, E. S., du Plessis, J. L., Vogel, S. W. and Allsopp, B. A., 1997. Different organisms associated with heartwater as shown by analysis of 16S ribosomal RNA gene sequences. *Vet. Parasitol*. 71, 283-300.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J., 1990. Basic local alignment search tool. *J. Mol. Biol*. 215, 403-10.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zang, J., Zang, Z., Miller, W. and Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389-3402.

Ambrosio, R. E., du Plessis, J. L. and Bezuidenhout, J. D., 1987. The construction of genomic libraries of *Cowdria ruminantium* in an expression vector, λgt11. *Onderstepoort J Vet Res* 54: 255-256.

Anon., 1994. Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711.

Atassi, M. Z., 1975. Antigenic structure of myoglobin: the complete immunochemical anatomy of a protein and conclusions relating to antigenic structure. *Immunochem*. 12, 423-38.

Atassi, M. Z., 1984. Antigenic structures of proteins. *Eur. J. Biochem*. 145, 1-20.

Atassi, M. Z., Lee, C. L. and Pai, R. C., 1976. Enzymatic and immunochemical properties of lysozyme. XVI. A novel synthetic approach to an antigenic reactive site by direct linkage of the relevant conformationally adjacent residues constituting the site. *Biochem. Biophys. Acta.* 427, 745-51.

Atassi, M. Z. and Smith, J. A., 1978. A proposal for the nomenclature of antigenic sites in peptides and proteins. Immunochemistry. 15, 609-10.

Audu, P. A., Agbede, R. I. S. and Esievo, K. A. N., 1995. Studies of experimental infection and cross-immunity between stocks of *Cowdria ruminantium* in Sokoto red goats. *Vet. Parasitol.*. 59, 23-8.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J .A. and Struhl, K., 1998. Current protocols in Molecular Biology. John Wiley & Sons, Inc, Canada.

Bajari, T. M., Lindstedt, K. A., Riepl, M., Mirsky, V. M., Nimf, J., Wolfbeis, O. S., Dresel, H, A., Bautz, E. K. and Schneider, W. J., 1998. A minimal binding domain of the low density lipoprotein receptor family. *Biol. Chem.* 379, 1053-62.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. *Cur. Opin. Biotechnol.* 10, 411-21.

Barbet, A. F. and Allred, D. R., 1991. The *msp1β* multigene family of *Anaplasma marginale*: Nucleotide sequence analysis of an expressed copy. *Infect. Immun.* 59, 971-6.

Barbet, A. F., Semu, S. M., Chigagure, N., Kelly, P. J., Jongejan, F. and Mahan, S. M., 1994. Size variation of the major immunodominant protein of *Cowdria ruminantium*. *Clin. Diagn. Lab. Immunol.* 1, 744-6.

Barbet, A. F., Whitmire, W. M., Kamper, S. M., Simbi, B. H., Ganta, R. R., Moreland, A. L., Mwangi, D. M., McGuire, T. C. and Mahan, S. M., 2001. A subset of *Cowdria ruminantium* genes important for immune recognition and protection. *Gene*. 275, 287-98.

Barré, N., Uilenberg, G., Morel, P. C. and Camus, E., 1987. Danger of introducing heartwater onto the American mainland: Potential role of indigenous and exotic *Amblyomma* ticks. *Onderstepoort J. Vet. Res.* 54, 405-17.

Bartoli, F., Nuzzo, M., Urbanelli, L., Bellintani, F., Prezzi, C., Cortese, R. and Monaci, P., 1998. DNA-based selection and screening of peptide ligands. *Nature Biotechnol.* 16, 1068-73.

Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R., Hannum, C., Leach, S. J., Margoliash, E., Michael, J. G., Miller, A., Prager, E. M., Reichlin, M., Sercarz, E. E., Smith-Gill, S. J., Todd, P. A. and Wilson, A. C., 1984. The antigenic structure of proteins: a reappraisal. *Annu. Rev. Immunol.* 2, 67-101.

Bentley, L., Fehrsen, J., Jordaan, F., Huismans, H. and du Plessis, D. H., 2000. Identification of antigenic regions on VP2 of African horsesickness virus serotype 3 using phage displayed epitope libraries. *J. Gen. Virol.* 81, 993-1000.

Berzofsky, J. A., 1985. Intrinsic and extrinsic factors in protein structure. *Science*. 229, 932-40.

Bezuidenhout, J. D. and Bigalke, R. D., 1987. The control of heartwater by means of tick control. *Onderstepoort J. Vet. Res.* 54, 525-8.

Bezuidenhout, J. D., Paterson, C. and Barnard, B. J., 1985. *In vitro* cultivation of *Cowdria ruminantium. Onderstepoort J. Vet. Res.* 52, 113-20.

Blüthner, M., Bautz, E. K. F. and Bautz, F., 1996. Mapping of epitopes recognised by PM/Scl autoantibodies with gene-fragment phage display libraries. *J. Imm. Methods*. 198, 187-98.

Blüthner, M., Schäfer, C. Schneider, C. and Bautz, F. A., 1999. Identification of major linear epitopes on the Nuclear PBC autoantigen by gene-fragment phage-display technology. *Autoimmunity*. 29, 33-42.

Bona, C. A., 1996. Internal image concept revisited. *Proc. Soc. Exp. Biol. Med.* 213, 32-42.

Bonnycastle, L. L. C., Mehroke, J. S., Rashed, M., Gong, X. and Scott, J. K., 1996. Probing the basis of antibody reactivity with a panel of constrained peptide libraries displayed by filamentous phage. *J. Mol. Biol.* 258, 747-62.

Bowie, M. V., Reddy, G. R., Semu, S. M., Mahan, S. M. and Barbet, A. F., 1999. Potential value of major antigenic protein 2 for serological diagnosis of heartwater and related ehrlichial infections. *Clin. Diagn. Lab. Immunol.* 6, 209-15.

Brayton, K. A., de Villiers, E. P., Fehrsen, J., Nxomani, C., Collins, N. E. and Allsopp, B. A.,1999. *Cowdria ruminantium* DNA is unstable in a SuperCos1 library. *Onderstepoort J. Vet. Research.* 66, 111-7.

Brayton, K. A., Fehrsen, J., de Villiers, E. P., van Kleef, M. and Allsopp, B. A., 1997. Construction and initial analysis of a representative λZAPII expression library of the intracellular rickettsia *Cowdria ruminantium*: cloning of *map1* and three other *Cowdria* genes. *Vet. Parasitol*. 72, 185-99.

Brayton, K. A., van der Walt, M., Vogel, S. W. and Allsopp, B. A., 1998. A partially protective clone from *C. ruminantium* identified by using a *Salmonella* vaccine delivery system. *Ann. New York Acad. Sci.* 849, 247-52.

Burton, D. R., 1995. Phage display. *Immunotechnology.* 1, 87-94.

Byrom, B., Mahan, S. M. and Barbet, A. F., 1993. The development of antibody to *C. ruminantium* in mice and its role in heartwater disease. *Rev. Élev. Méd. vét. Pays trop.* 46, 197-201.

Cahill, D. J., 2001. Protein and antibody arrays and their medical applications. *J. Immunol. Methods.* 250, 81-91.

Capra, J. D. and Kehoe, J. M., 1975. Hypervariable regions, idiotypy and the antibody-combining site. *Adv. Immunol.* 20, 1-40.

Chung, C. T., Niemela, S. L. and Miller, R. H., 1989. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA.* 86, 2172-5.

Church, G. M. and Gilbert, W., 1984. Genomic sequencing. *Proc. Nat. Acad. Sci. USA.* 81, 991-1995.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D. and Winter, G., 1991. Making antibody fragments using phage display libraries. *Nature*. 352, 624-8.

Clark, M. F. and Adams, A. N., 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol*., 34, 475-483.

Collins, N. C., de Villiers, E. P., Brayton, and Allsopp, B. A., 1998. DNA sequence of a cosmid clone of *Cowdria ruminantium*. *Ann. New York Acad. Sci.* 849, 365-8.

Cortese, I., Tafi, R., Grimaldi, L. E., Martino, G and Nicosia, A., 1996. Identification of peptides specific for cerebrospinal fluid antibodies in multiple sclerosis by using phage libraries. *Proc. Nat. Acad. Sci. USA.* 93, 11063-7.

Cowdry, E. V., 1925. Studies on the aetiology of heartwater. II *Rickettsia ruminantium* (n. sp.) in the tissues of infected animals. *J. Exp. Med.* 42, 231-52.

Craig, L., Sanschagrin, P. C., Rozek, A., Lackie, S., Kuhn, L. A. and Scott, J. K., 1998. The role of structure in antibody cross-reactivity between peptides and folded proteins. *J. Mol. Biol.* 281, 183-201.

Crameri, R., Jaussi, R., Menz, G. and Blaser, K., 1994. Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. *Eur. J. Biochem.* 226, 53-8.

Crameri, R. and Suter, M., 1993. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for the selection of functional gene products linked to the genetic information responsible for their production. *Gene*. 137, 69-75.

Cwirla, S. E., Peters, E. A., Barrett, R. W. and Dower, W. J., 1990. Peptides on phage: A vast library of peptides for identifying ligands. *Proc. Nat. Acad. Sci. USA.* 87, 6378-82.

Dale E. M., Binns, A. N. and Ward, J. E., 1993. Construction and characterisation of Tn5virB, a transposon that generates nonpolar mutations, and its use to define virB as an essential virulence gene in *Agrobacterium tumefaciens*. *J. Bacteriol*. 175, 887-91.

Dame, J. B., Mahan, S. M. and Yowell, C. A., 1992. Phylogenetic relationship of *Cowdria ruminantium*, agent of heartwater, to *Anaplasma marginale* and other members of the order *Rickettsiales* determined on the basis of 16S rRNA sequence. *Int. J. Syst. Bacteriol.* 42, 270-4.

Davenport, M. P., Quinn, C. L., Valsasnini, P., Sinigaglia, F. and Hill, A. V. S., 1996. Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. *Immunol.* 88, 482-6.

Davies, E. L., Smith, J. S., Birkett, C. R., Manser, J. M., Anderson-Dear, D. V. and Young, J. R., 1995. Selection of specific phage-displayed antibodies using libraries derived from chicken immunoglobin genes. *J. Immunol. Methods*. 186, 125-35.

Day, E. D., 1990. Advanced Immunochemistry, Second Edition. John Wiley and Sons, New York.

De Berardinis, P., Sartorius, R., Fanutti, C., Perham, R. N., Del Pozzo, G. and Guardiola, J., 2000. Phage display of peptide epitopes from HIV-1 elicits strong cytolytic responses. *Nat. Biotech.* 18, 873-6.

Deem, S. L., Donachie, P. L. and Norval, R. A. I., 1996. Colostrum from dams living in a heartwater-endemic area influences calfhood immunity to *Cowdria ruminantium*. *Vet. Parasitol.* 61, 133-44.

De Greef, A., van Alphen, L. and Smith, H. E., 2000. Selection of recombinant antibodies specific for pathogenic *Streptococcus suis* by subtractive phage display. *Infect. Immun.* 68, 3949-55.

De la Cruz, V., Lal, A. A. and McCutchan, T. F., 1988. Immunogenicity and epitope mapping of foreign sequences via genetically engineered filamentous phage. *J. Biol. Chem.* 263, 4318-22.

De Villiers, E. P., 2001. Genome Analysis of *Cowdria ruminantium*. PhD thesis, Utrecht University.

De Villiers, E. P., Brayton, K. A., Zweygarth, E. and Allsopp, B. A., 1998. Purification of *Cowdria ruminantium* organisms for use in genome analysis by pulsed-field gel electrophoresis. *Ann. New York Acad. Sci*. 849, 313-20.

De Villiers, E. P., Brayton, K. A., Zweygarth, E. and Allsopp, B. A., 2000a. Macrorestriction fragment profiles reveal genetic variation of *Cowdria ruminantium* isolates. *J. Clin. Micro.* 38, 1967-70.

De Villiers, E. P., Brayton, K. A., Zweygarth, E. and Allsopp, B. A., 2000b. Genome size and genetic map of *Cowdria ruminantium. Microbiology*. 146, 2627-34.

Devlin, J. J., Panganiban, L. C. and Devlin, P. E., 1990. Random peptide libraries: a source of specific protein binding molecules. *Science*. 249, 404-6.

Dübel, S. web page, http://www.mgen.uni-heidelberg.de/SD/SDscFvSite.html

Dumler, J. S., Barbet, A. F., Bekker, C. P., Dasch, G. A., Palmer, G. H., Ray, S. C., Rikihisa, Y. and Rurangirwa, F. R., 2001. Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales:* unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia,* descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int. J. Syst. Evol. Microbiol*. 51, 2145-65.

Du Plessis, D. H. and Jordaan, F., 1996. Phage libraries displaying random peptides derived from a target sequence. In *Phage Display of Peptides and Proteins. A Laboratory Manual.* Edited by Kay, B. K., Winter, J. and McCafferty, J., Academic Press, San Diego.

Du Plessis, D. H., Romito, M. and Jordaan, F. A., 1995. Identification of an antigenic peptide specific bluetongue virus using phage display expression of NS1 sequences. *Immunotechnology.* 1, 221-30.

Du Plessis, D. H., Wang, L-F., Jordaan, F. A. & Eaton, B. T., 1994. Fine mapping of a continuous epitope of bluetongue virus using overlapping synthetic peptides and a random epitope library. *Virology.* 198, 346-9.

Du Plessis, J. L., 1970. Immunity to heartwater: 1. A preliminary note on the role of serum antibodies. *Onderstepoort J. Vet. Res.* 3, 147-50.

Du Plessis, J. L., 1985. The natural resistance of cattle to artificial infection with *Cowdria ruminantium*: the role played by conglutinin. *Onderstepoort J. Vet. Res.* 52, 273-7.

Du Plessis, J. L., 1993. An *in vitro* test to demonstrate the inhibitory effect of homologous serum on the infectivity of *Cowdria ruminantium*. *Onderstepoort J. Vet. Res.* 60, 69-73.

Du Plessis, J. L., Berche, P. and van Gas, L., 1991. T cell-mediated immunity to *Cowdria ruminantium* in mice: the protective role of Lyt-2^+ T cells. *Onderstepoort J. Vet. Res.* 58, 171-9.

Du Plessis, J. L. and Kümm, N. A. L., 1971. The passage of *Cowdria ruminantium* in mice. *J. South Afr. Vet. Med. Ass.* 42, 217-21.

Du Plessis, J. L. and Malan, L., 1987. The non-specific resistance of cattle to heartwater. *Onderstepoort J. Vet. Res.* 54, 333-6.

Du Plessis, J. L., Malan, L. and Kowalsky, Z. E., 1987. The pathogenesis of heartwater. *Onderstepoort J. Vet. Res.* 54, 313-8.

Du Plessis, J. L., van Gas, L., Olivier, J. A. and Bezuidenhout, J. D., 1989. The heterogenicity of *Cowdria ruminantium* stocks: cross-immunity and serology in sheep and pathogenicity to mice. *Onderstepoort J. Vet. Res.* 56, 195-201.

Edman, P. and Begg, G., 1967. A protein sequenator. *Eur. J. Biochem*. 1, 80-91.

Fack, F., Hügle-Dörr, B., Song, D, Queitsch, I., Petersen, G. and Bautz, E. K. F., 1997. Epitope mapping by phage display:

random versus gene-fragment libraries. *J. Immunol. Methods*. 206, 43-52.

Fieser, T. M., Tainer, J. A., Geysen, H. M., Houghten, R. A. and Lerner, R. A., 1987. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. Proc. Natl. Acad. Sci. USA. 84, 8568-72.

Folgori, A., Tafi, R., Meola, A., Felici, F., Galfre, G., Cortese, R., Monaci, P. and Nicosia, A., 1994. A general strategy to identify mimotopes of pathological antigens using random peptide libraries and human sera. *EMBO.* 13, 2236-43.

Freifelder, D., 1983. Molecular Biology. A comprehensive introduction to prokaryotes and eukaryotes. Jones and Bartlett Publishers, Inc., Boston.

Geysen, H. M., Meloen, R. H. and Barteling, S. J., 1984. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Nat. Acad. Sci. USA*. 81, 3998-4002.

Geysen, H. M., Rodda, S. J. and Mason, T. J., 1986. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* 23, 709-15.

Geysen, H. M., Rodda, S. J., Mason, T. J., Tribbick, G. and Schoofs, P. G., 1987. Strategies for epitope analysis using peptide synthesis. *J. Immunol. Methods.* 102, 259-74.

Gilmore, R. D., Cieplak, W., Policastro, P. F. and Hackstadt, T., 1991. The 120 kilodalton outer membrane protein (rOMP B) of *Rickettsia rickettsii* is encoded by an unusually long open reading frame: evidence for protein processing from a large precursor. *Mol. Microbiol*. 5, 2361-70.

Greenwood, J., Willis, A. E. and Perham, R. N., 1991. Multiple display of foreign peptides on a filamentous bacteriophage. Peptides from *Plasmodium falciparum* circumsporozoite protein as antigens. *J. Mol. Biol*. 220, 821-7.

Grothaus, M. C., Srivastava, N., Smithson, S. L., Kieber-Emmons, T., Williams, D. B., Carlone, G. M. and Westerink, M. A. J., 2000. Selection of an immunogenic mimic of the capsular polysaccharide of *Neiseria meningitidis* serogroup A using a peptide display library. *Vaccine.* 18, 1253-63.

Guichard, G., Benkirane, N., Zeder-Lutz, G., van Regenmortel, M. H. V., Briand, J-P, and Muller, S., 1994. Antigenic mimicry of natural L-peptides with retro-inverso-peptidomimetics. *Proc. Nat. Acad. Sci. USA*, 91, 9765-9.

Hackstadt, T., Messer, R., Cieplak, W. and Peacock, M. G., 1992. Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of Rickettsiae: Identification of an avirulent mutant deficient in processing. *Infect. Immun*. 60, 159-65.

Harlow, E. and Lane, D., 1988. Antibodies. A Laboratory Manual. Cold Spring Harbour. USA

Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. and Hughes, S. H., 1983. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. *Proc. Nat. Acad. Sci*. USA. 80, 31-5.

Holt, L. J., Bussow, K., Walter, G. and Tomlinson, I. M., 2000. By-passing selection: direct screening for antibody-antigen interactions using protein arrays. *Nucleic Acids Res*. 28, E72.

Hoogenboom, H. R. and Winter, G., 1992. By-passing immunisation. Human antibodies from synthetic repertoires of germline V_H gene segments rearranged *in vitro. J. Mol. Biol*. 227, 381-8.

Hopp, T.P. and Woods, K.R.,1983. A computer program for predicting protein antigenic determinants. *Mol. Immunol*. 20, 483-9.

Howell, C. J., de Vos, A. J., Bezuidenhout, J. D., Potgieter, F. T. and Barrowman, P. R., 1981. The role of chemical tick eradication in the control or prevention of tick-transmitted diseases of cattle. In: G. B. Whitehead and J. D. Gibson (Ed), Tick Biology and Control. Tick Research Unit, Rhodes University, Grahamstown, South Africa, 61-6.

Janeway, C. A., Travers, P., Walport, M. and Shlomchik, M., 2001. Immunobiology, 5th ed., Garland Publishing.

Jacobsson, K. and Frykberg, L., 1995. Cloning of ligand-binding domains of bacterial receptors by phage display. *Biotechniques*.18, 878-85.

Jacobsson, K. and Frykberg, L., 1996. Phage display shot-gun cloning of ligand-binding domains of prokaryotic receptors approaches 100% correct clones. *Biotechniques*. 20, 1070-81.

Jefferies, D., 1998. Selection of novel ligands from phage display libraries: an alternative approach to drug and vaccine discovery? *Parasitol. Today*. 14, 202-6.

Jerne, N. K., 1960. Immunological speculations. *Ann. Rev. Microbiol.* 14, 341-58.

Jespers, L. S., Messens, J. H., de Keyser, A., Eeckhout, D., van den Brande, I., Gansemans, Y. G., Lauwereys, M. J., Vlasuk, G. P. and Stanssens, P. E., 1995. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnol.* 13, 378-82.

Jongejan, F., 1991. Protective immunity to heartwater (*Cowdria ruminantium*) is acquired after vaccination with *in vitro*-attenuated rickettsiae. *Infect. Immun.* 59, 729-31.

Jongejan, F. and Thielemans, M. J., 1989. Identification of an immunodominant antigenically conserved 32-kilodalton protein from *Cowdria ruminantium*. *Infect. Immun.* 57, 3243-6.

Jongejan, F., Thielemans, M. J., de Groot, M., van Kooten, P. J., and van de Zeijst, B. A., 1991. Competitive enzyme linked immunosorbent assay for heartwater using monoclonal antibodies to a *Cowdria ruminantium* specific 32 kilodalton protein. *Vet. Microbiol.* 28, 199-211.

Jongejan, F., Uilenberg, G., Franssen, F. F. J., Gueye, A. and Nieuwenhuijs, J., 1988. Antigenic differences between stocks of *Cowdria ruminantium*. *Res. Vet. Sci*. 44, 186-9.

Karplus, M. and McCammon, J. A., 1983. Dynamics of proteins: elements and function. *Annu. Rev. Biochem*. 52, 263-300.

Kay, B. K., Adey, N. B., He, Y-S., Manfredi, J. P., Mataragnon, A. H. and Fowlkes, D. M., 1993. M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene*. 128, 59-65.

Klimka, A., Matthey, B., Roovers, R. C., Barth, S., Arends, J. W., Engert, A. and Hoogenboom, H. R., 2000. Human anti-CD30 recombinant antibodies by guided phage antibody selection using cell panning. *Br. J. Cancer.* 83, 252-60.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A. and Virnekas, B., 2000. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol*. 296, 57-86.

Kocan, K. M. and Bezuidenhout, J. D., 1987. Morphology and development of *Cowdria ruminantium* in *Amblyomma* ticks. *Onderstepoort J. Vet. Res.* 54, 177-82.

Kock, N. D., van Vliet, A. H. M., Charlton, K. and Jongejan, F., 1995. Detection of *Cowdria ruminantium* in the blood and bone marrow samples from clinically normal, free-ranging Zimbabwean wild ungulates. *J. Clin. Micro*. 33, 2501-4.

Kola, A., Baensch, M., Bautsch, W., Henneke, M, Klos, A., Casaretto, M. and Köhl, J., 1996. Epitope mapping of a C5a neutralizing mAb using a combined approach of phage display, synthetic peptides and site-directed mutagenesis. *Immunotechnology*. 2, 115-26.

Kolaskar, A. S. and Tongaonkar, P. C., 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS*. 276, 172-4.

Kouzmitcheva, G. A., Petrenko, V. A. and Smith, G. P., 2001. Identifying diagnostic peptides for Lyme disease through epitope discovery. *Clin. Diagn. Lab. Immunol.* 8, 150-60.

Kumar, R. B. and Das, A., 2001. Functional analysis of the *Agrobacterium tumefaciens* T-DNA transport pore protein VirB8. *J. Bacteriol*. 183, 3636-41.

Kuwabara, I., Maruyama, H., Kamisue, S., Shima, M., Yoshioka, A and Maruyama, I. N., 1999. Mapping of the minimal domain encoding a conformational epitope by λ phage surface display: factor VIII inhibitor antibodies from haemophilia A patients. *J. Immunol. Methods.* 224, 89-99.

Laemmli, U. K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227, 680-5.

Lally, N. C., Nicoll, S., Paxton, E. A., Cary, C. M. and Sumption, K. J., 1995. The *Cowdria ruminantium groE* operon. *Microbiology.* 141, 2091-100.

Landsteiner, K., 1962. The specificity of serological reactions. Dower Publications. New York. 330.

Li, H., Shen., D. T., Davis, W. C., Knowles, D. P., Gorham, J. R. and Crawford, T. B., 1995. Identification and characterisation of the major proteins of malignant catarrhal fever virus. *J. Gen. Virol*. 76, 123-9.

Light, J., Maki, R. and Asssa-Munt, N., 1996. Expression cloning of cDNA by phage display selection. *Nucleic Acids Res*. 24, 4367-8.

Liu, B. and Marks, J. D., 2000. Applying phage antibodies to proteomics: selecting single chain Fv antibodies to antigens blotted on nitrocellulose. *Anal. Biochem*. 286, 119-28.

Logan, L. L., 1987. *Cowdria ruminantium*: Stability and preservation of the organism. *Onderstepoort. J. Vet. Res.* 54, 187-91.

Louw, E., Brayton, K. A., Collins, N. E., Pretorius, A., van Strijp, F. and Allsopp, B. A., 2002. Sequencing of a 15kb *Ehrlichia ruminantium* clone and evaluation of the *cpg*1 open reading frame for protection against heartwater. *Ann N Y Acad Sci* 969, 147-50.

Lundin, K., Samuelsson, A., Jansson, M., Hinkula, J., Wahren, B., Wigzell, H. and Persson, M. A. A., 1996. Peptides isolated from random peptide libraries on phage elicit a neutralising anti-HIV-1 response: analysis of immunological mimicry. *Immunol.* 89, 579-86.

Lunt, R. A., White, J. R. and Blacksell, S. D., 1988. Evaluation of a monoclonal antibody blocking ELISA for the detection of group-specific antibodies to bluetongue virus in experimental and field sera. *J. Gen. Virol.* 69, 2729-40.

Luzzago, A., Felici, F., Tramontano, A., Pessi, A. and Cortese, R., 1993. Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage display library of constrained peptides. *Gene*. 128, 51-7.

Mahan, S. M., Andrew, H. R. and Tebele, N., 1995. Immunization of sheep against heartwater with inactivated *Cowdria ruminantium*. *Res. Vet. Sci.* 58, 46-9.

Mahan, S. M., Kumbula, D., Burridge, M. J. and Barbet, A. F., 1998a. The inactivated *C. ruminantium* vaccine for heartwater protects against heterologous strains and against laboratory and field challenge. *Vaccine*. 16, 1203-11.

Mahan, S. M., Mcguire, T. C., Semu, S. M., Bowie, M. V., Jongejan, F., Rurangirwa, F. R. and Barbet, A. F., 1994. Molecular cloning of a gene encoding the immunogenic 21 kDa protein of *Cowdria ruminantium. Microbiology*. 140, 2135-42.

Mahan, S. M., Semu, S. M., Peter, T. F. and Jongejan, F., 1998b. Evaluation of the MAP-1B ELISA for Cowdriosis with field sera from livestock in Zimbabwe. *Ann. New York Acad. Sci.* 849, 259-61.

Mahan, S. M., Tebele, N., Mukwedeya, D., Semu, S., Nyathi, C. B., Wassink, L. A., Kelly, P. J., Peter, T. and Barbet, A. F., 1993. An immunoblotting diagnostic assay for heartwater based on the immunodominant 32-kilodalton protein of *Cowdria ruminantium* detects false positives in field sera. *J. Clin. Microbiol*. 31, 2729-37.

Mäkelä, O., 1965. Single lymph node cells producing heteroclitic bacteriophage antibody. *J. Immunol.* 95, 378-86.

Malik, P. and Perham, R. N., 1996. New vectors for peptide display on the surface of filamentous bacteriophage. *Gene*. 17,

49-51.

Malik, P. and Perham, R. N., 1997. Simultaneous display of different peptides on the surface of filamentous bacteriophage. *Nucleic Acids Res*. 25, 915-6.

Malik, P., Terry, T. D., Gowda, L. R., Langara, A., Petukhov, S. A., Symmons, M. F., Welsh, L. C., Marvin, D. A. and Perham, R. N., 1996. Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol*. 260, 9-21.

Markland, W., Roberts, B. L., Saxena, M. J., Guterman, S. K. and Ladner, R. C., 1991. Design, construction and function of multicopy display vector using fusions to the major coat protein of bacteriophage M13. *Gene*. 109, 13-9.

Marks J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. and Winter, G., 1991. By-passing immunisation: Human antibodies from V-gene libraries displayed on phage. *J. Mol biol.* 222, 581-97.

Martinez, D., Maillard, J. C., Coisne, S., Sheikboudou, C. and Bensaid, A., 1994. Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium. Vet. Immunol. Immunopathol.* 41, 152-63.

Martinez, D., Perez, J. M., Sheikboudou, C., Debus, A. and Bensaid, A., 1996. Comparative efficacy of Freund's and Montanide ISA50 adjuvants for the immunisation of goats against heartwater with inactivated *Cowdria ruminantium*. *Vet. Parasitol.* 67, 175-84.

Maruyama, I. N., Maruyama, H. I. and Brenner, S., 1994. Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc. Nat. Acad. Sci. USA*. 91, 8273-7.

Mathe, C., Peresetsky, A., Dehais, P., van Montagu, M. and Rouze, P., 1999. Classification of *Arabidosis thaliana* gene sequences: clustering of coding sequences into two groups according to codon usage improves gene prediction. *J. Mol. Biol.* 285, 1977 -1991.

Matthews, L. J., Davis, R. and Smith, G. P., 2002. Immunogenically fit subunit vaccine components via epitope discovery from natural peptide libraries. *J. Immunol*. 169, 837-46.

Mazzarelli, J. M. and Ricciardi, R. P., 2001. The RITE Assay: Identifying effectors that target the transcriptional machinery using phage display technology. *BioTechniques*. 30, 380-6.

Menendez, A., Bonnycastle, L. C., Pan, O. C. C. and Scott, J. K., 2001. Screening peptide libraries. In *Phage Display. A Laboratory Manual*. Edited by Barbas, C. F., Burton, D. R., Scott, J. K. and Silverman, G. J. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Meola, M., Delmastro, P., Monaci, P., Luzzago, A., Nicosia, A., Felici, F., Cortese, R. and Galfrè, G., 1995. Derivation of vaccines from mimotopes. Immunologic properties of Human Hepatitis B Virus surface antigen mimotopes displayed on filamentous phage. *J. Immunol*. 154, 3162-72.

Mondry, R., Martinez, D., Camus, E., Liebisch, A., Katz, J. B., Dewald, R., van Vliet, A. H. M. and Jongejan, F., 1998. Validation and comparison of three enzyme-linked immunosorbent assays for the detection of antibodies to *Cowdria ruminantium* infection. *Ann. New York Acad. Sci.* 849, 262-72.

Moshkovski, S.D., 1947. Comments by readers. *Science.* 106, 62.

Mutunga, M., Preston, P. M. and Sumption, K. J., 1998. Nitric oxide is produced by *Cowdria ruminantium*-infected bovine pulmonary endothelial cells in vitro and is stimulated by gamma interferon. *Infect. Immun*. 66, 2115-21.

Mwangi, D. M., Mahan, S. M., Nyanjui, J. K., Taracha, E. L. N. and McKeever, D. J., 1998a. Immunization of cattle by infection with *Cowdria ruminantium* elicits T lymphocytes that recognize autologous, infected endothelial cells and monocytes. *Infect. Immun*. 66, 1855-60.

Mwangi, D. M., McKeever, D. J., Nyanjui, J. K., Barbet, A. F. and Mahan, S. M., 1998b. Major antigenic proteins 1 and 2 of *Cowdria ruminantium* are targets for T-lymphocyte responses of immune cattle. *Ann. New York Acad. Sci*. 849, 372-4.

Nagesha, H. S., Yu, M. and Wang, L-F., 1996. Application of linker-ligation-PCR for the construction of phage display epitope libraries. *J. Virol. Methods*. 60, 147-54.

Neitz, A. W. H., Vermeulen, N. M. J and Viljoen, G. J., 1987. Purification of *Cowdria ruminantium* by density gradient centrifugation. *Onderstepoort J. Vet. Res.* 54, 223-31.

Neitz, W. O. and Alexander, R. A., 1945. Immunisation of cattle against heartwater and the control of tick-borne diseases, redwater, gallsickness and heartwater. *Trop. Anim. Health Prod*. 16, 56-62.

Nissim, A., Hoogenboom, H. R., Tomlinson, I. M., Flynn, G., Midgley, C., Lane, D. and Winter, G., 1994. Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO*. 13, 692-8.

Novotny, J., 1987. A static accessibility model of protein antigenicity. *Int. Rev. Immunol*. 2, 379-89.

Nyika, A., Mahan, S. M., Burridge, M. J., Mcguire, T. C., Rurangirwa, F. and Barbet, A. F., 1998. A DNA vaccine protects mice against the rickettsial agent *Cowdria ruminantium. Parasite Immunol*. 20, 111-9.

Oberem, P. T. and Bezuidenhout, J. D., 1987. The production of heartwater vaccine. *Onderstepoort J. Vet. Res.* 54, 485-8.

Oberle, S. M., Palmer, G. H. and Barbet, A. F., 1993. Expression and immune recognition of the conserved MSP4 outer membrane protein of *Anaplasma marginale*. *Infect. Immun.* 61, 5245-51.

Odorico, M. and Pellequer, J. L., 2003. BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J. Mol. Recogn*. 16, 20-2.

Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y., 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. *Infect. Immun*. 66, 132-9.

Osbourn, J. K., Earnshaw, J. C., Johnson, K. S., Parmentier, M., Timmermans, V. and McCafferty, J., 1998. Direct selection of MIP-1α neutralizing CCR5 antibodies from a phage display human antibody library. *Nat. Biotechnol.* 16, 778-781.

Padmalayam, I., Karem, K., Baumstark, B and Massung, R., 2000. The gene encoding the 17 kDa antigen of *Bartonella henslae* is located within a cluster of genes homologous to the virB virulence operon. *DNA Cell. Biol*. 19, 377-82.

Palzkill, T., Huang, W. and Weinstock, G. M., 1998. Mapping protein-ligand interactions using whole genome phage display libraries. *Gene*. 221, 79-83.

Parmley, S. F. and Smith, G. P., 1988. Antibody-selectable filamentous fd vectors: affinity purification of target genes. *Gene*. 73, 305-18.

Pereboeva, L. A., Pereboeva, A. V., Wang, L. F. and Morris, G. E., 2000. Hepatitis C epitopes from phage-displayed cDNA libraries and improved diagnosis with a chimeric antigen. *J. Med. Virol*. 60, 144-51.

Pereira, S., Maruyama, H., Siegel, D., van Belle, P., Elder, D., Curtis, P. and Herlyn, D., 1997. A model system for detection and isolation of a tumour cell surface antigen using antibody phage display. *J. Immunol. Methods*. 203, 11-24.

Perez, J-M., Martinez, D., Debus, A., Sheikboudou, C. and Bensaid, A., 1997. Detection of genomic polymorphisms among isolates of the intracellular bacterium *Cowdria ruminantium* by random amplified polymorphic DNA and Southern blotting. *FEMS Microbiol. Lett.* 154, 73-9.

Perez, J-M., Martinez, D., Sheikboudou, C., Jongejan, F. and Bensaid, A., 1998. Characterisation of variable immunodominant antigens of *Cowdria ruminantium* by ELISA and immunoblots. *Parasite Immunol*. 20, 613-22.

Peter, T. F., Anderson, E. C., Burridge, M. J. Perry, B. D. and Mahan, S. M., 1999a. Susceptibility and carrier status of impala, Sable and tsessebe for *Cowdria ruminantium*. *J. Parasitol*. 85, 468-7.

Peter, T. F., Bryson, N. R., Perry, B. D., O'Callaghan, C. J., Medley, G. F., Smith, G. E., Mlambo, G., Horak, I. G., Burridge, M. J. and Mahan, S. M., 1999b. *Cowdria ruminantium* infection in ticks in the Kruger National Park. *Vet. Rec*. 145, 304-7.

Peter, T. F., Deem, S. L., Barbet, A. F., Norval, A. I., Simbi, B. H., Kelly, P. J. and Mahan, S. M., 1995. Development and evaluation of PCR assay for detection of low levels of *Cowdria ruminantium* infection in *Amblyomma* ticks not detected by DNA probe. *J. Clin. Microbiol*. 33, 166-72.

Petersen, G., Song, D., Hügle-Dörr, B., Oldenburg, I. and Bautz, E. K., 1995. Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragment phage display libraries. *Mol. Gen. Genet*. 249, 425-31.

Pfaff, E., Mussgay, M., Bohm, H. O., Schulz, G. E., Schaller, H., 1982. Antibodies against a preselected peptide recognize and neutralize foot and mouth disease virus. *EMBO.* 1, 869-74.

Pincus S. H., Smith M. J., Jennings, H. J., Burritt, J. B. and Glee, P. M., 1998. Peptides that mimic the group Bstreptococcal type III capsular polysaccharide antigen. *J. Immunol.* 160, 293-8.

Prickett, D. and Hall, R., 2000. *Theileria annulata*: Epitope mapping neutralizing antibodies using phage display. *Exp. Parasitol*. 94, 42-4.

Provost, A. and Bezuidenhout, J. D., 1987. The historical background and global importance of heartwater. *Onderstepoort J. Vet. Res*. 54, 165-9.

Prozesky, L., 1987a. Heartwater. The morphology of *Cowdria ruminantium* and its staining characteristics in the vertebrate host and *in vitro. Onderstepoort J. Vet. Res*. 54, 173-6.

Prozesky, L. 1987b. The pathology of heartwater. III. A review. *Onderstepoort J. Vet. Res*. 54, 281-6.

Purchase, H. S., 1945. A simple and rapid method for demonstrating *Rickettsia ruminantium* (Cowdry, 1925) in heartwater brains. *Vet. Res*. 57, 413-4.

Reddy, G. R., Sulsona, C. R., Harrison, R. H., Mahan, S. M., Burridge, M. J. and Barbet, A. F., 1996. Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. *Clin. Diagn. Lab. Immunol*. 3, 417-22.

Reichmann, L. and Winter, G., 2000. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Nat. Acad Sci. USA.* 97, 10068-73.

Rodi, D. J., Janes, R. W., Sanganee, H. J., Holton, R. A., Wallace, B. A. and Makowski, L., 1998. Screening of a library of phage-displayed peptides identifies human Bcl-2 as a taxol-binding protein. *J. Mol. Biol.* 285, 197-203.

Rodi, D. J. and Makowski, L., 1999. Phage-display technology–finding a needle in a fast molecular haystack. *Curr. Opin. Biotechnol*.10, 87-93.

Roitt, I., 1997. Essential Immunology. Ninth Edition. Blackwell Science Ltd. Oxford.

Rondot, S., Koch, J., Breitling, F. and Dübel, S., 2001. A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnology*. 19, 75-8.

Rossouw, M., Neitz, A. W. H., de Waal, D. T., du Plessis, J. L., van Gas, L. and Brett. S., 1990. Identification of the antigenic proteins of *Cowdria ruminantium. Onderstepoort J. Vet. Res.* 57, 215-21.

Sambrook, J., Fritsch, E. F. and Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

Sanna, P. P., Williamson, R. A., De Logu, A., Bloom, F. E. and Burton, D. R., 1995. Direct selection of recombinant human monoclonal antibodies to herpes virus glycoproteins from phage display libraries. *Proc. Nat. Acad. Sci. USA.* 92, 6439-43.

Saunders, N. J. and Moxon, E. R., 1998. Implications of sequencing bacterial genomes for pathogenesis and vaccine development. *Curr. Opin. Biotechnol.* 9, 618-23.

Scott, J. K., 1992. Discovering peptide ligands using epitope libraries. *TIBS.* 17, 241-5.

Scott, J. K. and Barbas, C. F., 2000. Phage display vectors. In *Phage display. A Laboratory Manual*. (ed. Barbas, C. F., Burton, D. R., Scott, J. K. and Silverman, G. J.) Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Scott, J. K. and Smith, G. P., 1990. Searching for peptide ligands with an epitope library. *Science.* 249, 386-90.

Shompole, S., Rurangirwa, F. R., Wambugu, A., Sitienei, J., Mwangi, D. M., Musoke, A. J., Mahan, S., Wells, C. W. and McGuire, T. C., 2000. Monoclonal antibody binding to a surface-exposed epitope on *Cowdria ruminantium* that is conserved among eight strains. *Clin. Diagn. Lab. Immunol*. 7, 983-6.

Sibille, P. and Strosberg, A. D., 1997. A FIV epitope defined by a phage peptide library screened with a monoclonal anti-FIV antibody. *Immunol. Letters*. 59, 133-7.

Skilton, R. A., Bishop, R. P., Wells, C. W., Spooner, P. R., Gobright, E., Nkonge, C., Musoke, A. J., Macklin, M. and Iams, K. P., 1998. Cloning and characterisation of a 150 kDa microsphere antigen of Theileria parva that is immunologically cross-reactive with the polymorphic immunodominant molecule (PIM). *Parasitology*. 117, 321-30.

Smith, G. P., web page, http://www.biosci.missouri.edu/smithgp/index.html.

Smith, G. P., 1985. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science*. 228, 1315-7.

Smith, G. P., 1992. Cloning in fUSE vectors. February 10, 1992 ed. Division of Biological Sciences, University of Missouri, CO.

Sparks, A. B., Adey, N. B., Cwirla, S. and Kay, B. K., 1996. Screening phage-displayed random peptide libraries. In *Phage display of peptides and proteins. A Laboratory Manual* . Edited by Kay, B. K., Winter, J. and McCafferty, J., Academic Press, San Diego.

Staden, R., 1994. The Staden package, Vol 25, p. 9-170. *In* A.M. Griffin and H. G. Griffin (Eds.), "Methods in molecular biology" Humana Press Inc., Totawa, NJ.

Stephen, C. W., Helminen, P. and Lane, D. P., 1995. Characterisation of epitopes on human p53 using phage-displayed peptide libraries: Insights into antibody-peptide interactions. *J. Mol. Biol*. 248, 58-78.

Stoute, J. A., Ballou, W. P., Kolodny, N., Deal, C. D., Wirtz, R. A. and Lindler, L. E., 1995. Induction of humoral immune response against *Plasmodium falciparum* sporozoites by immunization with a synthetic peptide mimotope whose sequence was derived from screening a filamentous phage epitope library. *Infect. Immun.* 63, 934-9.

Sulsona, C. R., Mahan, S. M. and Barbet, A. F., 1999. The *map1* gene of *Cowdria ruminantium* is a member of a multigene family containing both conserved and variable genes. *Biochem. Biophys. Res. Commun*. 257, 300-5.

Thalhamer, J. and Freund, J.,1985. Passive immunisation: a method of enhancing the immune response against antigen mixtures. *J. Immunol. Methods*. 80, 7-13.

Theisen, M., Soe, S., Jessing, S. G., Okkels, L. M., Danielsen, S., Oeuvray, C., Druilhe, P. and Jepsen, S., 2001. Identification of a major B-cell epitope of the *Plasmodium falciparum* glutamine rich protein (GLURP), targeted by human antibodies mediating parasite killing. *Vaccine*. 19, 204-12.

Totté, P., Nyanjui, J.,Bensaid, A. and McKeever, D., 1999. Bovine $CD4^+$ T-cell lines reactive with soluble and membrane

antigens of *Cowdria ruminantium*. *Vet. Immunol. Immunopathol*. 70, 269-76.

Totté, P., McKeever, D., Martinez, D. and Bensaid, A., 1997. Analysis of T-cell responses in cattle immunized against heartwater by vaccination with killed elementary bodies of *Cowdria ruminantium*. *Infect. Immun*. 65, 236-41.

Totté, P., Vachiery, N., Martinez, D., Trap, I., Ballingall, K. T., MacHugh, N. D., Bensaid, A. and Wérenne, J., 1996. Recombinant bovine interferon gamma inhibits the growth of *Cowdria ruminantium* but fails to induce major histocompatibility complex class ll following infection of endothelial cells. *Vet. Immunol. Immunopathol*. 53, 61-71.

Towbin, J., Staehelin, T. and Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Nat. Acad. Sci. USA.* 76, 4350-4.

Uilenberg, G., 1983. Heartwater (*Cowdria ruminantium*) infection: current status. *Adv. Vet. Sci. Com. Med.* 27, 427-80.

Underwood, P. A., 1985. Theoretical considerations of the ability of monoclonal antibodies to detect antigenic differences between closely related variants, with particular reference to heterospecific reactions. *J. Immunol. Methods.* 85, 295-307.

Urbanelli, L., Fortugno, P., Bartoli, F., Nuzzo, M., de Tomassi, A., Felici, F. and Monaci, P., 2000. 'Affinity maturation' of ligands for HCV-specific serum antibodies. *J. Immunol. Methods*. 236, 167-76.

Vachiéry, P., Trap, I., Totté, P., Martinez, D. and J.,Bensaid, A., 1998. Inhibition of MHC class I and class II cell surface expression on bovine endothelial celss upon infection with *Cowdria ruminantium*. *Vet. Immunol. Immunopathol*. 61, 37-48.

Valadon, P., Nussbaum, G., Boyd, L. F., Mergulies, D. H. and Scharff, M. D., 1996. Peptide libraries define the fine specificity of anti-polysaccharide antibodies to *Cryptococcus neoformans. J. Mol. Biol*. 261, 11-22.

Van de Pypekamp, H. E. and Prozesky, L., 1987. Heartwater. An overview of the clinical signs, susceptibility and differential diagnoses of the disease in domestic ruminants. *Onderstepoort J. Vet. Res*. 54, 263-6.

Van Kleef, M., Gunter, N. J., Macmillan, H., Allsopp, B. A., Shkap, V. and Brown, W. C., 2000. Identification of *Cowdria ruminantium* antigens that stimulate proliferation of lymphocytes from cattle immunised by infection and treatment or inactivated organisms. *Infect. Immun*. 68, 603-14.

Van Kleef, M., Neitz, A. W. H. and de Waal, D. T., 1993. Isolation and characterization of antigenic proteins of *Cowdria ruminantium*. *Rev. Élev. Méd. vét. Pays trop*. 46, 157-64.

Van Regenmortel, M. H. V., 1988. Molecular dissection of protein antigens and the prediction of epitopes. In *Synthetic polypeptides as antigens*. Edited by van Regenmortel, M. H. V., Briand, J. P., Muller, S. and Plaué, S.

Van Regenmortel, M. H. V., 1995. Transcending the structuralist paradigm in immunology-affinity and biological activity rather than purely structural considerations should guide the design of synthetic peptide epitopes. *Biomed. Pept. Proteins Nucleic Acids.* 1, 109-16.

Van Regenmortel, M. H. V., 1998. From absolute to exquisite specificity. Reflections on the fuzzy nature of species, specificity and antigenic sites. *J. Immunol. Methods.* 216, 37-48.

Van Regenmortel, M. H. V., 1999. The use of peptides for diagnosing viral infections. In *Synthetic peptides as antigens*. Edited by van Regenmortel, M. H. V. and Muller, S.

Van Vliet, A. H. M., Jongejan, F. and van der Zeijst, B. A., 1992. Phylogenetic position of *Cowdria ruminantium* (Rickettsiales) determined by analysis of amplified 16S ribosomal DNA sequences. *Int. J. Syst. Bacteriol*.42, 494-8.

Van Vliet, A. H. M., Jongejan, F., van Kleef, M. and van der Zeijst, B. A. M., 1993. Cloning and partial characterisation of the Cr32 gene of *Cowdria ruminantium*. *Rev. Élev. Méd. vét. Pays trop*. 46, 167-70.

Van Vliet, A. H. M., Jongejan, F., van Kleef, M. and van der Zeijst, B. A. M., 1994. Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium. Infect. Immun.* 62, 1451-6.

Van Vliet, A. H. M., van der Zeijst, B. A. M., Camus, E., Mahan, S. M., Martinez, D. and Jongejan, F., 1995. Use of a specific immunodominant region on the *Cowdria ruminantium* MAP1 protein in a serological assay. *J. Clin. Microbiol*. 33, 2405-10.

Van Wezenbeek P. M., Hulsebos, T. J. and Schoenmakers, J. G., 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene.* 11, 129-48.

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. and Johnson, K. S., 1996. Human antibodies with sub-nanomolar affinities isolated from a large non-immunised phage display library. *Nat. Biotechnol.*14, 309-314.

Verhaert, R. M., van Duin, J. and Quax, W. J., 1999. Processing and functional display of the 86kDa heterodimeric penicillin G acylase on the surface of phage fd. *Biochem. J.* 342, 415-22.

Visser, E. S., McGuire, T. C., Palmer, G. H ., Davis, W. C., Shkap, V., Pipano, E. and Knowles, D. P., 1992. The *Anaplasma marginale msp*5 gene encodes a 19-kilodalton protein conserved in all recognized *Anaplasma* species. *Infect. Immun.* 60, 5139-44.

Viti, F., Nilsson, F., Demartis, S., Huber, A. and Neri, D., 2000. Design and use of phage display libraries for the selection of antibodies and enzymes. *Methods Enzymol*. 326, 480-505.

Waghela, S. D., Rurangirwa, F. R., Mahan, S. M., Yunker, C. E., Crawford, T. B., Barbet, A. F., Burridge, M. J. and Mcguire, T. C., 1991. A cloned DNA probe identifies *Cowdria ruminantium* in *Amblyomma variegatum* ticks. *J. Clin. Microbiol*. 29, 2571-7.

Wan, Y., Wu, Y., Bian, J., Wang, X., Zhou, W., Jai, Z., Tan, Y. and Zhou, L., 2001. Induction of hepatitis B virus-specific cytotoxic T lymphocytes response in vivo by filamentous phage display vaccine. *Vaccine*. 19, 2918-23.

Wang, L. F., du Plessis, D. H., White, J. R., Hyatt, A. D. and Eaton, B. T., 1995. Use of a gene-targeted phage display random epitope library to map an antigenic determinant on the bluetongue virus outer capsid VP5. *J. Immnunol. Methods*. 178, 1-12.

Webster, R., 1996. Biology of the Filamentous Bacteriophage. In *Phage display of peptides and proteins. A Laboratory Manual* . Edited by Kay, B. K., Winter, J. and McCafferty, J., Academic Press, San Diego.

Webster, R., 2001. Filamentous Phage Biology. In *Phage Display. A Laboratory Manual*. Edited by Barbas, C. F., Burton, D. R., Scott, J. K. and Silverman, G. J. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Wells, J. A., 1996. Hormone mimicry. *Science*. 273, 449-50.

Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A. and Van Regenmortel, M. H. V., 1984. Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature*. 311, 123-126.

Wilkins, M. R. and Gooley, A. A., 1997. Protein identification in proteome projects. In *Proteome Research: New Frontiers in Functional Genomics*. Edited by Wilkins, M. R., Williams, K. L., Appel, R. D. and Hochstrasser, D. F., Springer-Verlag, Berlin.

Winkler, H. H., 1990. *Rickettsia* species (as organisms). *Annu. Rev. Microbiol*. 44, 131-53.

Woods, S. A., Miles, J. S., Roberts, R. E. and Guest, J. R., 1986. Structural and functional relationships between fumarase

and aspartase. *Biochem. J*., 237, 547-557.

Young, R. A. and Davis, R. W., 1983. Efficient isolation of genes by using antibody probes. *Proc. Nat. Acad. Sci. USA*. 80, 1194-8.

Yu, M., Wang, L-F., Shiell, B. J., Morrissy, C. J. and Westbury, H. A., 1996. Fine mapping of a C-terminal linear epitope highly conserved among the major envelope glycoprotein E2 (gp51 to gp54) of different pestiviruses. *Virology*. 222, 289-92.

Zhong, G., Smith, G. P., Berry, J. and Brunham, R. C., 1994. Conformational mimicry of a chlamydial neutralization epitope on filamentous phage. *J. Biol. Chem*. 269, 24183-8

Zweygarth, E., Josemans, A. I. and Horn, E., 1998. Serum-free media for the *in vitro* cultivation of *Cowdria ruminantium*. *Ann. N. Y. Acad. Sci*. 849, 307-12.