UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS
UV

Identification of Antigenic Regions and Linear B cell Epitopes on Yellow Fever Virus

---

Shannon Lucrecia Smouse
M. Med. Sc.

Dissertation submitted in fulfilment of the requirements for the M.Med.Sc. Virology degree in the Faculty Health Sciences, at the University of the Free State

Promoter: Professor FJ Burt
Department of Medical Microbiology and Virology,
Faculty of Health Sciences, University of the Free State

University of the Free State, Bloemfontein Campus
February 2013

Table of Contents

DECLARATION i

PUBLICATIONS AND PRESENTATIONS ii

ABSTRACT iii

LIST OF FIGURES vi

LIST OF TABLES x

LIST OF ABBREVIATIONS xii

Chapter 1 1

LITERATURE REVIEW 1

1.1. Introduction 1

1.2. Viral genome structure and characteristics 1

1.3. Replication 4

1.4. Aetiology and Epidemiology 5

1.5. Molecular epidemiology of yellow fever 10

1.6. Clinical features of yellow fever 12

1.7. Pathogenesis 13

1.8. Diagnosis 14

1.9. Prevention measures and control 15

1.10. Vaccines 15

1.11. Epitope mapping 18

1.12. Recombinant antigens 20

1.13. Problem identification, aims and objectives 21

Chapter 2 23

CLONING AND OPTIMIZATION OF EXPRESSION OF YELLOW FEVER VIRUS EDIII, C AND NS4a PROTEINS 23

2.1. Introduction 23

2.2. Materials and methods 26

2.2.1. Amplification of genes encoding EDIII, C and NS4 proteins of YFV 26

2.2.1.1. RNA isolation 26

2.2.1.2. Reverse transcriptase-polymerase chain reaction (RT-PCR) amplification of viral RNA 27

2.2.1.3. Agarose gel electorophoresis 29

2.2.1.4. Purification of PCR product 29

2.2.1.5. Preparation of Luria Bertani (LB) media and agar 30

2.2.2. Cloning of EDIII, C and NS4a genes in pGEM®-T Easy 30

2.2.2.1. Plasmid DNA purification 33

2.2.2.2. Identification of positive transformants by restriction enzyme digestion 33

2.2.3. Cloning of EDIII, C and NS4 genes into the pQE-80L bacterial expression vector 34

2.2.3.1. Preparation of chemically competent cells for protein expression 34

2.2.3.2. Transformation of OverExpress cells 35

2.2.3.3. Identification of positive transformants 37

2.2.3.4. Sequencing of plasmid DNA 38

2.2.4. Optimization of expression using pQE-80L constructs 39

2.2.5. Cloning of EDIII gene into the pCold TF DNA bacterial expression vector 41

2.2.6. Induction and optimization of expression using pCold™ TF DNA construct 42

2.2.7. SDS-PAGE 42

2.2.8. Protein purification 44

2.2.9. Refolding purified protein after denaturation 46

2.2.10. Protein Concentration 46

2.2.11. Cleavage of TF from the recombinant pCold TF EDIII protein 47

2.2.12. Characterization of expressed proteins by Western blot analysis 48

2.3. Results 49

2.3.1. One step RT-PCR amplification of the genes encoding the C, NS4a and EDIII of YFV 49

2.3.2. A/T cloning of the YFV C, NS4a and EDIII amplicons into the pGEM®-T Easy vector 51

2.3.3. Cloning of the YFV EDIII into pQE-80L 55

2.3.4. Cloning of the YFV C gene into pQE-80L 56

2.3.5. Sequence analysis of the gene encoding YFV C in pQE-80L vector 58

2.3.6. Cloning of the YFV NS4a gene into pQE-80L 58

2.3.7. Sequence analysis of the gene encoding YFV NS4a in pQE-80L vector 60

2.3.8. Cloning of YFV EDIII gene into pCold TF 60

2.3.9. Sequence analysis of the gene encoding YFV EDIII in pCold TF 63

2.3.10. Expression, solubility, purification and characterization of recombinant EDIII protein frompQE-80L-YFVEDIII construct 64

2.3.11. Expression, solubility, purification and characterization of expressed EDIII protein using the pCold TF DNA vector 75

2.3.12. Expression of the C and NS4a constructs using the pQE-80L vector 79

2.4. Summary 81

Chapter 3 83

EVALUATION OF RECOMBINANT YFV EDIII PROTEIN FOR USE IN AN INDIRECT ELISA AND IDENTIFICATION OF LINEAR B CELL EPITOPES USING PEPTIDE LIBREARIES COVERING THE C, EDIII AND NS4a 83

3.1. Introduction 83

3.2. Materials and methods 84

3.2.1. Identification of flavivirus diverse regions 84

3.2.2. Human serum samples 84

3.2.3. *In-vitro* neutralisation assay 85

3.2.4. Preparation of YFV cell lysate antigen 86

3.2.5. Preparation of the recombinant YFV EDIII antigens using pQE-80L and pCold TF 87

3.2.6. ELISA 87

3.2.6.1. ELISA for the detection of IgG antibodies using the recombinant YFV pQE-80L-EDIII antigen 88

3.2.6.2. ELISA for the detection of IgG antibodies using YFV whole viral cell lysate antigen 88

3.2.7. Determination of ELISA cut-off values 89

3.2.8.1. Immunisation of mice 90

3.2.10. Mapping linear B cell epitopes 91

3.2.10.1. Prediction of antigenic and hydrophilic regions 91

3.2.10.2. Peptide libraries 91

3.2.10.3. ELISA for epitope identification 91

3.3. Results 92

3.3.1. Alignment of coding nucleotide and amino acid sequence data for the EDIII protein of YFV 17D vaccine strain, WNV Lineage 1 strain NY99, DENV-2 and JEV 92

3.3.2. Epitope prediction, antigenicity and hydrophilicity 92

3.3.3. Virus neutralisation 97

3.3.4. Detection of antibody against YFV using ELISA and whole cell lysate antigen 98

3.3.5. Detection of antibody against YFV using ELISA with recombinant YFV pQE-80L-EDIII antigen 98

3.3.6. Detection of antibody against recombinant YFV pColdTF-EDIII antigen in human sera 99

3.3.7. ELISA for detection of IgG antibodies in immunized mice 101

3.3.8. Serological cross-reactivity between YFV and WNV 102

3.3.9. Detection of antibody against C, NS4a and EDIII peptide libraries 103

3.4. Summary 106

Chapter 4 107

DISCUSSION 107

REFERENCES 112

APPENDIX A 119

Nucleotide and amino acid sequence data 119

APPENDIX B 138

ELISA Raw data 138

APPENDIX C 145

Abstract of poster presented at the South African Society for Biochemistry and Molecular Biology (SASBMB), 18-20th January 2010, Bloemfontein 145

Abstract of poster presented at the 13th International Congress on Infectious Diseases (ICID) 9 – 12th March 2010, Miami, USA 146

Abstract of poster presented at the Faculty of Health Sciences, Research Forum 26-27th August 2010, University of the Free State, Bloemfontein 147

Abstract of poster presented at the Virology Africa 2011 Congress, UCT Graduate School of Business, V&A Waterfront, Cape Town, South Africa 148

Abstract of poster presented at the Faculty of Health Sciences, Research Forum 23-24th August 2012, University of the Free State, Bloemfontein 149

OPSOMMING 150

DECLARATION

I certify that the dissertation hereby submitted by me for the M.Med.Sc. Virology degree at the University of the Free State is my independent effort and had not previously been submitted for a degree at another university/faculty. I furthermore waive copyright of the dissertation in favour of the University of the Free State.

_____

Shannon Lucrecia Smouse

PUBLICATIONS AND PRESENTATIONS

S Smouse and FJ Burt. Preparation of antigenically active recombinant yellow fever viral envelope domain III protein. Internat J Infect Dis. 2010;14 (Supplement 1): e2-e190.

Smouse SL and Burt FJ. Identification of human defined yellow fever virus linear B cell epitopes using peptide libraries. (Manuscript in Preparation).

Poster presentations

- Preparation of antigenically active recombinant yellow fever viral envelope domain III protein. 14th International Conference on Infectious Diseases (ICID), Miami, Florida from 9-12th March 2010.
- Recombinant yellow fever viral envelope protein for the detection of antibodies. South African Society for Biochemistry and Molecular Biology (SASBMB) Congress, Bloemfontein, South Africa from 18-20 January 2010.
- Preparation of recombinant yellow fever viral envelope domain III protein for use as a diagnostic reagent and induction of neutralising antibody response. Virology Africa 2011, UCT Graduate School of Business, V&A Waterfront, Cape Town, South Africa from 29 November – 2 December 2011.

Oral presentations

- Induction of humoral immunity from linearized epitopes on the domain III region of yellow fever virus envelope protein. Faculty of Health Sciences Research Forum, UFS, Bloemfontein, South Africa from 26-27 August 2010.
- Identification of linear B cell epitopes in the capsid, NS4a and domain III region in the E glycoprotein of yellow fever virus. Faculty of Health Sciences Research Forum, UFS, Bloemfontein, South Africa from 24-25 August 2012.

ABSTRACT

Yellow fever virus (YFV) virus is an arthropod-borne virus that causes viral hemorrhagic fever in humans in the tropical parts of both Africa and South America. The virus belongs to the family *Flaviviridae*, of the genus *Flavivirus* comprising of approximately 70 viruses. It is transmitted to vertebrates by the bite of an infected female mosquito, primarily the *Aedes* species. It is a re-emerging pathogen with case-fatality rates that can exceed 50% in humans. YFV can cause an acute febrile illness in humans which can progress to severe disease with hepatic and renal failure. The diagnosis of infection and testing of the immune status of vaccinees require reagents that are prepared in biosafety level (BSL) three and four facilities. Therefore the development of a recombinant antigen that does not require BSL three facilities for preparation and is safe to use, would have an important role in a diagnostic laboratory for detecting antibodies in infected individuals and vaccinees. Despite the availability of a live-attenuated efficacious vaccine, it is not recommended for immunocompromised individuals, thus development of new generation vaccines would have important public health implications. Identification and mapping of antigenic regions and viral epitopes is important for development of subunit vaccines and improved diagnostics. Subunit vaccines focusing on antigens that induce a protective immune response provide a safe approach to the development of vaccines against diseases causing severe and frequently fatal haemorrhagic fevers. The aim of this study was to identify immunodominant viral proteins that induce detectable antibody responses that could be used for developing diagnostic assays and to identify linear B cell epitopes on selected viral proteins.

The complete open reading frame of the genes encoding the domain III (EDIII) region of the envelope protein, capsid (C) and NS4a proteins of YFV were amplified, from the 17D strain of YFV, by RT-PCR using primers specifically designed from sequence data retrieved from GenBank. Oligonucleotide primers were modified with *Bam*HI and *Hind*III restriction enzyme sites that facilitated downstream cloning. Each amplicon was cloned into the pGEM®-T Easy cloning vector using T/A cloning. Each gene was rescued from the recombinant plasmid using *Bam*HI and *Hind*III restriction enzyme sites and ligated into bacterial expression system, pQE-80L vector. In a previous study, the YFV EDIII gene was cloned into pQE-80L and expressed in JM109 *Escherichia coli* cells however extremely low yields were obtained. In this study the expression levels were improved using different cell lines and optimizing incubation conditions. An insoluble 13 kDa protein was expressed from the construct and confirmed by Western blot

analysis. The protein was expressed with a 6 x Histidine tag that was used to facilitate purification using a Ni^{2+} column under denaturing conditions. Attempts to express the YFV C and NS4a proteins were not successful and expression was abandoned. In an attempt to improve solubility the YFV EDIII gene was excised from the pGEM®-T Easy vector and subsequently cloned into pCold TF bacterial expression vector. A ~65 kDa soluble protein was expressed from the construct and purified under native conditions.

The functional activity of the recombinant antigens in ELISA was compared with whole cell lysate antigen prepared from cell cultures infected with YFV. The biological activity of the recombinant YFV pQE-80L-EDIII antigen was confirmed in immunoassays using serum samples from humans vaccinated with YFV vaccine. Positive sera failed to react in ELISA using pCold TF expressed antigen and this antigen was excluded from further assays. A total of 20/24 serum samples from human vaccinees collected at varying stages after vaccination reacted in an ELISA with the recombinant YFV pQE-80L-EDIII protein and 24/24 reacted in ELISA with whole cell lysate antigen. The EDIII region of the envelope protein was shown to be able to differentiate between West Nile Virus infection and YFV infection in a limited number of convalescent horse sera. The recombinant EDIII protein was used to immunize mice. Serum samples collected from the mice reacted against whole cell lysate antigen in ELISA and was shown to have neutralising antibodies using an *in vitro* neutralisation assay. Hence the EDIII region of the envelope protein likely induces an important protective immune response. Finally, bioinformatics was used to predict possible epitope regions and using peptide libraries spanning predicted sites, one potential epitopic region was identified in the EDIII protein. Putative epitopic and antigenic regions along the length of the C, NS4a and EDIII proteins of each strain were predicted using the BCPREDS and ABCpred software.

In conclusion, the EDIII protein, an immunodominant antigen of YFV, prepared in this study has some potential for differentiation of flavivirus antibodies although it lacks sensitivity for routine diagnosis. A potential epitope, TGHGTVVMQ, from amino acid 21 to 29 on the EDIII protein was identified using bioinformatics and was shown to have reactivity against immune sera. The significance of this epitope needs further investigation. Finally the EDIII region of the YFV protein shows potential as a target region for vaccine development as shown for other flaviviruses but which has not previously been published for YFV.

ACKNOWLEDGEMENTS

I would like to thank the following institutions:

- The Department of Medical Microbiology and Virology for providing me with the facilities to complete my M.Med.Sc studies.

- The Grow Our Own Timber (GOOT) Scholarship: Atlantic Philanthropies programme, a UFS initiative to encourage development of young academics, for their financial support during 2009-2010.

- The Stars of Academe and Research (SoAR) Scholarship for their financial support during 2011-2012.

- The Postgraduate School of Medicine Research Council for their financial support during 2011 and 2012 at the University of the Free State.

- The National Health Laboratory Service (NHLS) Research Trust fund for their financial assistance.

I would like to thank the following individuals:

- Professor Felicity J. Burt for being a wonderful supervisor throughout my project. Thank you for your continuous assistance whenever I experienced problems in the laboratory. Thank you for your determination, encouragement and vast knowledge of science that contributed to a quality project. Thank you for your motivation and inspiration. I have gained so much knowledge and I intend to learn more.

- My colleagues and friends for their neverending support. Thank you to Lehlohonolo Mathentheng, Kulsum Kondiah, Rudo Samudzi, Mitta Mamabolo, Hermanus Hanekom, Azeeza Rangunwala, Carina Combrinck and my amazing friends Shalane Kelderman, Carla Di Lillie, Premaloshni Naidoo, Beatrice Kyambadde and Eriva Kyambadde for their support.

- My parents, brothers and grandmother for their unconditional love and support. Thank you for your continuous prayers and for having so much faith in me. I will always be grateful.

- The Lord, my saviour for his unconditional love and continuous support during this challenging time.

LIST OF FIGURES

CHAPTER 1

Figure 1.1. The YFV virion illustrating the immature (intracellular) and mature (extracellular) infectious virion. 2

Figure 1.2. Schematic representation of the *flavivirus* genome illustrating the positions of the genes encoding the structural and NS proteins and the known function of each protein. 3

Figure 1.3. The Flavivirus replication cycle. 5

Figure 1.4. *Aedes aegypti* mosquito responsible for the transmission of YFV between humans. 6

Figure 1.5. Transmission cycles of YFV. The virus is maintained between monkeys and tree-hole breeding mosquitoes. 7

Figure 1.6. Geographic distribution of YFV in endemic zones. The maps represent areas that are at risk for YFV transmission since 2009, (A) Africa and (B) Americas. 8

Figure 1.7. Phylogenetic relationships among genotypes of YFV from Africa and South America using nucleotide sequences of the pr/M and E regions. 11

Figure 1.8. Stages of yellow fever infection, showing the major clinical and laboratory features of the disease. 12

CHAPTER 2

Figure 2.1. Vector map, promoter and multiple cloning site of pQE-80L, pQE-81L and pQE-82L. 24

Figure 2.2. Vector map, promoter and multiple cloning site of pCold™ TF DNA. 26

Figure 2.3. Vector map, promoter and multiple cloning site of pGEM®-T Easy. 31

Figure 2.4. Agarose gel electrophoresis analysis of YFV EDIII amplicon. 50

Figure 2.5. Agarose gel electrophoresis analysis of YFV C and NS4a amplicons. 50

Figure 2.6. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 52

Figure 2.7. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVC construct using *Bam*HI and *Hind*III restriction endonucleases. 52

Figure 2.8. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVNS4a construct using *Bam*HI and *Hind*III restriction endonucleases. 53

Figure 2.9. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested YFV EDIII gene. 54

Figure 2.10. Agarose gel electrophoresis analysis of restriction enzyme digestion of purified double digested YFV C gene. 54

Figure 2.11. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested YFV NS4a gene. 55

Figure 2.12. Agarose gel electrophoresis analysis of restriction enzyme digestions of the pQE-80L-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 56

Figure 2.13. Agarose gel electrophoresis analysis of YFV C amplicons on three pQE-80L colonies. 57

Figure 2.14. Agarose gel electrophoresis analysis of restriction enzyme digestions of the pQE-80L-YFVC construct using *Bam*HI and *Hind*III restriction endonucleases. 57

Figure 2.15. Agarose gel electrophoresis analysis of YFV NS4a amplicons on one pQE-80L colony. 59

Figure 2.16. Agarose gel electrophoresis analysis of restriction enzyme digestions of the pQE-80L-NS4a construct using *Bam*HI and *Hind*III restriction endonucleases. 59

Figure 2.17. Agarose gel electrophoresis analysis of restriction enzyme digestions of pCold TF plasmid DNA. 61

Figure 2.18. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested pCold TF plasmid. 61

Figure 2.19. Agarose gel electrophoresis analysis of YFV EDIII amplicons on three pCold TF colonies. 62

Figure 2.20. Agarose gel electrophoresis analysis of restriction enzyme digestions of the pColdTF-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 63

Figure 2.21. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3, T=4, T=5 hours and 16 hours post-induction, with a final IPTG concentration of 0.125 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 65

Figure 2.22. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 0.125 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 66

Figure 2.23. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3, T=4, T=5 hours and 16 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 67

Figure 2.24. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 68

Figure 2.25. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress *E.coli* BL21 (DE3) cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 69

Figure 2.26. SDS-PAGE analysis of proteins expressed at T=4, T=5, T=16 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress *E.coli* BL21 (DE3) cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 70

Figure 2.27. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress *E.coli* BL21 (DE3) cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 71

Figure 2.28. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (0.125 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 72

Figure 2.29. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (1 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 73

Figure 2.30. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (2.5 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 73

Figure 2.31. Detection of recombinant EDIII protein (0.125 mM and 1 mM IPTG) by Western blot analysis using an anti-His tag antibody. Expressed protein was detected at 16 hours post-induction after purification under denaturing conditions. 74

Figure 2.32. Detection of recombinant EDIII protein (2.5 mM IPTG) by Western blot analysis using an anti-His tag antibody. Expressed protein was detected at 16 hours post-induction after purification under denaturing conditions. 74

Figure 2.33. SDS-PAGE analysis of proteins expressed at T=0, T=4, T=16 and T=24 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pColdTF-YFVEDIII. 76

Figure 2.34. SDS-PAGE analysis of proteins at T=16 hours and T=24 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pColdTF-YFVEDIII. 77

Figure 2.35. SDS-PAGE analysis of purified proteins at T=24 hours post-induction (1 mM IPTG), expressed using the recombinant expression construct, pColdTF-YFVEDIII. 78

Figure 2.36. SDS-PAGE analysis of purified pColdTF-YFVEDIII protein cleaved using Factor Xa. 79

Figure 2.37. SDS-PAGE analysis of proteins expressed at T=4 and T=5 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVC and pQE-80L-YFVNS4a. 80

CHAPTER 3

Figure 3.1. Parker Hydrophilicity Prediction plots of the A. C, B. EDIII and C. NS4a proteins (YFV 17D vaccine strain) generated using the IEDB software. 95, 96

Figure 3.2. Detection of IgG antibodies against YFV using whole cell lysate antigen vaccines. 98

Figure 3.3. Detection of IgG antibodies against recombinant YFV pQE-80L-EDIII antigen using human sera from vaccinees. 99

Figure 3.4. Detection of IgG antibodies against recombinant YFV pColdTF-EDIII antigen. 100

Figure 3.5. Assessment of the biological activity of the recombinant YFV pQE-80L antigen over four months. 101

Figure 3.6. Antibody reactivity against YFV whole cell lysate antigen using sera from mice immunized with the YFV 17D vaccine and recombinant YFV pQE-80L-EDIII antigen. 102

Figure 3.7. Detection of IgG antibodies against YFV cell lysate and recombinant YFV pQE-80L-EDIII antigens using convalescent sera form WNV confirmed horses. 103

Figure 3.8. Reactivity of sera from YFV vaccinees against synthesized overlapping peptides spanning the NS4a protein (NS4a_1-NS4a5), C protein (C6-C9) and EDIII protein (ED10-ED28) in an indirect ELISA. 105

LIST OF TABLES

CHAPTER 1

Table 1.1. Countries in Africa and the Americas at risk of yellow fever. 9

CHAPTER 2

Table 2.1. Oligonucleotide primers used to amplify the genes encoding the EDIII, NS4a and C proteins of YFV. 28

Table 2.2. Reaction mix for One Step RT-PCR for the amplification of genes encoding YFV EDIII, C and NS4a proteins. 28, 29

Table 2.3. Ligation reaction mixture used for ligation of genes encoding EDIII, C and NS4a proteins and pGEM®-T Easy vector. 32

Table 2.4. Restriction enzyme reaction mix for identification of positive transformants. 34

Table 2.5. Restriction enzyme digestion of pQE-80L plasmid. 36

Table 2.6. Ligation reaction mixture used for ligation of genes encoding EDIII, C and NS4a proteins and pQE-80L plasmid. 36

Table 2.7. Colony PCR for the detection of YFV EDIII, C and NS4a genes. 37, 38

Table 2.8. Restriction analysis for positive transformant identification. 38

Table 2.9. Restriction enzyme digestion of pCold™ TF DNA. 41

Table 2.10. Ligation reaction mixture used for ligation of gene encoding EDIII protein and linearized pCold™ TF DNA vector. 41

Table 2.11. Preparation of an 8% resolving gel for SDS-PAGE analysis. 43

Table 2.12. Preparation of a 4% stacking gel for SDS-PAGE analysis. 43

Table 2.13. Restiction enzyme digestion of TF fusion protein. 47

Table 2.14. Protein concentrations of purified recombinant YFV EDIII protein from cultures induced with 0.125 mM, 1mM and 2,5mM IPTG. 72

CHAPTER 3

Table 3.1. *In vitro* virus neutralisation test using YFV 17D vaccine strain. 86

Table 3.2. Nucleotide and amino acid sequence changes among YFV vaccine strains 17D, 17DD, 17D-213, 17D-204 and the YFV Asibi (wild type) strain. 93

Table 3.3. Common regions identified from linear epitope prediction methods as likely epitopic and antigenic regions on the C, NS4a and EDIII proteins. 94

Table 3.4. Serum samples from volunteers with a history of YFV vaccination and antibody results. 97

Table 3.5. Peptide libraries of 8 – 9 mers with 3 offset residues were synthesized. The libraries spanned predicted epitopic regions on C, NS4a and EDIII proteins. 104

.

LIST OF ABBREVIATIONS

ABTS – 2,2'-azino diethyl-benzothiazoline-sulfonic acid peroxidase substrate

Ae. – Aedes

AP61 – Ae. Pseudoscutellaris

ATCC – American Tissue Culture Collection

BBS – borate buffered saline

bla – beta-lactamase gene

bp – base pairs

BSA – bovine serum albumin

BSL – biosafety level

C++ - high positive serum control

C+ - low positive serum control

C- - negative serum control

C – capsid

cfu – colony forming units

*csp*A - cold-shock protein A

CPE – cytopathic effect

DENV2 – dengue virus serotype 2

DIC – disseminated intravascular coagulation

DMSO – dimethyl sulfoxide

DNA – deoxyribonucleic acid

cDNA – complementary DNA

DS – denaturing solubilisation

DTT - dithiothreitiol

E – envelope

EDIII – envelope domain III

ELISA – enzyme-linked immunosorbent assay

EMEM – Eagle's Minimal Essential Media

EPO - erythropoetin

ER – endoplasmic reticulum

FCS – foetal calf serum

FITC – fluorescein isothiocyanate

FNV – French neurotropic vaccine

GAGs - glycoaminoglycans

HI – hemagglutination inhibition

HIV – human immunodeficiency virus

HRPO – horse radish peroxidase

IEDB – Immune Epitope Database

IFA – indirect immunofluorescence assay

Ig - immunoglobulin

IPTG – isopropyl β-D-1-thiogalactopyranoside

IMAC – immobilized metal ion chromatography

JEV – Japanese encephalitis virus

kDa – kilo Dalton

LAV – live attenuated vaccine

LB – luria-bertani

LEW – lysis-equilibration-wash

LIV – Louping ill virus

M – membrane

MAbs – monoclonal antibodies

$MAbs^{R}$ – Mab neutralisation-resistant escape

MCS – multiple cloning site

MW – molecular weight

NCR – non-coding region

Ni^{2+} - nickel

NS – non-structural

NT – neutralisation test

NTA – nitroloacetic acid

OD – optical density

ORF – open reading frame

PBS – phosphate buffered saline

prM – pre-membrane

PCR – polymerase chain reaction

pmol – picomolar

PVDF – polyvinylidene difluouride

RBSII – synthetic ribosomal binding site

RNA – ribonucleic acid

RT-PCR – reverse transcriptase-PCR

SD – standard deviation

SDS – sodium dodecyl sulphate

SDS PAGE – SDS-polyacrylamide gel electrophoresis

SLEV – Saint Louis encephalitis virus

SOB – super optimal broth

SOC – super optimal catabolite-repression

SNF – supernatant fluid

TAE – Tris acetate EDTA

TBE – tick-borne encephalitis

TBS – Tris buffered saline

TCID – tissue culture infectious dose

TED – Tris-carboxymethyl ethylene diamine

TEE – translation enhancing element

TF – trigger factor

Tm – melting temperature

tPA – tissue plasminogen activator

WHO – World Health Organisation

WNV – West Nile virus

X-gal – 5-bromo-4-chloro-3-indolyl-[beta]-D-galactosidase

YFV – yellow fever virus

Chapter 1

LITERATURE REVIEW

1.1. Introduction

Yellow fever virus (YFV) was first described in the 17th century as a disease entity. In 1881, a Cuban doctor and scientist, Carlos Finlay, first proposed that yellow fever was a filterable agent transmitted by mosquitoes rather than direct contact with humans. In 1900, an army surgeon, Walter Reed, confirmed Finlay's theory by exposing human volunteers to body fluids and excrement of other yellow fever sufferers by exposing other volunteers to the virus through potentially infected mosquitoes (Frierson, 2010). Only those exposed to mosquito bites developed the disease (Monath, 1986). Despite a highly efficacious vaccine, YFV remains a significant public health concern.

YFV is the prototype member of the genus *Flavivirus* which consists of approximately 70 viruses, many of which are medically significant pathogens (Mutebi *et al.,* 2004). It belongs to the family *Flaviviridae.* YFV can cause an acute febrile illness in humans which can progress to severe disease with hepatic and renal failure, haemorrhage and shock related syndromes. It is an arboviral disease that has re-emerged in tropical parts of Africa and South America in the last two decades and is primarily transmitted by the bite of an infected female mosquito. The flavivirus genus is divided into three groups based on transmission route: mosquito-borne, tick-borne and those with "no known vector" (Kuno *et al.,* 1998). An estimated 80% of *Flaviviruses* are arthropod-borne.

1.2. Viral genome structure and characteristics

Flaviviruses are RNA viruses with a single-stranded positive sense RNA genome. Mature virions have a nucleocapsid with icosahedral symmetry that encloses the RNA (Figure 1.1). The genome of flaviviruses is approximately 11-kb in length, and includes a single open reading frame (ORF) flanked by 5' and 3' non-coding regions (NCR's) that serves as a messenger for translation of a precursor polyprotein, which is processed by proteolytic cleavage. The viral genome encodes three structural proteins capsid (C), pre-membrane (prM), envelope (E) and seven non-structural (NS) proteins, NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5, that are numbered in order of their synthesis (Figure 1.2) (Monath, 1989). The NS proteins are involved

in replication of the RNA. NS5 is an RNA-dependent RNA polymerase, NS3 has protease and helicase activity and NS2B is a co-factor for protease activity. These are essential for the replication process and may be useful targets for antiviral drug development. The structural proteins are responsible for the assembly of the flavivirus virion, incorporating a lipid bilayer and RNA genome (Patkar *et al.,* 2007). The C protein serves as a scaffold around which the envelope proteins and the lipid bilayer are organised. The C protein is the least conserved protein of the flaviviruses, but contains basic amino acid residues and a hydrophobic region. The hydrophobic region is involved in binding the C protein to the viral envelope protein and the basic amino acid residues function by interacting with negatively charged phosphate groups of the RNA genome. The small, hydrophobic NS4a protein plays a role in RNA replication and contains several transmembrane domains including a signal sequence (Acheson NH, 2007).

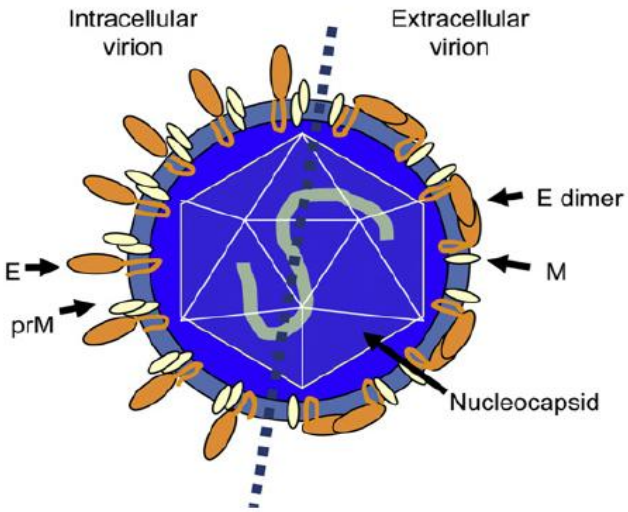


Figure 1.1. The YFV virion illustrating the immature (intracellular) and mature (extracellular) infectious virion (Gardner and Ryman, 2010).

The E protein comprises a single glycoprotein and is the major structural protein of the virus responsible for attachment to host cell receptors. It contains important type and group specific epitopes, and epitopes eliciting neutralising antibodies (Monath, 1989; Ryman *et al.,* 1997). The glycoprotein has a host membrane derived lipid bilayer with dimers of E protein anchored on the surface by hydrophobic tails. The viral E protein consists of 12 cysteine residues that are completely conserved amongst all *flaviviruses*. This suggests that regions of the E protein structure are fundamentally conserved while other portions of the E protein, exposed to the host

immune system, may be quite distinctive among the various *flaviviruses*. This is supported by the lack of cross-protection induced by flavivirus vaccines and among people infected with heterologous viruses, regardless of the ability to generate flavivirus cross-reactive antibodies (Holzmann *et al*., 1996). The E protein is divided into three distinct domains designated I, II, III that can be distinguished both serologically and within an X-ray crystal structure (Holbrook *et al.,* 2004). The Ig-like fold present in the domain III protein is commonly associated with structures that have an adhesion function (Chin *et al.,* 2007). The E domain III (EDIII) protein is highly antigenic and consists principally of linear-epitopes and has been proposed as the viral receptor-binding domain based on, neutralising monoclonal antibodies (MAbs) and cell binding sites (Chu *et al.,* 2007; Beasley and Barrett, 2002; Crill and Roehrig, 2001). Studies using competitive binding of MAbs have identified the presence of multiple distinct epitopes on the E protein of several different *flaviviruses* (Mason *et al.,* 1989).

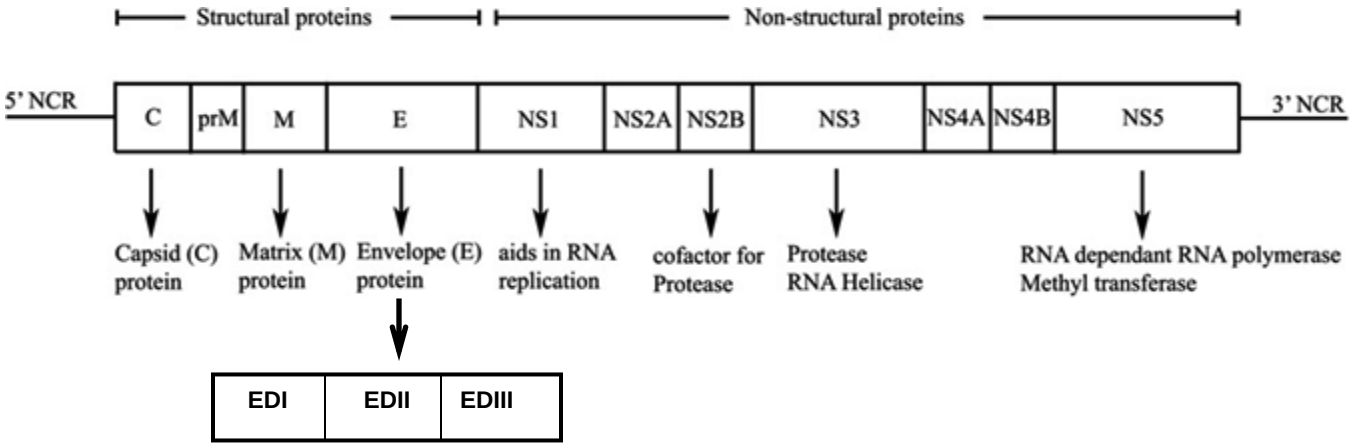


Figure 1.2. Schematic representation of the *flavivirus* genome illustrating the positions of the genes encoding the structural and NS proteins and the known function of each protein (Carrington and Auguste, 2012).

Determining the role of the E protein in the pathogenesis process has been facilitated by the selection of single-site virus variants with altered ability to induce encephalitis in the mouse model (Ryman *et al*., 1998). To select antigenic virus variants *in vitro*, neutralising MAbs can be used, which are termed MAb neutralisation-resistant escape (MAb^R) variants (Ryman *et al*., 1998). This approach has been used to isolate MAb^R variants for several *flaviviruses* including

tick-borne encephalitis (TBE), Japanese encephalitis virus (JEV), Louping ill virus (LIV), Dengue virus (DENV) serotype 2, Murray Valley encephalitis virus (MVEV) and YFV (Ryman *et al*., 1998). In several cases single amino acid mutations in the E protein affecting pathogenicity were observed and were associated with EDIII protein (Ryman *et al.,* 1998). The epitopes within each antigenic domain have similar structural properties and the domains usually correspond to different structural entities on the protein (Mandl *et al.,* 1989). Hence the EDIII protein possesses ideal properties for use as a diagnostic and research tool.

1.3. Replication

Replication of YFV occurs within the cytoplasm and is associated with the cellular membranes of the cell. All mosquito-borne *flaviviruses*, including YFV, share conserved RNA regions and structures (Bredenbeek *et al.,* 2003). The NS proteins assemble to form the viral RNA replicase, together with the genomic RNA template and host factors, on the cytoplasmic membranes. Replication begins through the synthesis of a complementary negative-stranded RNA, which serves as a template for positive-strand genomic RNA synthesis. The asymmetric RNA replication process leads to a 10- to 100-fold excess of positive strands compared to negative strands. The positive strands are translated, replicated or packaged into virions. These negative strands accumulate throughout the infection and have been isolated solely in double-stranded forms (Lindenbach *et al*., 2003). The replication process occurs via replicative intermediates, double stranded replicative forms consisting of only double stranded RNA, and finally the synthesis of full length genomic RNA. The virus particles are assembled within the rough endoplasmic reticulum, after which the immature virions are transported out of the cell as shown in Figure 1.3. The cleavage of prM occurs prior to release of the virion and is responsible for converting the particle to its mature form. Immature, non-infectious virions assemble within the endoplasmic reticulum (ER) and is packaged into an ER-derived lipid bilayer containing heterodimers of the prM and E proteins suggesting that budding occurs intracellularly through the host cell membrane (Lorenz et al., 2003; Mackenzie et al., 2001; Yu et al., 2008; Li et al., 2008). Immature flavivirus particles are transported to the cell surface by the translocation of immature virion-containing vesicles from membranous components of the cell to the plasma membrane. Maturation of the virion occurs in the trans-Golgi network, thus triggering rearrangements in the E protein that promote infectivity and as a result the mature, infectious virus particles are released into the extracellular medium by exocytosis (Chambers et al., 1990).

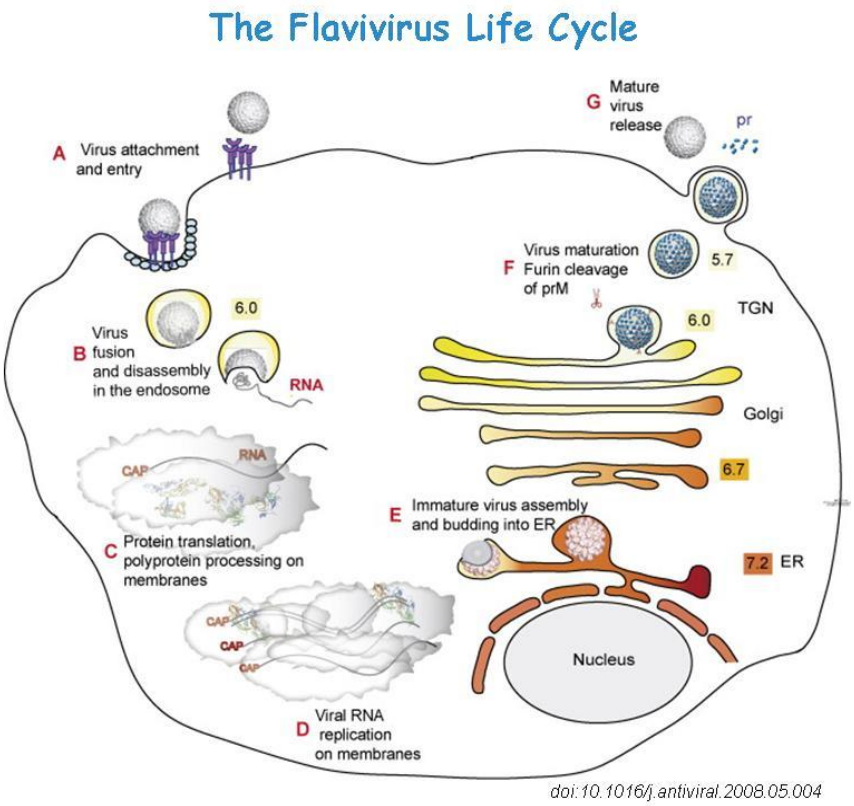


Figure 1.3. The Flavivirus replication cycle (http://www.infectionlandscapes.org/2011/07/yellow-fever.html).

1.4. Aetiology and Epidemiology

YFV is transmitted by a female mosquito as shown in Figure 1.4. YFV is maintained in three epidemiological patterns of transmission involving non-human primates and arboreal mosquitoes as illustrated in Figure 1.5. The virus is maintained in sylvatic (jungle) transmission cycles involving nonhuman primates, but utilises humans as the sole vertebrate host in urban epidemics. In South America, sylvatic vectors belong to the *Haemogogus* and *Sabethes* genera, and the urban vector is *Aedes (Ae.) aegypti*, while in Africa *Ae.* species mosquitoes serve as both sylvatic and urban vectors (Mutebi *et al*., 2001). Although several mosquito species are involved in the transmission cycles of YFV in Africa the most important species are *Ae. africanus*, *Ae. aegypti*, *Ae. simpsoni* complex, *Ae. furcifer*, *Ae. taylori* and *Ae. luteocephalus*. In Africa, the virus is maintained in nature via all three cycles of transmission, sylvatic or jungle, intermediate and urban cycle. The sylvatic cycle involves transmission among monkeys and

small mammals by the mosquito *Ae. africanus* in rain and gallery forests of adjacent humid savannahs (Mutebi and Barret, 2002). *Ae. africanus* is a forest mosquito that breeds primarily in tree-holes and plays a central role in YFV ecology in East, Central and West Africa (Mutebi and Barrett, 2002). The intermediate cycle has only been described in the moist savannah regions of Africa. It involves several mosquito species, including *Ae. fucifer, Ae. metallicus, Ae. opok, Ae. taylori, Ae. vittatus* and members of the *Ae. simpsoni* complex (Mutebi and Barrett, 2002). The intermediate cycle, in which the virus is transmitted from monkey to human and from human to human via mosquitoes, frequently occurs in small village settlements and communal herding areas (Mutebi and Barrett, 2002). In drier areas the virus spills over into urban regions and is maintained by human to human transmission via domestic mosquitoes.

Figure 1.4. *Aedes aegypti* mosquito responsible for the transmission of YFV between humans (Gardner and Ryman, 2010).

In South America, YFV is principally maintained in the rain forest, in a cycle involving non-human primates and one or more mosquito species of the genus *Haemagogus* or *Sabethes* (Baronti *et al*., 2011). The virus may periodically emerge in human populations when individuals infringe upon the habitat supporting enzootic or epizootic transmission. Enzootic cases are generally sporadic and scattered, but epizootic periods are characterized by high vector density and infection rates. In addition, the virus is maintained by vertical transmission in mosquitoes and transovarial transmission, when ova containing virus survive in dry tree-holes and hatch infectious progeny mosquitoes during rainfall (Monath, 2001).

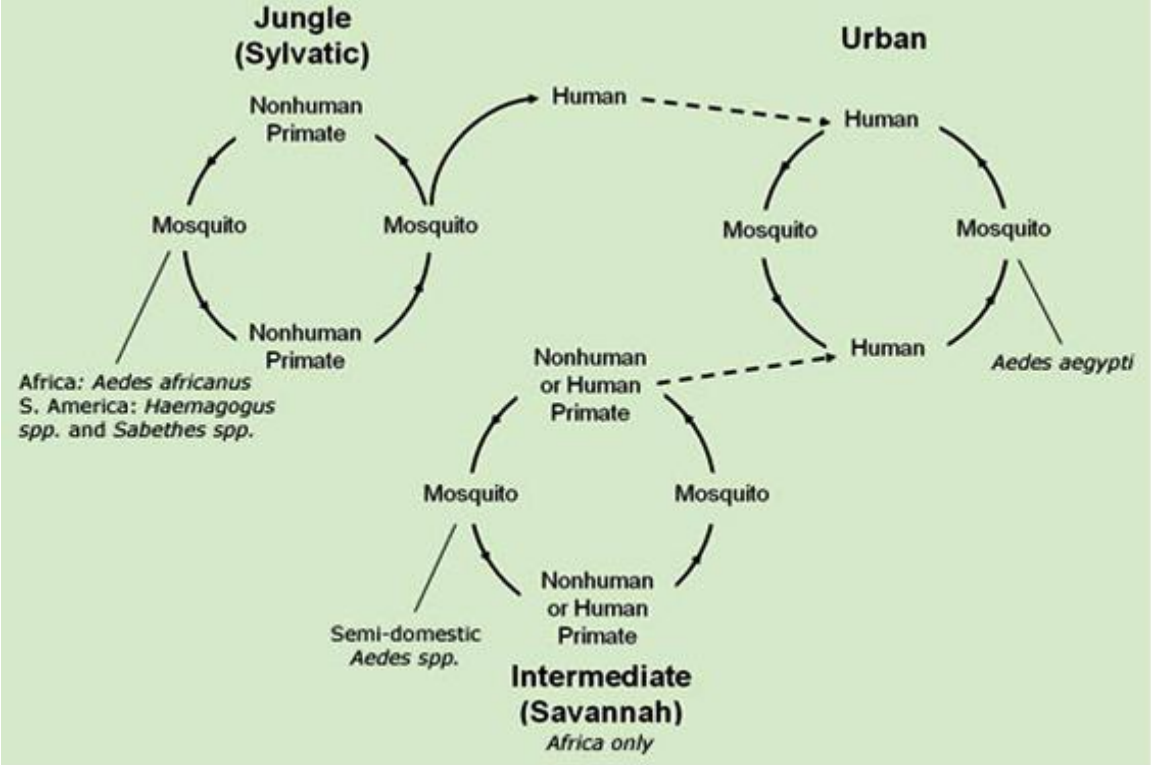


Figure 1.5. Transmission cycles of YFV. The virus is maintained between monkeys and tree-hole breeding mosquitoes (http://www.cdc.gov/yellowfever/transmission/index.html).

Strains of YFV were first isolated from humans in 1927 in Ghana and in Senegal (Lepineic *et al.,* 1994). YFV outbreaks are predominantly found in tropical regions of South America and sub – Saharan Africa and remains endemic in the equatorial forests of these areas as illustrated in Figure 1.6 (Lepineic *et al.,* 1994). The World Health Organization (WHO) estimates that there are approximately 200 000 human cases occurring annually in Africa, South America and Central America with approximately 30 000 fatalities. Fewer cases are documented because of under reporting of outbreaks (Barnett, 2007; Gardner and Ryman, 2010). In the past 15 years the incidence of YFV has steadily increased and at least one YFV outbreak is reported to the WHO every year (Mutebi and Barrett, 2002). Between 1990-1999, 11 297 YFV cases and 2648 deaths were reported to the WHO (Monath, 2005). Of these cases, approximately 9358 cases occurred in Africa with the largest number reported from Nigeria (Monath, 2005). Since 2000, notable outbreaks of YFV have occurred in Guinea, Sierra Leone, Liberia and Burkina Faso (Monath, 2005).

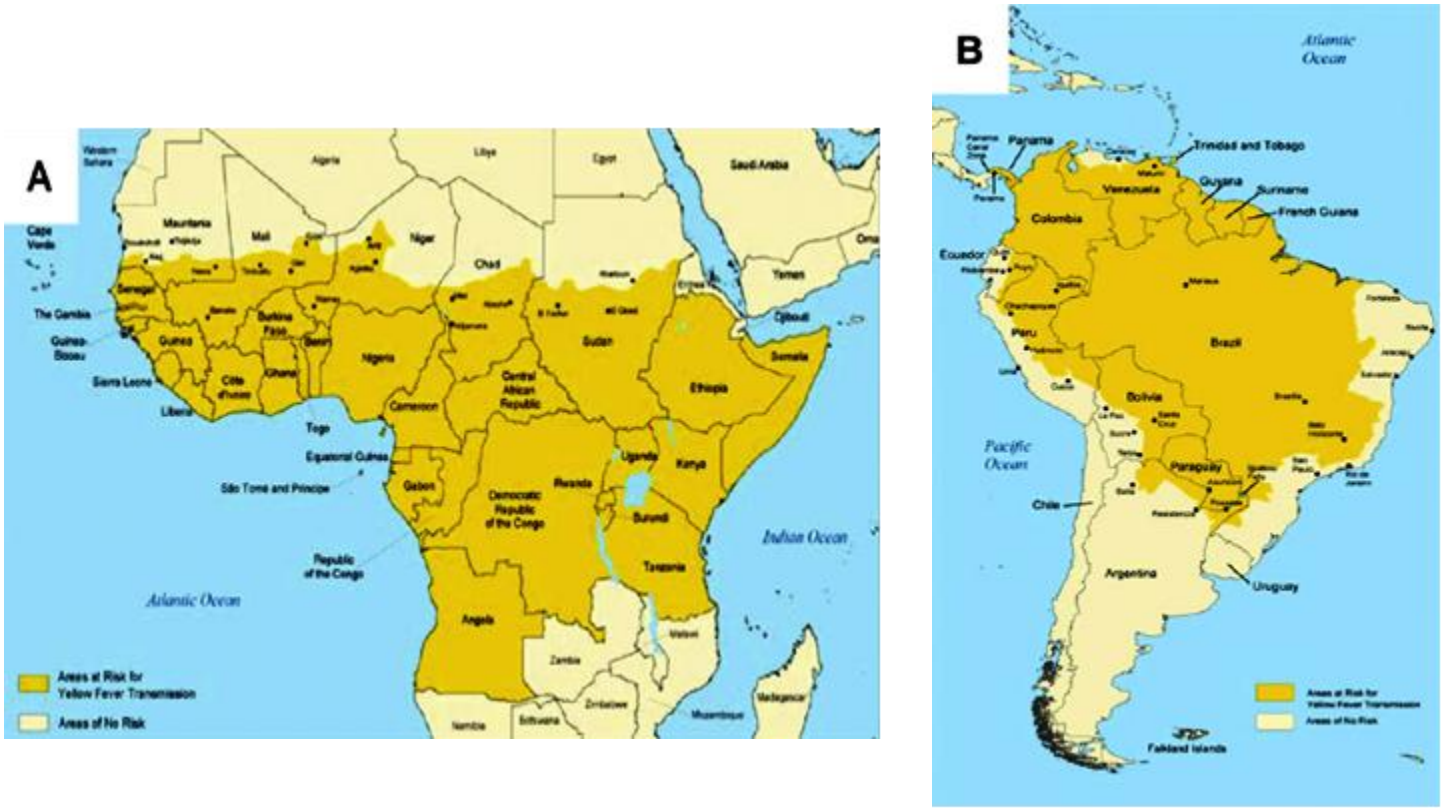


Figure 1.6. Geographic distribution of YFV in endemic zones. The maps represent areas that are at risk for YFV transmission since 2009, (A) Africa and (B) Americas (Gardner and Ryman, 2010).

In South America, between 1990-1999, approximately 200 cases were reported per year with a reported total of 1939 cases and 941 deaths. Evidence for the presence of YFV has been confirmed in 34 countries in Africa and 14 countries in South America as described in Table 1.1. There is no evidence that the virus is or has been present in Asia. Specific outbreaks usually occur in remote areas where confirmation of the cause is recognized late in the outbreak and the diagnosis is made on the basis of clinical disease with retrospective laboratory confirmation.

The spread of yellow fever in Africa is maintained by a high density of vector mosquito populations that are in close proximity to largely unvaccinated human populations. Despite the incorporation of yellow fever vaccines into childhood immunization programs in some countries, vaccine coverage is not optimal (Barnett, 2007). The rate of transmission of yellow fever is lower in South America than Africa because high vaccine coverage occurs primarily as part of mass immunization campaigns in response to outbreaks of the disease. The largest outbreak of yellow fever in South America since the 1950s occurred in Peru in 1995 and the re-emergence

of the disease in Brazil during the late 1990s and early 2000s prompted mass vaccination campaigns (Barnett, 2007).

Table 1.1. Countries in Africa and the Americas at risk of yellow fever (Gardner and Ryman, 2010).

Africa	
West Africa	Benin, Burkina Faso, Cape Verde, Côte d' Ivoire, Equatorial Guinea, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierre Leone, Togo
Central Africa	Angola, Burundi, Cameroon, Central African Republic, Chad, Democratic Republic of the Congo, Gabon, Rwanda
East Africa	Ethiopa, Kenya, Somalia, Sudan, Tanzania, Uganda
Americas	
Central America	Panama
South America	Argentina, Bolivia, Brazil, Colombia, Ecuador, Guyana, French Guyana, Paraguay, Peru, Suriname, Trinidad and Tobago, Venezuela

In recent years, there has been an alarming resurgence of virus circulation and expansion of the endemic zones in Africa and South America (Gardner and Ryman, 2010). In Africa during the 1940s, urban YFV was controlled, especially in French speaking West African countries, due to mass vaccination campaigns and efforts to remove *Ae. aegypti* breeding sites (Gardner and Ryman, 2010). In the 1980s, due to reduced or non-existent vaccine coverage an increase in enzootic transmission cycles were observed with an estimated 120 000 cases and a 20% fatality rate in Nigeria (Gardner and Ryman, 2010). Factors contributing to the re-emergence of YFV in Africa are inadequate vaccine coverage and rapid urbanization with cities becoming larger and urban populations increasing annually. In South America, jungle YFV continues to occur affecting young male forestry and agricultural workers in the Orinoco and Amazon river basins due to unsuccessful attempts to eradicate the *Haemagogus* species mosquito vectors (Gardner and Ryman, 2010). The Pan-American Health Organization reported an increase in jungle YFV affecting areas such as Argentina, Paraguay, Brazil, Colombia, Venezuela and Trinidad and Tobago in the Caribbean. YFV returned to Paraguay in 2008, after a 34-year absence, causing a cluster of possible urban YFV cases in Asunción (Gardner and Ryman, 2010). The outbreak was contained by mass vaccination campaigns of two million vaccine doses that were urgently requested from the global stockpile (Gardner and Ryman, 2010).

1.5. Molecular epidemiology of yellow fever

The first genetic studies of YFV identified three topotypes which corresponded with geographic distribution in West Africa, Central and East Africa, and South America (Deubel et al., 1986). Early studies on the genetic relationship of of YFV strains were performed using RNase T1 oligonucleotide fingerprinting, a technique where T1 ribonuclease digests virion RNA (Barrett, 2010; Carrington and Auguste, 2012; Deubel *et al.,* 1986). A characteristic "fingerprint" for each virus results from radiolabelled digested RNA separated by 2-dimensional electrophoresis based on the size and charge of fragments (Deubel et al., 1986; Lepineic et al., 1994). Studies using oligonucleotide fingerprinting described three variants of YFV of which two were found in Africa (Deubel et al., 1986). A different variant was described in the Central African Republic despite similarities between strains from Ivory Coast and Burkina Faso, thus demonstrating genetic variation between West Africa and East and Central African strains (Deubel et al., 1986). A third variant of YFV, based on analysis of three geographically distinct strains (Ecuador, Trinidad and Panama), was described in South America indicating that South American YFV strains were genetically differentiated from the African strains (Deubel et al., 1986; Lepineic et al., 1994). In a more recent study, which included a larger number of isolates, greater diversity was identified in Africa, with five genotypes designated Central/East Africa, East Africa, Angola, West Africa I and West Africa II (Von Lindern *et al*., 2006; Mutebi *et al*., 2001). The genetic stability and clustering according to the geographic location was further supported by later studies based on the analysis of nucleotide sequences from the E gene (Carrington and Auguste, 2012). Phylogenies were further confirmed from nucleotide sequence determinations targeting a 1320 nucleotide fragment from the 5' terminus, the region encoding the NS4a, 2K and NS4b, the 3'NCR, the prM/M and E junction and the complete ORF (Carrington and Auguste, 2012). Five genotypes are circulating in Africa and two genotypes are present in South America. Phylogenetic analysis has shown that sequences from West Africa form a separate clade distinct from a clade containing more divergent East and Central African sequences (Carrington and Auguste, 2012).

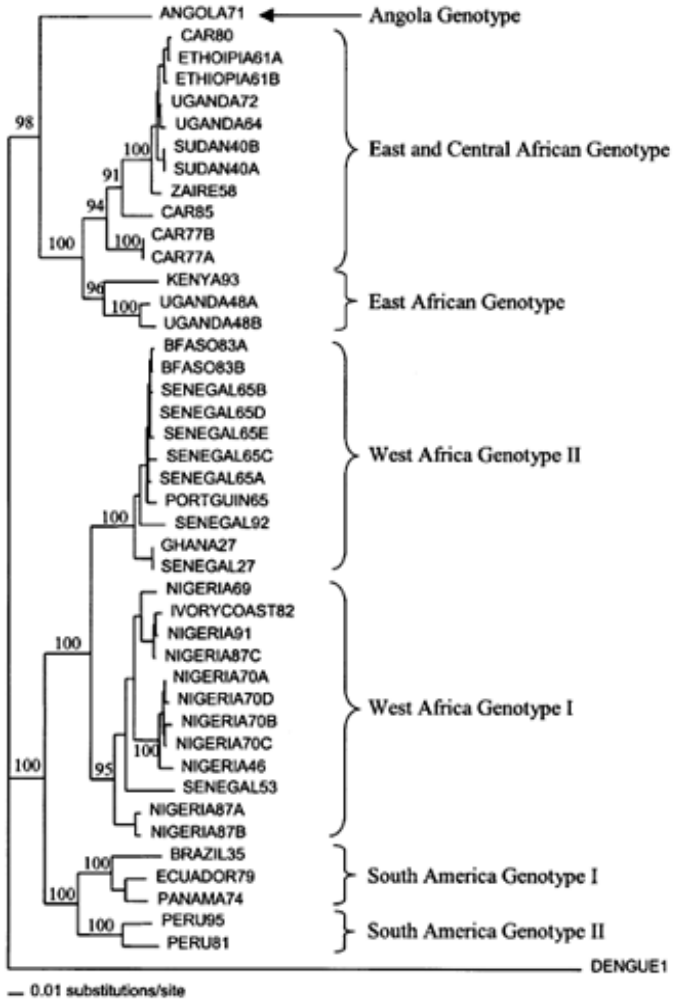


Figure 1.7. Phylogenetic relationships among genotypes of YFV from Africa and South America using nucleotide sequences of the pr/M and E regions (Mutebi and Barrett, 2002).

This evolutionary pattern suggests that YFV originated in East or Central Africa where three genotypes, Angola, East/Central Africa and East Africa have been identified (Von Lindern *et al.,* 2006; Mutebi *et al.,* 2001). Later YFV spread to West Africa where two genotypes were identified as West Africa genotype I and genotype II as shown in Figure 1.7 (Mutebi *et al*., 2001; Carrington and Auguste, 2012). YFV phylogenies indicate that South American sequences originated from a common ancestor shared with West African sequences suggesting that the virus was imported to the Americas during the slave trade (Carrington and Auguste, 2012). Currently two geographically defined genotypes exist in the Americas, namely genotype I found mainly in Brazil and neighbouring countries on the north side including Trinidad, Venezuela, Ecuador, Panama, and Colombia and genotype II found in countries to the East of the continent such as Peru and Bolivia overlapping into Brazil, Ecuador and Trinidad (Figure 1.7) (Carrington and Auguste, 2012).

1.6. Clinical features of yellow fever

Clinical disease varies from mild, febrile, non-specific illness to a severe and frequently fatal haemorrhagic fever including jaundice and renal failure. The incubation period after exposure to an infected mosquito ranges from 3-6 days. Viremia peaks 2-3 days after onset of illness but only 15% of individuals that are infected develop clinical yellow fever, whereas the rest develop a mild form and recover rapidly. The presentation of yellow fever disease ranges from subclinical infection to systemic disease including fever, jaundice, hemorrhage and renal failure (Barnett, 2007). As described in Figure 1.8 first phase symptoms, referred to as the "period of infection" during which virus is present in blood, is characterized by a sudden onset of fever, headache, generalized myalgia, muscle pain, backache, nausea and vomiting.

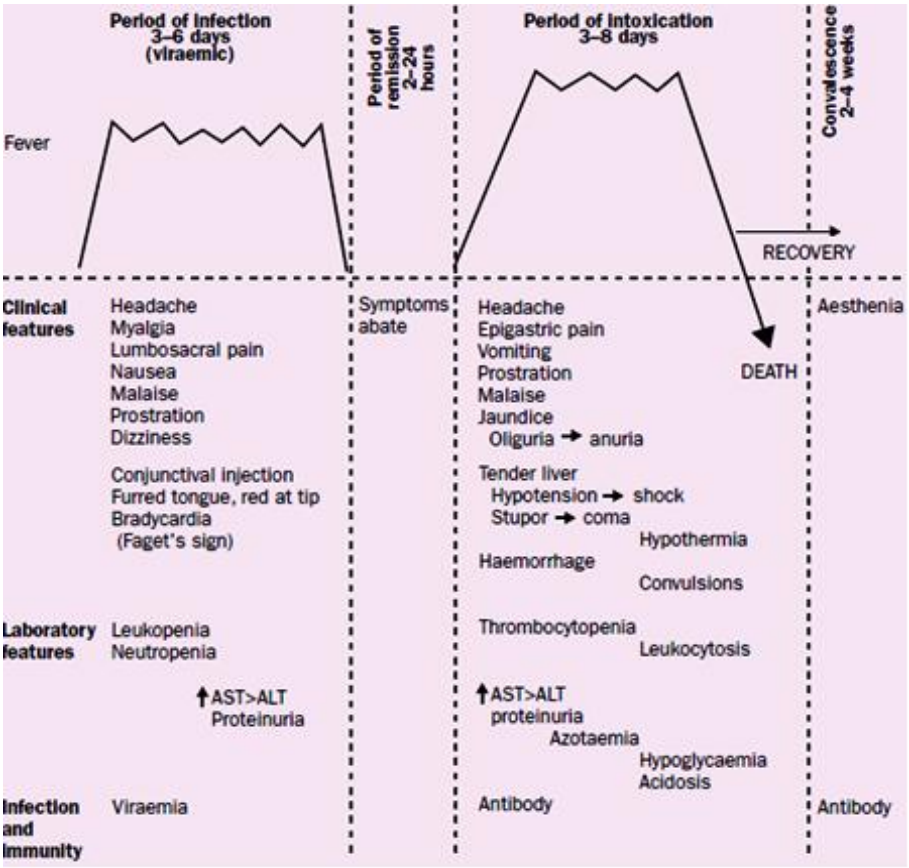


Figure 1.8. Stages of yellow fever infection, showing the major clinical and laboratory features of the disease (Monath, 2001).

Secondary phase referred to as the “period of remission” is characterized by an improvement in symptoms, including a reduction of fever. Some infected individuals recover at this phase without developing jaundice. The virus is frequently absent in the blood during the secondary phase which is more than eight days after onset of illness (Mutebi *et al*., 2002). At this stage anti-YFV antibodies can be detected (Mutebi *et al*., 2002). The virus is cleared by both the humoral and cellular immune response, although neutralising antibodies are known to play a significant role in protection. However 15-25% of cases develop a severe form of illness referred to as the “period of intoxication” and present with severe symptoms including jaundice, haemorrhagic tendancies and renal failure (Monath, 2001). Fatalities occur in 20-50% of people that develop renal failure. Patients that survive the infection have a prolonged weakness and fatigue.

1.7. Pathogenesis

In a variety of vertebrate hosts YFV produces both neurotropic and viscerotropic patterns of infection (Tomori, 2004). Pathology of human YFV infection reveals that the kidneys are generally enlarged, congested and edematous (Gardner and Ryman, 2010). The liver is the characteristic target of YFV infection. The appearance of an infected liver is normal or slightly enlarged in size and icteric with lobular markings. Pathological changes in the liver include swelling and necrosis of hepatocytes in the midzone of the liver lobule with sparing of cells surrounding the central veins and portal area (Monath 2001; Gardner and Ryman, 2010). Eosinophilic degeneration with condensed nuclear chromatin occurs due to the presence of infected hepatocytes leading to apoptotic cell death (Tomori, 2004). Viral antigen and RNA in cells demonstrating pathologic changes are detectable by immunocytochemistry and *in situ* nucleic acid hyrbidisation (Gardner and Ryman, 2010). Acute tubular necrosis is present in the kidneys, due to reduced perfusion of blood rather than direct viral injury (Gardner and Ryman, 2010). Decreased synthesis of clotting factors by the liver is a result of haemorrhaging and subsequent disseminated intravascular coagulation (DIC). Haemorrhagic symptoms and fatal outcomes are correlated with highly elevated pro- and anti-inflammatory cytokines during acute yellow fever infection (Gardner and Ryman, 2010). In addition, increased membrane permeability and apoptotic cell death is a result of mitochondrial damage (Marfin *et al*., 2005).

1.8. Diagnosis

Diseases most closely mimicking YFV that are also characterised by jaundice, haemorrhage, DIC and high case-fatality rates are leptospirosis and louse-borne relapsing fever (Monath, 2001). Viral hepatitis and severe malaria are other diseases that should also be considered in the differential diagnosis (Monath, 2001). Specific laboratory diagnostic tools rely on the detection of live virus, viral antigen or viral nucleic acid in blood or by serology (Monath, 2001; Wamala *et al.,* 2012). YFV may be isolated by intracranial inoculation of suckling mice, intrathoracic inoculation of mosquitoes or virus isolation by cell culture (Monath, 2001; Wamala *et al.,* 2012). Inoculation of cell cultures such as the use of *Ae. pseudoscutellaris* (AP61) and mammalian cell lines (Vero 76, BHK-21 cells) may be used in combination with PCR or detection of viral antigen in post mortem tissue by immunostaining (Monath, 2001). Viral nucleic acid has been detected by RT-PCR in clinical samples that are no longer detectable (Monath, 2001). Since human YFV infections are characterised by a phase of viremia of sufficient amount to infect vector mosquitoes, lasting several days to a week or more, it should be possible to directly detect viral antigen in blood by immunoassay (Monath *et al.,* 1986). However, antigen detection is not commonly used probably due to the introduction of molecular assays. Serological diagnosis of YFV is achieved principally by the detection of IgM antibodies by enzyme-linked immunosorbent assay (ELISA) (Monath, 2001; Wamala *et al.,* 2012). Traditional serological tests currently used for the detection of antibodies against YFV include hemagglutination-inhibition (HI), indirect immunofluorescence assay (IFA) and serum neutralisation test (NT) although HI and NT are unable to differentiate IgG and IgM responses and hence cannot distinguish between a recent or old infections. The presence of IgM antibodies in the late acute or early convalescent phase provides a presumptive diagnosis (Monath, 2001). All *flaviviruses* are serologically related and their cross-reactivity may lead to difficulties in interpretation of results the laboratory during diagnosis. In fatal cases YFV is commonly isolated from liver biopsies and blood specimens.

Due to extensive antigenic cross-reactivity among viruses, serological diagnosis of flavivirus infections are difficult, especially in regions where two or more of these human pathogenic viruses are endemic (Chávez *et al*., 2010). There have been attempts to develop serological tests that can be used diagnostically to distinguish between the various members of the *flavivurus* genus. The use of a recombinant, bacterially expressed West Nile virus (WNV)-EDIII antigen in an ELISA was described in Beasley et al. (2004). This study described the differentiation between antibody responses to WNV and other related *flaviviruses* such as Saint

Louis encephalitis virus (SLEV), JEV and MVEV. In discriminating specific antibody responses to WNV, the recombinant WNV EDIII antigen was superior to whole virus antigens using control mouse immune ascitic fluids raised against several JEV groups and other mosquito-borne flaviviruses (Beasley *et al*., 2004; Li *et al.,* 2005). In addition, an ELISA for detection of flavivirus antibody using recombinant EDIII antigen it was possible to differentiate the TBE flavivirus serocomplex flaviviruses from mosquito-borne flaviviruses (Holbrook *et al*., 2004). Recombinant EDIII peptides were also able to differentiate between specific viruses within the mosquito-borne flavivirus group using ELISA and Western blot assays (Holbrook *et al*., 2004). Recently, dos Santos *et al*. (2004) established a diagnostic test and evaluated serum immune responses against 20 recombinant polypeptides mimicking the entire genome of DENV-2 (dos Santos *et al*., 2004). The sensitivity with mouse brain whole virus antigen versus pD2-3 antigen (the most reactive peptide is located in the envelope protein) was 79% and 100% respectively when testing dengue acute and convalescent phase sera. The specificity using both antigens was 100% (dos Santos *et al*., 2004).

1.9. Prevention measures and control

Entry into YFV endemic countries or travel from YFV endemic countries to *Ae. aegypti*-infested countries, requires a certificate of vaccination under the International Health Regulations (Monath, 2001). The YFV 17D is a highly efficacious live attenuated vaccine that has been used for over 60 years. YFV epidemics may be prevented by increasing the use of vector control with pesticides or by the reduction and maintenance of domestic breeding to reduce virus transmission (Monath, 1989). These precautions include drainage or treatment of all mosquito breeding sites in surrounding areas (Monath *et al*., 2002). The use of protective clothing and repellents should be applied to exposed skin and to thin clothing that is penetrable by mosquito mouthparts, thus reducing the risk of exposure. There is no effective antiviral therapy, although supportive therapy is administered during the early onset of the disease.

1.10. Vaccines

The development of YFV vaccines began in 1927 after the isolation of YFV from the Ghanaian patient, Asibi. Researchers focused on a live attenuated virus vaccine (LAV) after failure with inactivated vaccines (Pulmanausahakul *et al*., 2010). The French neurotropic vaccine (FNV)

became the first effective YFV vaccine after passaging the French viscerotropic virus, Dakar strain, in suckling mice using the intracranial route (Pulmanausahakul *et al*., 2010). Despite its high level of efficacy the FNV was discontinued in 1982 due to the high incidence of adverse effects in vaccinees (Barnett, 2007). The YFV 17D became the second vaccine that was developed following 176 passages of the wild-type strain Asibi in both suckling mouse and chicken embryonic tissues (Pulmanausahakul *et al*., 2010). The 17D YFV vaccine is a live attenuated strain that was first developed in 1937 by Max Theiler (Theiler and Smith, 1937). Max Theiler was the son of Arnold Theiler that started the Ondersdepoort Veterinary Institute in South Africa. It has proven to be a highly effective and very safe vaccine that should be administered to travellers who enter areas of possible yellow fever activity (Monath, 1989). Since attenuated viruses may revert to virulence and the attenuated virus may adapt to host cells and increase in pathogenicity, safety has been the major concern of replication-competent vaccines. Two specific substrains are used currently and are both derived from the wild-type Asibi virus, substrains 17D-204, which are at passages 235-240, used in countries where it is produced and 17DD, which are at passages 287-289, used in Brazil (Barrett *et al*., 2009). The YFV vaccine used in southern Africa is based on the 17D-204 strain (Rossiter *et al.*, 2010).

Despite the availability of an efficacious vaccine and mosquito eradication programs, tropical South American countries such as Bolivia, Brazil, Colombia, Ecuador, French Guiana, Peru and Venezuela and much of the sub-Saharan Africa currently experience yellow fever epidemics despite a marked reduction in the world-wide incidence of yellow fever in the last five decades. Vaccination with YFV 17D strains elicits a protective immune response. The humoral immune response to the viral structural proteins is more significant in the protective effect induced by 17D vaccines. A single dose of the YFV vaccine provides protection in individuals producing 90% of neutralising antibodies within 10 days (Monath, 2001). The International Health regulations stipulate that the vaccination certificate for yellow fever is valid 10 days after administration of the vaccine because clinical trial studies have shown that 90% of individuals seroconverted within 10 days after vaccination and 100% of individuals seroconverted within 14 days. The WHO requires a revaccination after 10 years, although it has shown that immunity can last longer than 30 years (Barrett *et al*., 2009).

There are rare cases of serious adverse events following immunization with the 17D vaccine strain. Two types of serious adverse events have been reported to date, vaccine-associated neurotropic disease (YEL-AND) caused by neuroinvasion of the 17D virus and vaccine-associated viscerotropic disease (YEL-AVD) caused by a pansystemic infection starting often

with hepatic involvement, a condition very similar to wild-type yellow fever infection (Barrett *et al*., 2009). Despite the availability of a highly efficacious vaccine, it cannot be administered to the elderly, immunosuppressed or thymectomized individuals, infants under nine months of age and pregnant woman because of its potential to revert back to its wild-type (Monath *et al*., 2011). Serious adverse events could be circumvented with the use of a vaccine incapable of replicating in the host (Monath *et al*., 2011). The 17D vaccine also contains considerable amounts of chicken embryo proteins and allergic reactions contribute to the adverse events (Barrett *et al*., 2009).

Recently, a purified inactivated vaccine produced in Vero cells was developed and results from phase I clinical trials suggest good immunogenicity and adequacy (Monath *et al*., 2011). The development of cell culture vaccines should reduce the risk of allergic reactions, present in the 17D vaccine, since they do not contain the common allergens such as egg proteins (Monath *et al*., 2011). A single injection with this vaccine elicited antibody responses similar to those against the 17D vaccine, in hamsters and monkeys. The study showed that the immune responses against the inactivated cell culture vaccine were strongly dose dependant suggesting that stronger immune responses might be achieved with higher doses of the purified antigen (Monath *et al*., 2011). Although historical comparisons are compromised by differences in the neutralising antibody assays used, the safety and immunogenicity of the inactivated cell culture vaccine are similar to those of Japanese and tick-borne encephalitis vaccines (Monath *et al*., 2011).

The attenuated 17D YFV strain is one of the most effective and safe vaccines available for flaviviruses. The EDIII domain of the 17D strain binds more efficiently to the cell surface of glycosaminoglycans (GAGs). The interaction between EDIII and GAGs reduces the viremia and prevents viscerotropism of the YFV vaccine. This high affinity also leads to a rapid removal of virus from the bloodstream as a result of non-productive binding of virus to extracellular matrix components (rich in GAG) (Lee *et al*., 2008). Potential candidates for flavivirus vaccines are EDIII proteins because of their ability to induce the production of neutralising antibodies and cellular immune responses. Mice immunized with WNV-EDIII soluble protein showed production of high levels of IFN-Y and IL-2 cytokines, Th1-type cellular immune response and T lymphocyte proliferation (Chu *et al*., 2007). The MAb was mapped to an epitope localized on the EDIII of the WNV E protein and it protected >90% of the challenged mice (Chu *et al*., 2007; Chu *et al*., 2005). These findings show that EDIII proteins can induce limited cross-flavivirus protective effects (Wu *et al*., 2003). *In vivo* protection of WNV infections was observed when

anti-sera from mice immunized with a WNV-EDIII protein was pre-incubated with virus (Chu *et al*., 2005). Thus, protection was observed against WNV infection in suckling mice that were inoculated. The anti-sera produced with the WNV-EDIII protein were effective against JEV infection demonstrating a survival rate of 80% in animals.

1.11. Epitope mapping

Problems associated with laboratory diagnosis such as serological cross reactivity and the interest in developing a safe subunit vaccine increase the importance of identifying immunodominant epitopes. Antigenic analysis is important in identifying regions that induce a detectable antibody response. Activation of the host humoral immune response during virus infection results in antibody production against specific viral protein epitopes. Identification and mapping of viral epitopes is important for development of subunit vaccines and improved diagnostics (Wu *et al.*, 2003). Prediction of epitopes in a given protein may reduce experimental efforts in determining epitopes needed for vaccine development and immunodiagnostics. Antibody defined epitopes are classified as linear (non-conformational) or discontinuous (conformational) and are primarily composed of a single stretch of the polypeptide chain. Discontinous (conformational) epitopes are composed of different parts of the polypeptide chain that are brought into close proximity by the folding of the protein. Continuous (linear) epitopes comprise of approximately 10% of all epitopes, where the linear peptide fragment of the epitope cross-reacts with the corresponding antibodies (Larsen *et al.,* 2006).

There are various methods for mapping linear and discontinuous epitopes. One approach to map linear B-cell epitopes is the use of synthetic peptides. The amino acid sequence of a protein can be readily predicted from nucleotide sequence data for the preparation of recombinant proteins or synthesis of peptides. Propensity scale methods are used for predicting linear B-cell epitopes. Based on studies of the epitopes physico-chemical properties these methods assign a propensity value to every amino acid (Larsen *et al.*, 2006). Similarly, in 1981 Hopp and Woods proposed a method for identifying epitopes based on calculating hydrophilicity residues and assumed that hydrophilic regions in the protein were predominantly located on the surface and thus potentially antigenic (Ponomarenko *et al*., 2009). Regions of the parent protein containing antigenic regions or epitopes can be predicted using antigenicity and hydrophilicity plots. Regions of hydrophilicity are likely to be surface exposed and therefore potentially antigenic. The Immune Epitope Database (IEDB) version 2.0

(www.ImmuneEpitope.org) provides a central database for submission of epitope information and computation tools developed for epitope prediction that is accessible to all researchers (Vaughan *et al.,* 2010). The IEDB contains data related to more than 1 900 structures and more than half of these structures have been identified as B- or T-cell epitopes defined in several host species (Vaughan *et al.,* 2010).

Epitope based vaccines provide a novel approach to prophylactic and therapeutic treatment of viral infections (Sollner *et al.,* 2008). In addition, polyvalent peptides based on significant epitopes may be a safe alternative to native antigens in diagnostic assays (Amexis *et al*., 2007; de Groot *et al*., 2003; Wu *et al*., 2003). There has been a recent increase in interest in mapping of epitopes and use of peptides. To enhance the sensitivity and specificity of diagnostic systems, peptide sequences that mimic immunodominant epitopes of the viral proteins have been used (Gómara *et al.,* 2000). The use of well-defined antigens such as peptides may circumvent difficulties in reproducing results and variations in the sensitivity of ELISA. The advantage of synthetic peptides compared with bacterially expressed recombinant proteins is that they eliminate non-specific reactions that result from the cross-reactivity of antibodies in *Escherichia coli* (*E.coli)*-derived recombinant products (González *et al*., 1997). Due to their weak binding to solid surfaces, the main disadvantage however, is that short synthetic peptides lacking hydrophobic side chains are poorly antigenic in solid-phase based immunoassays (Gómara *et al*., 2000). In addition, because the essential antigenic side chains are hidden when the peptide binds to the solid surface, synthetic peptides occasionally lose their antigenicity (Gómara *et al.,* 2000). Despite this, peptides have been successfully used in ELISA and as candidate vaccines. The effectiveness of the amount of synthetic peptide that is used to coat with is dependent on the hydrophobic interaction between the peptide and the plastic surface (Gómara *et al.,* 2000). The size and charge of the peptide also plays a role.

To identify epitopic regions, overlapping peptides of a defined length can be synthesized and reacted with antibody, monoclonal or polyclonal, in an ELISA format. Studies have documented B-cell epitopes of DENV-2 using overlapping synthetic peptides to analyse antisera (Wu *et al.,* 2003). After identifying peptides using overlapping regions it is possible to identify the exact region using shorter overlapping peptide sequences. Confirmation of the relevance of epitopes identified using polyclonal sera can be obtained relatively simply using antigen competition assays to exclude possible cross-reactivity or detection of antibodies against denatured antigen. An obvious shortcoming of this approach is the inability to detect conformational epitopes.

However linear epitopes can also be responsible for inducing neutralising antibody and have an important role in inducing immune response. Identifying viral protein targets of the immune response is important for development of novel vaccines and improved diagnostics. Subunit vaccines focusing on antigens that induce a protective immune response provide a safer approach to development of vaccines against diseases causing severe and frequently fatal haemorrhagic fevers.

1.12. Recombinant antigens

Recombinant antigens are useful tools for diagnosis and research. The use of recombinant DNA technology has a major advantage in that it is capable of producing specific therapeutic proteins in a heterologous expression system. A variety of different expression systems have been used for the production of therapeutic proteins for use in the prevention or treatment of human and animal diseases (Ahn *et al*., 2008). Recombinant DNA technology in mammalian, plant, insect, yeast and bacterial cells including transgenic organisms have been used to establish expression systems for therapeutic proteins (Ahn *et al*., 2008). These expression systems are available commercially for producing recombinant proteins depending on requirements, yield and cost.

Among the variety of available heterologous expression systems, yeasts are used to produce recombinant proteins that are not competently produced in *E.coli* because of problems with folding or glycosylation. Yeast expression systems are known to perform post-translational modifications, they are easier and less expensive than insect and mammalian expression systems and they are genetically well-characterized (Demain and Vaishnav, 2009; Verma *et al*., 1998). The main advantage of mammalian cell expression is that proteins are properly and efficiently recognized by mammalian cells and usually signal for synthesis, processing and secretion of eurkaryotic proteins (Verma *et al*., 1998). The need for erythropoietin (EPO) and tissue plasminogen activator (tPA) production in the early 1980s of the biopharmaceutical effort initiated the use of mammalian cell culture (Demain and Vaishnav, 2009). Proteins are often produced in a properly folded and glycosylated form (Demain and Vaishnav*,* 2009).

Insect cell expression systems as opposed to mammalian systems have emerged in the last few years as attractive choices for the expression of recombinant molecules (Verma *et al*., 1998). These expression systems have the best machinery for the folding of mammalian proteins, thus

suitable for producing soluble protein of mammalian origin (Agathos, 1991). Proteins are often glycosylated in insect cells and although the N-linked glycosylation sites are the same as in the mammalian systems, there appears to be a difference in the nature of the oligosaccharide chains (Verma *et al*., 1998). These cells are not capable of processing the mature oligosaccharides to the forms found in mammalian cells (Verma *et al*., 1998). Insect cells are easier to grow in shaker culture.

E.coli expression systems are widely used hosts for the expression of heterologous proteins (Terpe, 2006). *E.coli* is an attractive system because of its well-characterized genetics, its ability to grow rapidly and at high density on inexpensive substrates (Baneyx, 1999). This system is used for massive production of commercialized proteins and for functional expression of non-glycosylated proteins. Heterologous proteins produced in *E.coli* as inclusion bodies are usually inactive, aggregated and insoluble, possessing non-native intra- and inter-molecular disulfide bonds (Fischer *et al*., 1993). Active proteins can be obtained by the removal of inclusion bodies from the cell, solubilization of proteins using denaturants which unfold the proteins and the elimination of disulfide bonds using reducing agents. Refolding is achieved by removing the denaturant and reducing agent, followed by renaturing the protein (Demain and Vaishnav, 2009). High yields of protein are usually produced in the cytoplasm rather than in periplasmic space.

1.13. Problem identification, aims and objectives

YFV is a highly pathogenic virus that has re-emerged and has become a public health concern in epidemic regions. Despite the availability of a highly efficacious vaccine, there are still many outbreaks and the vaccine cannot be administered to immuno-competent individuals. Preparation of diagnostic reagents require biosafety level (BSL) three or BSL four facilities. In addition, diagnostic kits are not readily available commercially, which restricts many laboratories in their ability to rapidly diagnose YFV and to test the immune status of vaccinated individuals.

The serological cross-reactivity between *flaviviruses* is well known and does complicate the interpretation of serological tests. The development of a recombinant antigen that does not require BSL three facilities for preparation and is safe to use, would have an important role in a diagnostic laboratory for detecting antibodies in infected individuals and vaccinees. In addition it may be possible to prepare recombinant antigens or even peptides representing specific epitopes that can differentiate between different *flaviviruses*. Hence the identification of

immunodominant antigenic regions and linear B cell epitopes would be useful for the development of tools for detection and contribute to vaccine development.

The aim of this study was to identify immunodominant viral proteins that induced detectable antibody responses that could be used for developing diagnostic assays and to identify linear B cell epitopes on selected viral proteins.

The objectives of this study were as follows:

1. To express selected YFV C, EDIII and NS4a recombinant proteins using a bacterial expression system.
2. To determine the usefulness of these proteins for detection of specific antibody against YFV in ELISA.
3. To use epitope prediction software to predict epitopic regions on the C, EDIII and NS4a proteinsand use peptide libraries covering epitopic regions to identify possible linear B cell epitopes.

Chapter 2

CLONING AND OPTIMIZATION OF EXPRESSION OF YELLOW FEVER VIRUS EDIII, C AND NS4a PROTEINS

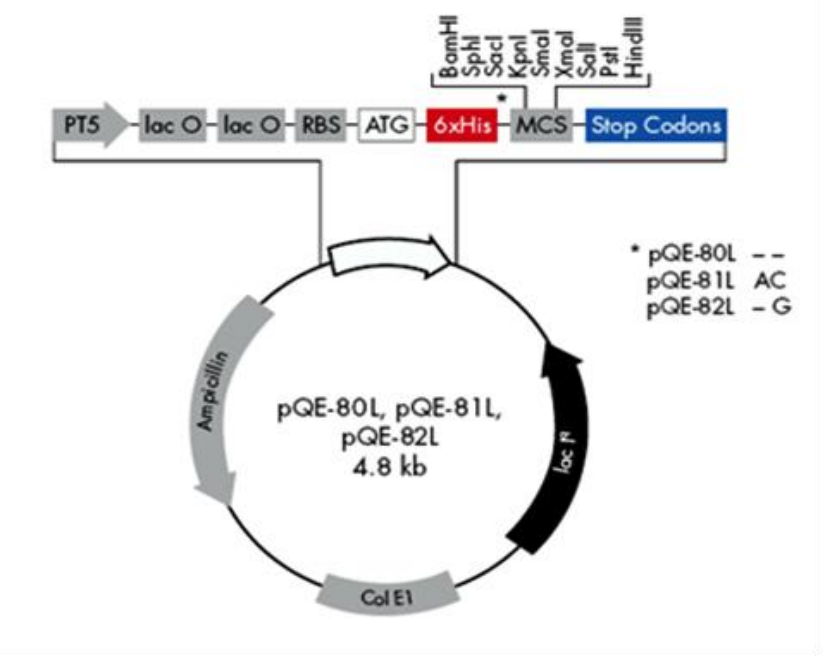
2.1. Introduction

Recombinant proteins are useful tools for research and diagnosis that can be readily produced in the laboratory. Briefly, the gene encoding the protein of interest is ligated into a suitable vector for expression. Proteins can be expressed in various host organisms including bacteria, insect cells, yeast and mammalian cells. Factors to consider when selecting the appropriate expression system for recombinant protein production include the quality and functionality of the protein and the production speed, cost and yield (Demain and Vaishnav, 2009). *E. coli* expression systems are widely used for the production of heterologous proteins (Terpe, 2006). They offer rapid growth and expression, high culture yields and are very cost effective. The *E. coli* system however does have a few disadvantages, such as high cell densities that can result in toxicity due to acetate formation and formation of inclusion bodies. There are various methods for the retrieval of proteins from inclusion bodies that require denaturing and refolding of the proteins. Baculovirus expression systems have the advantage that mammalian proteins are expressed with post-translational modifications, excluding glycosylation, but yield can be lower than a bacterial system and the process is more time-consuming and more challenging technically. Mammalian expression systems have low yield generally insufficient for preparing antigens for application in diagnostics and surveillance. To obtain recombinant yellow fever antigens in this study for application in ELISA to detect IgG antibody, two bacterial expression systems were investigated. The EDIII protein of YFV was selected to optimise the expression protocols and subsequently, attempts were made to express the YFV, C and NS4 proteins using similar protocols.

Bacterial expression: There are a vast number of bacterial expression plasmids available each with their specific properties and characteristics. To prepare recombinant YFV EDIII proteins for downstream applications, we selected two vectors, pQE-80L (Qiagen, Hilden, Germany) and pCold TF (Takara, Japan), and compared the yield and functional usefulness of the recombinant protein expressed from each construct.

The pQE-80L vector (illustrated in Figure 2.1) has a phage T5 promoter and two *lac* operator sequences that increase *lac* repressor binding and ensure efficient repression of the T5

promoter. The vector has a synthetic ribosomal binding site (RBSII) for high translation rates, a 6x His-tag coding sequence to facilitate protein purification, a multiple cloning site (MCS) and a ColE1 origin of replication. Translational stop codons are present in all reading frames, and to ensure the stability of the expression construct and prevent read-through transcription the vector has two transcriptional terminators t_0 (phage lambda) and T1 (*rrn*B operon from *E.coli*). The vector also codes for the β-lactamase gene (*bla*) that confers resistance to ampicillin. Recombinant protein expression using pQE-80L vectors is induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) which binds to the *lac* repressor protein and inactivates it. The host cell's RNA polymerase can transcribe the sequences downstream from the promoter once the *lac* repressor is inactivated. The resultant transcripts are then translated into recombinant proteins.



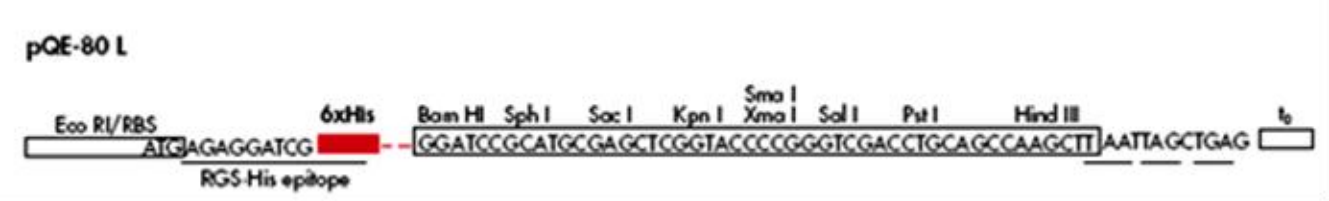


Figure 2.1. Vector map, promoter and multiple cloning site of pQE-80L, pQE-81L and pQE-82L (Qiagen, Hilden, Germany 2003).

The pCold™TF vector (illustrated in Figure 2.2) uses cold shock technology for expression of the TF chaperone as a soluble fusion tag to facilitate correct protein folding resulting in efficient soluble protein production. The system provides increased *in vivo* protein yield, purity and solubility for expressed recombinant proteins. TF is a ~52 kDa prokaryotic ribosome-associated chaperone that facilitates co-translational modifications. TF is expressed in large amounts in *E. coli* expression systems because of its *E. coli* origin. Expression occurs at low temperatures optimally 15ºC and the lower incubation temperature facilitates up-regulation of target protein production, slows down cell growth and suppresses the expression of non-specific cellular proteins. Compared to conventional *E. coli* expression systems this method should allow expression of target proteins at a higher yield and purity and increase solubility. To obtain increased amounts of soluble recombinant protein, co-expression of chaperone proteins during expression has proven to be effective. The pCold™ DNA TF vector has a cold-shock protein A (*csp*A) promoter, a MCS, a His-Tag sequence, and a translation enhancing element (TEE). Expression is regulated by a lac operator inserted downstream of the *csp*A promoter. Recognition sites for HRV 3C Protease, Thrombin and Factor Xa provide cleavage sites for removal of TF fusion protein from the expressed protein if required.

In a previous study, the YFV EDIII gene was cloned into pQE-80L and expressed in JM109 *E.coli* cells. Extremely low yields were obtained and were insufficient for use in ELISA hence in this study optimization of expression to obtain greater yield was investigated using alternative cell lines and an alternative expression plasmid and expression system. This chapter describes the methodology used to clone the YFV EDIII gene, and the NS4a and C genes, into different vectors and the various methods used to optimize and characterize expression of the protein.

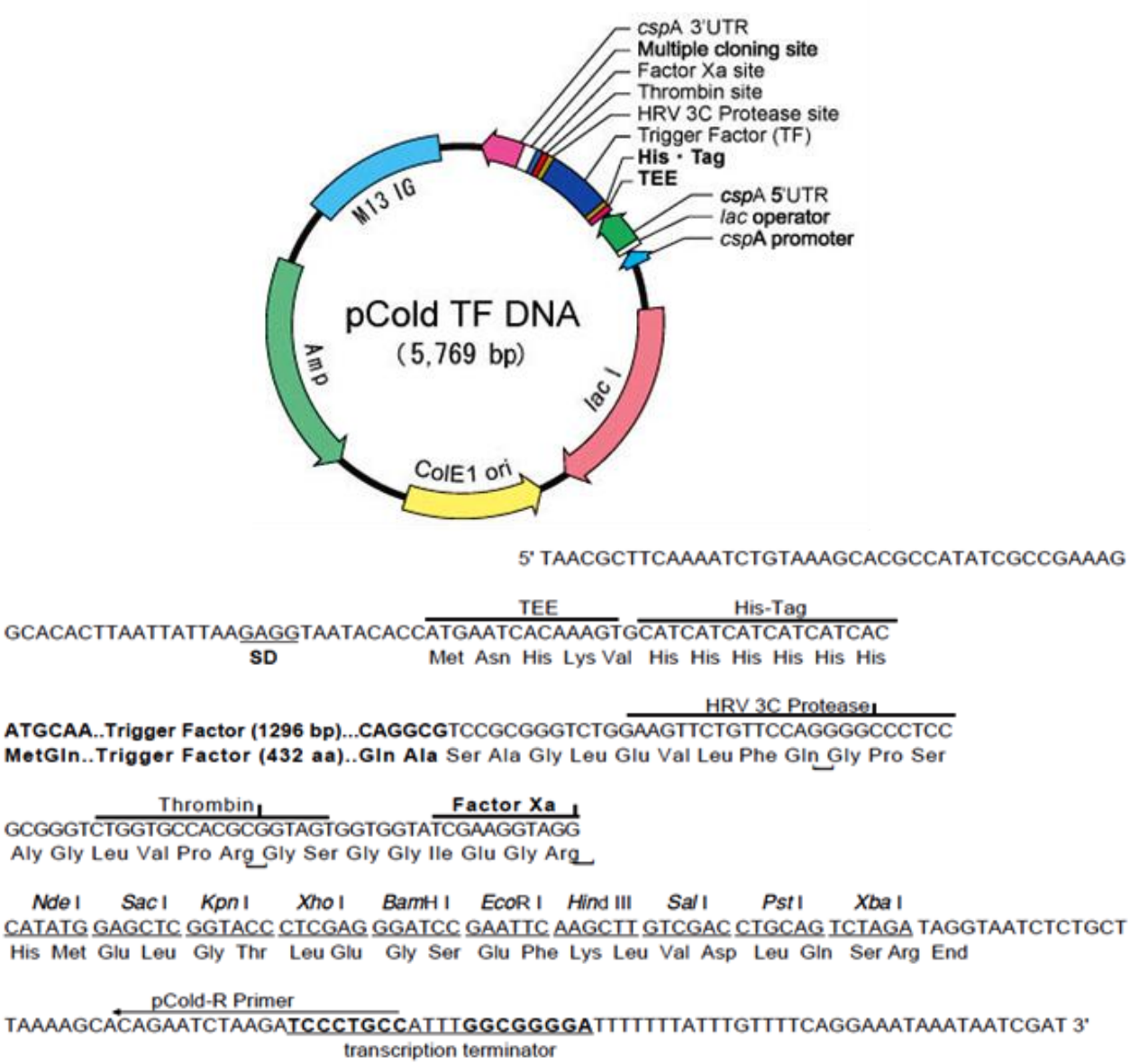


Figure 2.2. Vector map, promoter and multiple cloning site of the pCold™ TF plasmid (Takara, Japan).

2.2. Materials and methods

2.2.1. Amplification of genes encoding EDIII, C and NS4 proteins of YFV

2.2.1.1. RNA isolation

Stocks of YFV 17D vaccine strain were prepared as 10% mouse brain suspension. The strain used in this study was given to the Department of Medical Microbiology and Virology by Dr

Robert Tesh from the Arbovirus Reference Center in Texas. The YFV 17D vaccine strain was selected as it can be handled within a BSL 2 laboratory provided the researcher has previously been vaccinated.

Viral RNA was extracted from an aliquot of 10% mouse brain suspension in Eagle's Minimal Essential Media (EMEM) containing 1% penicillin and streptomycin with no other additives, using the QIAamp® Viral RNA Mini Spin Kit, according to the manufacturer's instructions. Briefly, a 560 µl aliquot of lysis Buffer AVL containing 5,6 µl of carrier RNA was added to a 1.5 ml microcentrifuge tube. A 140 µl aliquot of the mouse brain suspension was added to the Buffer AVL and carrier RNA and mixed by pulse-vortexing for 15 seconds. The sample was incubated at room temperature for 10 minutes and briefly centrifuged to remove droplets from the inside of the lid. A 560 µl aliquot of ethanol (96-100%) was added to the sample and mixed by pulse-vortexing for 15 seconds. A 630 µl aliquot of the solution was applied to the QIAamp Mini column and centrifuged for 1 minute at 8 000 x g. A 500 µl aliquot of wash Buffer AW1 was added to remove any residual contaminants and the minicolumn was centrifuged at 8 000 x g for 1 minute. A second wash was performed with the addition of 500 µl of wash Buffer AW2 and centrifuged at 14 000 x g for 3 minutes. An extra spin was performed at 14 000 x g for 1 minute to avoid any chance of Buffer AW2 carryover and another spin to remove ethanol that may affect downstream applications. RNA was eluted in a 50 µl aliquot of nuclease-free water and stored at -20°C.

2.2.1.2. Reverse transcriptase-polymerase chain reaction (RT-PCR) amplification of viral RNA

The Titan One Tube RT-PCR System (Roche, Mannheim, Germany) was used for the amplification of viral RNA, according to manufacturer's instructions. The system consists of AMV reverse transcriptase for complementary DNA (cDNA) synthesis and Expand High Fidelity enzyme blend for amplification of cDNA. Primers that had previously been identified in our laboratory for the amplification of the gene encoding the EDIII protein of YFV were modified to include *Bam*HI and *Hind*III sites at the 5' ends of both the forward and reverse respectively as illustrated in Table 2.1. Primers were also identified for the amplification of the genes encoding the C and NS4a of YFV and were modified to include *Bam*HI and *Hind*III sites as described above. Each primer was modified with "T" overhangs to facilitate digestion of 5' to 3' ends if required. The reaction mixtures were set up as described in Table 2.2. The RT-PCR reactions were cycled on a Perkin Elmer GeneAmp 9600 Thermocycler (Applied Biosystems, London,

England) and the following cycling conditions were used: 50°C for 30 minutes followed by a denaturation step at 94°C for 2 minutes, 30 cycles of 94°C for 10 seconds, annealing at 48°C for 30 seconds, elongation at 68°C for 45 seconds and a final extension step at 68°C for 7 minutes. The samples were stored at 4ºC.

Table 2.1. Oligonucleotide primers used to amplify the genes encoding the EDIII, NS4a and C proteins of YFV. Restriction sites *Bam*HI and *Hind*III were added to the forward and reverse primers respectively as indicated in bold font. The genomic positions are relative to the complete genomic sequence of the YFV 17D vaccine strain (GenBank accession numbers X03700 K02749).

Primer	Nucleotide sequence	Genomic position	Tm
Forward YFDEDIIIF	5'-TTT**GGATCC**AAGGGGACATCCTACAAAATATGC-3'	1849-1872	54ºC
Reverse YFDEDIIIR	5'-TTT**CTTCTC**TTTGTGCCACTGGTAAGTG-3'	2149-2128	55ºC
Forward YFCF	5'-TTT**GGATCC**TCTGGTCGTAAAGCTCAGG-3'	122-140	62.9ºC
Reverse YFCR	5'-TTT**AAGCTT**TCCACCCGTCATCAACAGCATT-3'	481-458	60.3ºC
Forward YFNS4AF	5'-TTT**GGATCC**GGAGCTGCTGAAGTGCTAGTTG-3'	6440-6461	64.3ºC
Reverse YFNS4AR	5'-TTT**AAGCTT**CCTTCTCTGTGCAAGCTTTGAC-3'	7300-7279	60.3ºC

Table 2.2. Reaction mix for One Step RT-PCR for the amplification of genes encoding YFV EDIII, C and NS4a proteins.

Reaction reagents (final concentration)	Volume (µl) per tube
Forward Primer (0.4 µM)	0.75 µl
Reverse Primer (0.4 µM)	0.75 µl
RNase Inhibitor (10U)	0.25 µl
Dithiothreitol (DTT) (5 mM)	2.5 µl
5 x RT-PCR buffer (7.5 mM $MgCl_2$ and Dimethyl sulfoxide (DMSO))	10 µl
Deoxynucleotides (dNTPs) (0.2 mM)	1 µl
Enzyme	1 µl
RNA Template (1 µg)	2 µl

Nuclease free water	31.75 µl
Final Volume	**50 µl**

2.2.1.3. Agarose gel electorophoresis

Amplicons were separated by electrophoresis using a 1% agarose gel prepared in 1 x Tris acetate EDTA (TAE) buffer containing ethidium bromide (20 µg/ml). Gel electrophoresis was performed using a Bio-Rad PowerPac Basic system (Bio-Rad, California, United States of America) at 100 volts (V) for 40 minutes. The size of the amplicons was determined by comparison with the O'Generuler^{TM} DNA ladder mix (Thermo Scientific, Massachusetts, United States of America) (100 bp–10 000 bp) or O'Generuler^{TM} 100 bp DNA ladder (Thermo Scientific, Massachusetts, United States of America) (100 bp–1000 bp). A 5 x gel loading buffer was used (10 mM Tris-HCl, pH 7.6, 60 mM EDTA, 0.03% bromophenol blue and 60% glycerol in water) for loading DNA samples.

2.2.1.4. Purification of PCR product

The Wizard® SV Gel and PCR Clean-Up System (Promega, Wisconsin, United States of America) was used for the purification of amplicons, according to the manufacturer's instructions. The system allows the binding of DNA to silica membranes in the presence of chaotropic salts. The gene of interest was excised from the agarose gel to remove non-specific bands and dissolved by adding a ratio of 10 µl of membrane binding solution per 10 mg of agarose gel slice. The mixture was vortexed and incubated at 65ºC until the gel slice was completely dissolved. The dissolved gel mixture was transferred to the SV Minicolumn assembly and incubated at room temperature for 1 minute. The SV Minicolumn assembly was centrifuged in a 1 minute at 14 000 x g. The SV Minicolumn was removed from the spin column assembly and the liquid in the collection tube was discarded. The column was washed by adding 700 µl of membrane wash solution to the SV Minicolumn assembly and centrifuged for 1 minute at 14 000 x g. This step was repeated by adding 500 µl membrane wash solution and centrifuged for 5 minutes at 14 000 x g. The column assembly was centrifuged with the microcentrifuge lid open for 1 minute to allow evaporation of any residual ethanol. The SV Minicolumn was carefully transferred to a clean 1.5 ml microcentrifuge tube, 30 µl of nuclease-free water was added and the columns allowed to stand at room temperature for 1 minute and centrifuged for 1 minute at 14 000 x g. The eluted DNA was stored at – 20 ºC. DNA

concentrations were measured using a NANODROP spectrophotometer at an optical density of 260 nm and the DNA purity was measured using the 260/280 nm ratio.

2.2.1.5. Preparation of Luria Bertani (LB) media and agar

LB media was prepared as follows: 10 g Bacto®-tryptone, 5 g Bacto®-yeast extract and 5 g NaCl to 1 litre of distilled water. The media was sterilized by autoclaving at 121°C for 30 minutes and allowed to stand until the temperature reached 50ºC before adding ampicillin to a final concentration of 100 µg/ml. Media was stored at 4ºC.

LB/ampicillin plates were prepared by addition of 15 g agar to 1 liter of LB broth. The media was sterilized by autoclaving at 121°C for 30 minutes and allowed to stand until the temperature reached 50ºC before adding ampicillin to a final concentration of 100 µg/ml. Plates were prepared by pouring warm media (30 – 35 ml) into 85 mm petri dishes and allowed to stand until the agar hardened. The agar plates were stored at 4ºC.

LB/ampicillin/IPTG/5-bromo-4-chloro-3-indolyl-[beta]-D-galactosidase (X-gal) plates were prepared by adding a 100 µl aliquot of 100 mM IPTG (Thermo Scientific, Massachusetts, United States of America) and 40 µl of 50 mg/ml X-gal (Thermo Scientific, Massachusetts, United States of America) spread over the surface of an LB/ampicillin plate and the liquid allowed to adsorb for 30 minutes at room temperature before use.

2.2.2. Cloning of EDIII, C and NS4a genes in pGEM®-T Easy

The genes encoding YFV EDIII, C and NS4a proteins were each cloned into the pGEM®-T Easy (Promega, Wisconsin, United States of America) vector according to manufacturer's instructions. The vector provides a useful system for cloning and sequencing PCR products. Genes can subsequently be rescued from the plasmids using restriction enzyme digestion and cloned into expression plasmids using the specific restriction sites. The pGEM®-T Easy vector has single 3'-T overhangs at the insertion site to facilitate TA cloning of amplicons with overhanging 3' adenines and prevents re-circularization of the vector. The pGEM®-T Easy Vector contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase as shown in Figure 2.3. The

α-peptide allows recombinant clones to be directly identified by colour screening on indicator plates after insertional inactivation. For ligation a 3:1 (insert:vector) molar ratio was used.

$$\frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}} \times \text{Insert:vector molar ratio} = \text{ng of insert}$$

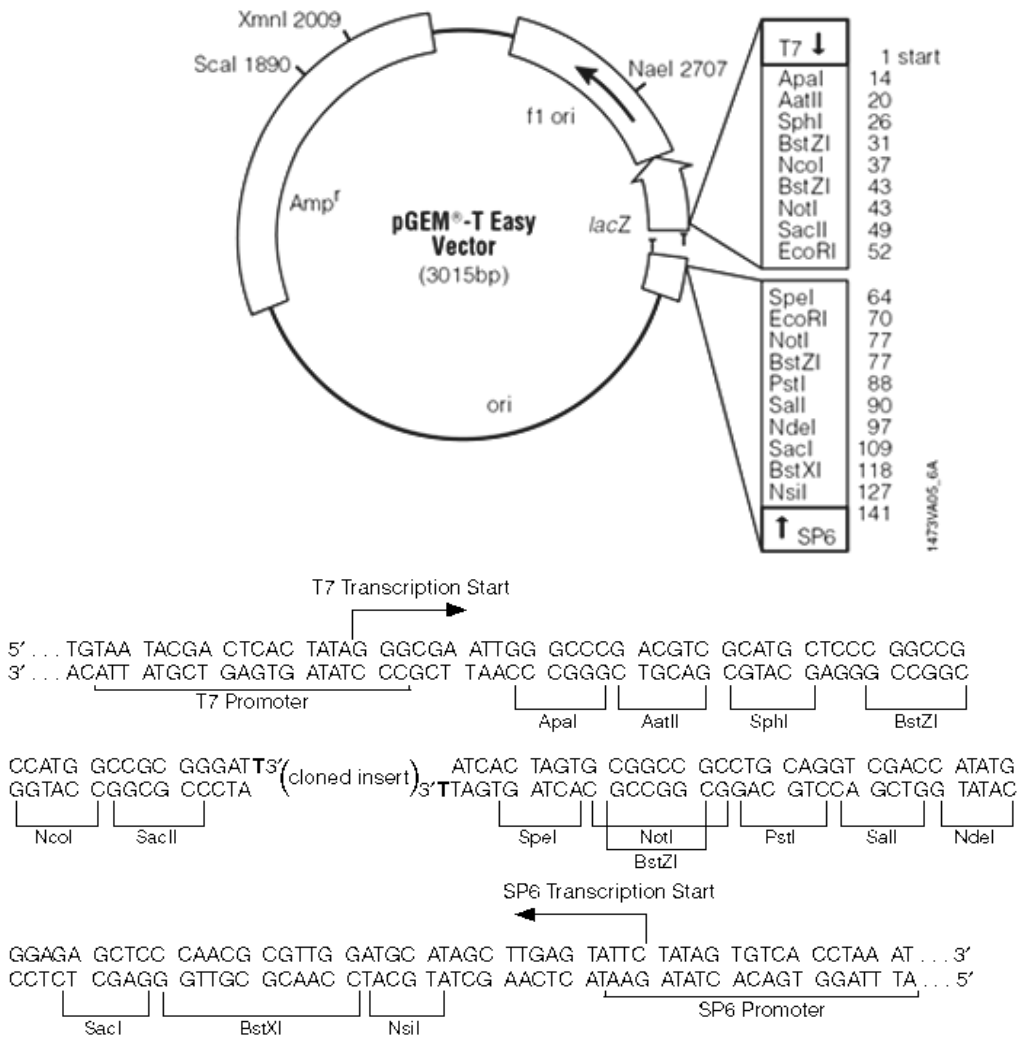


Figure 2.3. Vector map, promoter and multiple cloning site of pGEM®-T Easy (Promega, Wisconsin, United States of America)

Briefly, ligation reactions were carried out according to manufacturer's instructions as illustrated in Table 2.3.

Table 2.3. Ligation reaction mixture used for ligation of genes encoding EDIII, C and NS4a proteins and pGEM®-T Easy vector.

Reaction reagents	Reaction volume
10 x T4 DNA Ligase Reaction Buffer	1 µl
pGEM®-T Easy Vector (50 ng)	1 µl
YFV PCR product*	5 µl
T4 DNA Ligase	1 µl
Deionized water	2 µl
Final Volume	10 µl

* Total amount (ng) of DNA in ligation reaction: EDIII 130 ng, C 130 ng, NS4a 172 ng

Ligation reactions (as described in Table 2.3) were incubated for 1 hour at room temperature. Chemically competent JM109 cells (Promega, Wisconsin, United States of America) with a transformation efficiency of 1 x 10^8 cell forming units (cfu)/µg DNA were transformed using 2 µl of the ligation reaction. The cells were mixed gently and incubated on ice for 20 minutes. The cells were heat-shocked for 45 seconds in a water bath at 42ºC and transferred to ice for 2 minutes. The heat-shock method facilitates the opening of the pores of the cell membrane and allows the entry of plasmid DNA. A 900 µl aliquot of Super optimal catabolite-repression (SOC) media (Super optimal broth (SOB) media with the addition of catabolite-repression which indicates that glucose has been added to the media) was added to the transformation culture and incubated at 37ºC for 1.5 hours with shaking at 200 rpm. A 200 µl aliquot was plated out onto LB/ampicillin/IPTG/X-gal agar plates. The remaining transformation culture was centrifuged at 14 000 x g, the pellet was re-suspended in 100 µl SOC media and plated out onto LB/ampicillin plates. The plates were incubated overnight at 37ºC, blue/white colony selection was performed and the selected transformants were grown overnight and purified. The MCS is located in the *lacZ* gene and successful cloning of the genes into pGEM®-T Easy should disrupt the *lacZ* gene. Beta-galactosidase converts the colourless substrate X-gal to produce blue colonies. Disruption of the gene results in white colonies. Hence white colonies are likely the result of positive transformation. It is necessary to confirm positive transformants as cloning may result in blue colonies when the insert is a multiple of three bases and if there are no stop codons in the gene inserted.

2.2.2.1. Plasmid DNA purification

To identify positive transformants, white colonies were selected from LB/ampicillin/IPTG/X-gal plates and each colony was grown overnight in a 5 ml volume of LB/ampicillin at 37°C with shaking at 200rpm. Plasmid was purified from overnight cultures using the PureYield™ Plasmid Miniprep System (Promega, Wisconsin, United States of America) according to the manufacturer's instructions. The Miniprep System can be used to purify 1.5 – 7.5 µg of plasmid DNA which is purified using a silica-membrane column. The protocol includes an endotoxin removal wash to remove protein, RNA and endotoxin contaminants. Briefly, a 1.5 ml aliquot of bacterial culture grown in LB/ampicillin media was transferred to a 1.5 ml microcentrifuge tube. The culture was centrifuged at 14 000 x g in a microcentrifuge for 1 minute. The supernatant was discarded and the step was repeated to obtain a higher yield of plasmid DNA. The cell pellet was re-suspended in 600 µl of ultra pure water. A 100 µl aliquot of cell lysis buffer (contains sodium hydroxide, NaOH) was added and mixed by inverting the tube 6 times. The NaOH assists in rupturing the cells and denaturing the bacterial genomic DNA into single strands. A 350 µl aliquot of cold neutralisation solution was added and mixed thoroughly by inverting the tube. The sample was centrifuged at 14 000 x g for 3 minutes. The supernatant was transferred into a PureYield™ Minicolumn, without disturbing the cell debris pellet and centrifuged at 14 000 x g. A 200 µl aliquot of endotoxin removal wash (containing chaotropic agents) was added to the minicolumn and centrifuged at 14 000 x g for 1 minute. A 400 µl aliquot of column wash solution was added to the minicolumn and centrifuged at 14 000 x g for 1 minute. The column was placed in a PureYield™ collection tube and centrifuged for 1 minute at 14 000 x g in a microcentrifuge. The minicolumn was transferred to a clean 1.5 ml microcentrifuge tube and 30 µl of nuclease free water was added to the minicolumn matrix and allowed to stand at room temperature for 1 minute. The microcentrifuge tube was centrifuged for 1 minute at 14 000 x g to elute the plasmid DNA. The eluted plasmid DNA was stored at – 20 ºC.

2.2.2.2. Identification of positive transformants by restriction enzyme digestion

Restriction enzyme digestions of purified plasmids for identification of positive transformants were performed at 37ºC for 2 hours in a water bath. *Bam*HI (Promega, Wisconsin, United States of America) and *Hind*III (Promega, Wisconsin, United States of America) restriction enzymes were used for digesting plasmid DNA. All restriction enzyme digestions were performed in 20 µl reactions as described in Table 2.4. Reaction mixtures were analyzed by gel

electrophoresis. Aliquots of digested and undigested plasmid DNA were separated on a 1% agarose gel containing 1 x TAE and ethidium bromide as described in Section 2.2.1.3. The size of each band was determined by comparison with the O'Generuler^{TM} DNA ladder mix (100 bp–10 000 bp) or O'Generuler^{TM} 100 bp DNA ladder (100 bp–1000 bp). A 5 x gel loading buffer was used for loading the samples.

Table 2.4. Restriction enzyme reaction mix for identification of positive transformants.

Reaction components	Reaction volume
*Bam*HI 10U/µl	1 µl
*Hind*III 10U/µl	1 µl
10 x Buffer E	2 µl
Plasmid DNA (2-5 µg)	5 µl
Nuclease free water	11 µl
Final Volume	**20 µl**

2.2.3. Cloning of EDIII, C and NS4 genes into the pQE-80L bacterial expression vector

2.2.3.1. Preparation of chemically competent cells for protein expression

For expression experiments chemically competent cells were prepared using the calcium chloride method. Sterile 15 ml centrifuge tubes, 1.5 ml microcentrifuge tubes and tips were stored at 4ºC for at least 24 hours. Briefly, a 5 ml overnight culture of the OverExpress *E.coli* BL21 (DE3) strain (OverExpress) (Lucigen® Corporation, Wisconsin, United States of America) was grown in LB media in a 50 ml centrifuge tube. The culture was grown at 37ºC with shaking at 200 rpm. A 250 µl aliquot of the overnight culture was used to inoculate 25 ml of LB media. The culture was grown at 37ºC with shaking until reaching an absorbance at 600 nm of between 0.2-0.4. The culture was divided equally between two 15 ml centrifuge tubes and cells were collected by centrifugation at 4000 x g for 10 minutes. The cells were kept on ice in all subsequent steps. The cells were gently re-suspended in half the culture volume (12.5 ml) of ice-cold 0.1 M $CaCl_2$ and the 12.5 ml volume was split between two 15 ml centrifuge tubes. The $CaCl_2$ solution was prepared fresh and sterile filtered. The cells were kept on ice at 4ºC for 1 hour and cells were collected by centrifugation at 4000 x g for 10 minutes at 4ºC. The cells

were gently re-suspended on ice in a 1/10 culture volume (2.5 ml) of ice-cold 0.1 M $CaCl_2$. The cells from both tubes were combined and kept on ice at 4ºC for 1 hour. A 375 µl aliquot of ice-cold sterile glycerol was added to the cells to make a final concentration of 15% v/v. The cells were kept on ice at 4ºC for 30 minutes, aliquoted into cold cryotubes and stored at -70ºC. The transformation efficiency of the chemically competent cells was determined by transforming the cells with uncut plasmid and calculating cfu/µg DNA. Briefly, 100 µl aliquot of cells was transformed with 0.1 ng uncut pUC19 plasmid DNA. The cells were mixed gently and incubated on ice for 20 minutes. The cells were heat-shocked for 45 seconds in a water bath at 42ºC and directly transferred to ice for 2 minutes. A 900 µl aliquot of SOC media was added to the ligation reaction and incubated at 37ºC for 1.5 hours with shaking at 200 rpm. A 200 µl aliquot was plated out onto LB/ampicillin agar plates. The remaining transformation culture was centrifuged at 14 000 x g and the pellet was re-suspended in 100 µl SOC media and plated out. The plates were incubated overnight at 37ºC and white colonies indicated positive transformants. The calculation used to determine the transformation efficiency in cfu/ µg DNA is as follows: $\frac{\text{Number of colonies on a plate (cfu)}}{\text{ng DNA}}$ X 1000 cfu/µg DNA

2.2.3.2. Transformation of OverExpress cells

The YFV EDIII, C and NS4a genes were rescued from the pGEM®-T Easy vector using *Bam*HI and *Hind*III restriction enzymes and subsequently cloned into the pQE-80L vector. Briefly, the pGEM®-T Easy constructs containing the YFV EDIII, C and NS4a genes (designated pGEM-YFVEDIII, pGEM-YFVC and pGEM-YFVNS4a respectively) were grown overnight in 5 ml LB/ampicillin media at 37ºC with shaking at 200 rpm. Plasmid was purified using the PureYield™ Plasmid Miniprep System and digested with *Bam*HI and *Hind*III and the digestion products separated by electrophoresis. After separation on agarose gel, the bands representing the YFV EDIII, C and NS4a genes were excised and purified from the gel using the Wizard® SV Gel and PCR Clean-Up System. The pQE-80L expression plasmid was grown overnight in 5 ml LB/ampicillin media at 37ºC with shaking at 200 rpm and purified. Plasmid DNA was purified and digested with *Bam*HI and *Hind*III restriction enzymes to linearize the plasmid and produce vector DNA with *Bam*HI and *Hind*III compatible 5' and 3' cloning ends respectively (Table 2.5). Digestion products were separated and visualized on a 1% agarose gel. The linearized pQE-80L plasmid DNA was excised and purified from the gel and the DNA concentration measured.

The expression constructs were created by ligating the EDIII, C and NS4a genes into the linearized pQE-80L plasmid. Ligation reactions were prepared as described in Table 2.6 using insert: vector Molar ratios of approximately 3:1.

Table 2.5. Restriction enzyme digestion of pQE-80L plasmid.

Reaction components	**Reaction volume**
*Bam*HI 10U/µl	1 µl
*Hind*III 10U/µl	1 µl
10 x Buffer E	2 µl
pQE-80L plasmid DNA (2-5 µg)	5 µl
Nuclease free water	11 µl
Final Volume	**20 µl**

Table 2.6. Ligation reaction mixture used for ligation of genes encoding EDIII, C and NS4a proteins and pQE-80L plasmid.

Reaction reagents	**Reaction volume**
10 x T4 DNA Ligase Reaction Buffer	1 µl
pQE-80L DNA (14-50 ng/µl)	1 µl
YFV DNA *	6 µl
T4 DNA Ligase	1 µl
Deionized water	1 µl
Final Volume	**10 µl**

***Total amount (ng) of DNA in ligation reaction: EDIII 60 ng, C 42.16 ng, NS4a 124 ng

The ligation reaction mixtures were incubated at 4ºC overnight. Chemically competent OverExpress cells with a transformation efficiency of 3 x 10^7 cfu/µg were transformed with the ligation reaction using the heat-shock method. Briefly, a total of 2 µl of the ligation reaction mixture was added to a 100 µl aliquot of chemically competent cells. The cells were mixed

gently and incubated on ice for 20 minutes. The cells were heat-shocked for 45 seconds in a water bath at 42ºC and directly transferred to ice for 2 minutes. A 900 µl aliquot of SOC media was added to the transformation culture and incubated at 37ºC for 1.5 hours with shaking at 200 rpm. A 200 µl aliquot was plated out onto LB/ampicillin agar plates. The remaining transformation culture was centrifuged at 14 000 x g and the pellet was re-suspended in 100 µl SOC media. The plates were incubated overnight at 37ºC and white colonies indicated positive transformants. For each ligation reaction, three colonies were selected for analysis and grown overnight in 5 ml LB/ampicillin media at 37ºC with shaking. Plasmid DNA was purified from overnight cultures as described in Materials and Methods 2.2.2.1. pQE-80L plasmid DNA was prepared similarly as a negative control for expression of mock antigen for use in downstream applications.

2.2.3.3. Identification of positive transformants

Positive transformants were identified by performing a PCR using DNA from selected colonies as template and specific primers to confirm the insertion of the gene of interest. In addition, positive transformants were confirmed by digesting the plasmid with *Bam*HI and *Hind*III to determine if the plasmids contained a gene of expected size. GoTaq ® DNA Polymerase (Promega, Wisconsin, United States of America) was used for the PCR amplifications according to the manufacturer's instructions. The kit contains GoTaq® DNA Polymerase and 5 x Green GoTaq® Reaction Buffer. Briefly, the template was prepared by inoculating 25 µl of nuclease free water with a single colony selected from an LB/ampicillin plate. A fraction of the colony was inoculated into LB/amp broth for further propagation. The sample was boiled for 5 minutes at 95ºC causing the cell wall to break. Samples were centrifuged for 30 seconds at 5000 x g and 25 µl was used as template. The YFV primers described in Table 2.1 were used for amplification.

Table 2.7. Colony PCR for the detection of YFV EDIII, C and NS4a genes

Final concentration of reaction components	Reaction volume
Forward Primer (0.4 µM)	0.75 µl
Reverse Primer (0.4 µM)	0.75 µl

5 x Green GoTaq® Reaction Buffer	10 µl
dNTPs (0.2 mM)	1 µl
GoTaq DNA Polymerase (5u/µl)	1 µl
Template	25 µl
Nuclease free water	11.5 µl
Final Volume	**50 µl**

The reaction mixtures were set up as described in Table 2.7. The following cycling conditions were used: denature at 95ºC for 2 minutes, 30 cycles of 95ºC for 1 minute, 45ºC for 30 seconds, 72ºC for 1 minute, a final elongation at 72ºC for 5 minutes. PCR products were separated by gel electrophoresis on a 1% agarose gel. In addition, positive transformants were confirmed using restriction enzyme digestions performed at 37ºC for 2 hours in a water bath. The plasmid DNA was digested with *Bam*HI and *Hind*III restriction enzymes. All restriction enzyme digestions were performed in 20 µl reactions as described in Table 2.8. Reaction mixtures were separated on a 1% agarose gel.

Table 2.8. Restriction enzyme analysis for positive transformant identification.

Reaction components	**Reaction volume**
*Bam*HI 10U/µl	1 µl
*Hind*III 10U/µl	1 µl
10 x Buffer E	2 µl
Plasmid DNA (2-5 µg)	5 µl
Nuclease free water	11 µl
Final Volume	**20 µl**

2.2.3.4. Sequencing of plasmid DNA

To confirm that the genes were cloned in frame and that no mutations were introduced during amplification of the gene, DNA sequences were determined using the BigDye® Terminator version 3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems, London, England) and

analysed on the GeneAmp® PCR system 9700 (Applied Biosystems, London, England). All sequencing reactions were performed in duplicate. Cycle sequencing reactions were setup as follows: 4 µl of 0.8 pmol/µl of primer pQE forward 5'-CCCGAAAAGTGCCACCTG-3', 2 µl 5 x sequencing buffer, 2 µl ready reaction mix, 2 µl (3-10 ng) plasmid DNA template. Twenty-five cycles were performed at 96ºC for 1 minute, 96ºC for 10 seconds, 50ºC for 5 seconds, 60ºC for 4 minutes.

The sequencing reaction was cleaned up using the EDTA/Ethanol precipitation protocol according to the BigDye® Terminator version 3.1 Ready Reaction Cycle Sequencing Kit. Briefly, the sequencing reaction volume was adjusted to 20 µl with nuclease free water and added to a 5 µl aliquot of EDTA (125 mM) in a 60 µl aliquot of ethanol and incubated at room temperature for 15 minutes to precipitate. After centrifugation at 14 000 x g for 20 minutes the supernatant was aspirated and the pellet was washed in 500 µl of 70% ethanol, centrifuged for 10 minutes at 14 000 x g and dried at 37ºC. Electrophoresis was performed at the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State on the 3130xl Genetic Analyser HITACHI (Applied Biosystems, London, England). Sequence editing analysis and alignments were performed using the ChromasPro (http://www.technelysium.com.au/ChromasPro.html) version 1.5, Clustal X version 2.1 (http://www.clustal.org/clustal2/) and BioEdit Sequence Alignment Editor version 7.0.9.1 (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

2.2.4. Optimization of expression using pQE-80L constructs

Various expression experiments were performed to optimize expression of recombinant EDIII using the plasmid designated pQE-80L-YFVEDIII. A 5 ml (LB/ampicillin media) overnight starter culture was inoculated with a positive colony or a 10 µl aliquot of glycerol stock and incubated with shaking at 200 rpm at 37ºC. An additional 5 ml overnight starter culture was inoculated with a 10 µl aliquot of glycerol stock from pQE-80L with no insert which was used a negative control. An induction study was performed by diluting the starter culture 1:20 (2 ml in 38 ml) in pre-warmed LB/ampicillin media and grown at 37ºC with shaking until the culture reached an absorbance at 600 nm of between 0.6 - 0.8. Protein expression was induced by adding different concentrations of IPTG. IPTG was added to give a final concentration of 0.125 mM, 1 mM and 2.5 mM and the cultures were incubated at 25ºC or 37ºC. Samples were collected at the following intervals after induction T=0, 1, 2, 3, 4, 5 and 16 hours. Optical densities (OD) were

determined at each interval, samples were centrifuged at 13 000 x g and re-suspended in a calculated volume of phosphate buffered saline (PBS), pH 7.4 (Volume of PBS = A_{600nm} /0.5 x 150).

In addition, to prepare recombinant pQE-80L-YFVC and pQE-80L-YFVNS4a antigens, OverExpress cells were transformed with the pQE-80L plasmid DNA containing the C and NS4a genes. A 5 ml overnight starter culture was prepared in LB/ampicillin media at 37ºC with shaking at 200 rpm. An induction study was performed by diluting the starter culture 1:20 in pre-warmed LB/ampicillin media and grown at 37ºC until the culture reached an absorbance at 600 nm of between 0.6 – 0.8. Protein expression was induced by adding different concentrations of IPTG. IPTG was added to give a final concentration of 0.05 mM and 1 mM and the cultures were incubated at 30ºC or 37ºC. Protein expression was induced for 16 hours with a final concentration of 1 mM IPTG at 37ºC with shaking at 200 rpm. Samples were collected at T=0, 1, 2, 3, 4, 5 and 16 hours. OD were determined at each interval, samples were centrifuged at 13 000 x g and re-suspended in a calculated volume of PBS as shown above.

To determine in which phase the protein was present, a protein solubility study was performed. Briefly, a 1 ml sample of the culture was collected at T=4 hours, T=5 hours and T=16 hours after induction. The samples were centrifuged at 13000 x g and the pellets were re-suspended in 500 µl cold PBS. Sarcosyl, a mild detergent that releases membrane bound protein was added to the samples at a final concentration of 7.5%. The cells were sonicated using a Branson 220 ultrasonic cleaner (SmithKline Company, USA) using 10 x 15 second bursts with a 15 second cooling period on ice between each burst. An 80 µl aliquot of the cell suspension was collected to represent the total protein (soluble and insoluble fraction) and added to 20 µl of 5 x protein loading buffer (0.313 M Tris-HCl, pH 6.8, 10% SDS, 0.05% Bromophenol blue, 50% glycerol) and 1 µl of a 2 M solution DTT (Thermo Scientific, Massachusetts, United States of America). The remaining cell suspension was centrifuged at 13000 x g for 20 minutes. The supernatant was collected (soluble fraction) and added to 20 µl of the loading buffer. The pellet (insoluble fraction) was re-suspended in 500 µl cold PBS, an 80 µl aliquot was collected and added to 20 µl loading buffer. All the samples were heated for 5 minutes at 95ºC and separated by electrophoresis on a sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) gel (8% resolving, 4% stacking).

2.2.5. Cloning of EDIII gene into the pCold TF DNA bacterial expression vector

The YFV EDIII gene was digested from the pGEM®-T Easy vector using *Bam*HI and *Hind*III restriction enzymes and subsequently cloned into the pCold TF™ (Takara, Japan) plasmid as described in Materials and Methods 2.2.3.2. The pCold™ TF expression plasmid was linearized for ligation reactions from an overnight culture as follows: 5 ml LB/ampicillin media was inoculated with pCold™ TF (from glycerol stocks) and incubated overnight at 37ºC with shaking at 200 rpm. pCold™ TF DNA plasmid was purified using the PureYield™ Plasmid Miniprep system and digested using *Bam*HI and *Hind*III restriction enzymes as shown in Table 2.9 to obtain 5' and 3' ends compatible for cloning.

Table 2.9. Restriction enzyme digestion of pCold™ TF DNA.

Reaction components	Reaction volume
*Bam*HI 10U/µl	1 µl
*Hind*III 10U/µl	1 µl
10 x Buffer E	2 µl
pCold™ TF plasmid (2-5 µg)	5 µl
Nuclease free water	11 µl
Final Volume	**20 µl**

Table 2.10. Ligation reaction mixture used for ligation of gene encoding EDIII protein and linearized pCold™ TF DNA vector.

Reaction components	Reaction volume
10 x T4 DNA Ligase Reaction Buffer	1 µl
pCold TF DNA (1 µg – 10 µg)	2 µl
YFV EDIII DNA (5 ug – 15 ug)	6 µl
T4 DNA Ligase	1 µl
Deionized water	_
Final Volume	**10 µl**

Positive transformants were identified by performing a PCR using plasmid DNA as template and YFV EDIII primers to confirm the presence of the gene of interest. In addition, positive transformants were confirmed by restriction digestion analysis as described in Materials and Methods 2.2.3.3. Sequences were determined using the BigDye® Terminator version 3.1 Ready Reaction Cycle Sequencing Kit and analysed on the GeneAmp® PCR system 9700. All sequencing reactions were performed in duplicate. Cycle sequencing reactions were setup as described in Materials and Methods 2.2.3.4. The reaction mix included a 4 µl aliquot of (0.8 pmol/µl) primer pCold F1 5'-CCACTTTCAACGAGCTGATG-3', 2 µl 5 x sequencing buffer, 2 µl ready reaction mix, 2 µl (3-10 ng) plasmid DNA template.

2.2.6. Induction and optimization of expression using pCold™ TF DNA construct

A colony from OverExpress cells positively transformed with the pCold™ TF DNA plasmid containing the EDIII gene was selected for propagation. A 5 ml overnight starter culture was prepared in LB/ampicillin media at 37ºC with shaking at 200 rpm. An additional 5 ml LB/ampicillin starter culture was inoculated with 10 µl glycerol stock of pCold TF with no gene insert and used as a negative control. An induction study was performed by diluting the starter culture 1:20 (2 ml in 38 ml LB/ampicillin media) in pre-warmed LB/ampicillin media and grown at 37ºC until the culture reached an absorbance at 600 nm of between 0.4 – 0.5. The bacterial culture was allowed to stand for 30 minutes at 16ºC. Protein expression was induced for 24 hours with a final concentration of 1 mM IPTG at 16ºC with shaking at 200 rpm. Samples were collected at T=0, 4, 16 and 24 hour intervals. OD values were determined at each interval, samples were centrifuged at 13 000 x g and re-suspended in a calculated volume of PBS, pH 7.4 (Volume of PBS = A_{600nm} /0.5 x 150). To determine in which phase the protein was present, a protein solubility study was performed as described in Materials and Methods 2.2.4.

2.2.7. SDS-PAGE

For analysis of protein samples by SDS-PAGE, a modified method as described by Laemmli (1970) was used. The resolving gel was prepared using 30% acrylamide solution/0.8% bisacrylamide stock solution (Merck, New Jersey, United States of America) as shown in Table 2.11. The resolving gel was transferred to the pre-assembled electrophoresis apparatus (Bio-Rad, California, United States of America) and a layer of amyl alcohol was added to the top before polymerization. Once the resolving gel solidified, the amyl alcohol was discarded. The

stacking gel was prepared as illustrated in Table 2.12. The stacking gel was poured on top of the resolving gel and the comb was inserted before polymerization at room temperature.

Table 2.11. Preparation of an 8% resolving gel for SDS-PAGE analysis.

Reaction component	Reaction volume
30% acrylamide solution/0.8% bisacrylamide stock solution	21.4 ml
1M Tris-HCl pH 8.8 stock solution	30 ml
10% SDS	0.8 ml
1.5% ammonium persulfate (freshly prepared)	4 ml
TEMED	0.02 ml
Distilled water	23.8 ml
Total	**80 ml**

Table 2.12. Preparation of a 4% stacking gel for SDS-PAGE analysis.

Reaction component	Reaction volume
30% acrylamide solution/0.8% bisacrylamide stock solution	2 ml
1M Tris-HCl pH 6.8 stock solution	1.9 ml
10% SDS	0.15 ml
Glycerol	1 ml
1.5% ammonium persulfate (freshly prepared)	0.7 ml
TEMED	0.02 ml
Distilled water	9.25 ml
Total	**15 ml**

The Mini-PROTEAN® Tetra cell electrophoresis apparatus was filled with 1 x Tris-glycine-SDS electrophoresis buffer (25 mMTris, 192 mM glycine and 0.1% SDS, pH 8.3) (Bio-Rad, California, United States of America). The protein samples were prepared as follows: 80 µl of each sample was re-suspended in 20 µl 5 x protein loading buffer. The samples were denatured at

95ºC for 5 minutes and 25 µl aliquots of each sample were loaded into each well. The size of the protein bands were estimated using the Pageruler Prestained protein marker (10 kDa – 170 kDa). All protein samples were separated at 175 V for 90 minutes. SDS-PAGE gels were stained overnight with Coomassie Brilliant blue stain (45% methanol, 10% glacial acetic acid and 0.2% Coomassie brilliant blue). The gels were destained with approximately 600 ml destaining solution containing 45% methanol and 7% glacial acetic acid. The gels were then dried under vacuum in a gel dryer for approximately 3 hours at 70ºC.

2.2.8. Protein purification

Protein purification was done according to the manufacturer's instructions, using a commercially available Protino® Ni-TED kit (Macherey-Nagel, Düren, Germany). All clarification steps were performed at 10 000 x g for 30 minutes. This kit enables fast and convenient purification of recombinant polyhistidine-tagged proteins by immobilized metal ion chromatography (IMAC). Protino® Ni-TED is a dry silica-based resin precharged with Ni^{2+} ions. The interaction between the polyhistidine tag of the recombinant protein and immobilized Ni^{2+} ions is based on how the proteins bind. The bacterial cells were lysed by re-suspending the cell pellet in 5 ml Bugbuster Protein Extraction Reagent (Novagen, Wisconsin, United States of America) at a final concentration of 200 mg/ml. A 1 µl aliquot of r-lysozyme (30KU/µl) (Novagen, Wisconsin, United States of America) was added to achieve a final concentration of 1 mg/ml and 50 units/ml of benzonase (Novagen, Wisconsin, United States of America) was added to the cell suspension. The BugBuster Protein Extraction Reagent along with Benzonase Nuclease provides an efficient combination for releasing target proteins and distinctly reducing extract viscosity prior to downstream applications. The cell suspension was incubated at room temperature for 30 minutes with gentle agitation. The lysed cells were further broken down by sonication on ice using the Soniprep 150 with 10 x 15 second bursts with a 15 second cooling period on ice between each burst. This step was shown to increase yield of the expressed protein. After sonication the suspension was clarified at 4ºC and the pellet was re-suspended in 2.5 ml 1 x lysis-equilibration-wash (LEW) buffer (supplied in the kit) and washed as previously described.

2.2.8.1. Protein purification under denaturing conditions

For purification of recombinant proteins expressed from pQE-80L-YFVEDIII under denaturing conditions, the pellet was re-suspended in 2.5 ml denaturing solubilisation (DS) buffer (pH 8.0) containing 8M urea to solubilize the inclusion bodies. The proteins were then pooled and incubated at room temperature for 60 minutes with gentle agitation. The cell extract obtained was clarified by centrifugation at 10 000 x g for 30 minutes at 20 ºC. The supernatant was transferred to a clean tube. Protino® Ni-TED columns were equilibrated with 2 ml DS buffer and allowed to drain by gravity. The supernatant was added to the pre-equilibrated column and allowed to drain by gravity. The column was washed 2 x 2 ml with DS Buffer to reduce non-specific binding of *E.coli* proteins. The protein was eluted with 2.5 ml denaturing elution buffer containing 8M urea and 0.25M imidazole and the eluates were collected in 500 µl aliquots.

2.2.8.2. Protein purification under native conditions

For purification of pColdTF-YFVEDIII recombinant proteins under native conditions the cells were lysed as described above excluding the use of DS with addition of urea. After resuspending the cells in LEW buffer, samples were centrifuged at 10 000 x g for 30 minutes at 4ºC. The supernatant was stored on ice at 4ºC until applied to the Protino Ni-TED columns. Protino® Ni-TED columns were equilibrated with 2 ml LEW buffer. The supernatant was added to the pre-equilibrated column. The column was washed 2 x 2 ml with LEW. The protein was eluted with 2.5 ml elution buffer containing 0.25M imidazole and the eluates were collected in 500 µl aliquots.

2.2.8.3. Scale-up of recombinant protein preparation

To obtain a sufficient yield of protein for use in an ELISA, culture volumes were scaled up to 250 ml for expression of recombinant YFV pQE-80L-EDIII and pColdTF-EDIII proteins. A 25 ml volume of LB/ampicillin media was inoculated with a 250 µl aliquot of glycerol stock of pQE-80L-YFVEDIII or pColdTF-YFVEDIII and incubated overnight with shaking at 200 rpm at 37ºC. An induction study was performed by diluting the starter culture 1:10 (25 ml in 225 ml LB/ampicillin media) in pre-warmed LB/ampicillin media and grown at 37ºC with shaking until the culture reached an absorbance at 600 nm as described in Materials and Methods 2.2.4 (pQE-80L) and

2.2.6 (pColdTF). Protein expression was induced by adding a final IPTG concentration of 1 mM to each culture and incubating for 16 hours at 37ºC (pQE-80L) or 24 hours at 16ºC (pCold TF). Cells were harvested and purified as described above.

2.2.9. Refolding purified protein after denaturation

To obtain the correctly folded proteins after solubilisation of the inclusion bodies, excess denaturing and reducing agents have to be removed (Misawa *et al*., 1999). Renaturation or refolding of solubilised inclusion bodies may be initiated by several methods such as dilution, dialysis, diafiltration, gel filtration, and immobilization onto a solid support (Clark, 1998). The simplest method is dilution of the denatured solution in a renaturation buffer (Clark, 1998). Eluates were analysed to determine in which fraction the recombinant protein was eluted. In most instances the protein was eluted in the second fraction, hence protein eluate 2 was diluted 1:2 in LEW buffer (500 µl protein + 1000 µl LEW buffer). Ultrafiltration was used to separate small particles and dissolved molecules from fluid, to separate proteins from buffer components for buffer exchange and the rapid change of ionic or pH environment. A Millipore filter with a 10 kDa molecular weight cut off (Millipore Corporation, Massachusetts, United States of America) was used to retain materials greater than 10 kDa, while salts and water were allowed to pass through. Materials smaller than the pore size rating pass through compared to materials larger than the pore size rating are retained by the filter and either concentrated or separated from low molecular weight contaminants. The diluted eluate was ultra-filtered to a final volume of approximately 500 µl per eluate. The control antigen (pQE-80L with no insert) was prepared similarly. The purified proteins were separated on SDS PAGE to confirm the purity of the proteins and were stored at -70ºC.

2.2.10. Protein Concentration

Protein concentrations were determined using the Quant-iT™ Protein Assay Kit (Life Technologies, New York, United States of America), according to the manufacturer's instructions and analysed using the Qubit™ fluorometer (Life Technologies, New York, United States of America). The Quant-iT Protein Assay is supplied with three standards. All assay reagents were incubated at room temperature before use. Briefly, a Quant-iT™ working solution was prepared by diluting the Quant-iT™ Protein reagent (supplied in the kit) 1:200 in Quant-iT™ Protein buffer (supplied in the kit). An aliquot of 190 µl of Quant-iT working solution

was required for each standard and a 10 µl aliquot of each Quant-iT standard (supplied in the kit) was added to the appropriate tube and mixed by vortexing for 2-3 seconds. Protein samples diluted 1:200 in Quant-iT working solution were briefly mixed by vortexing for 2-3 seconds. All samples were allowed to incubate at room temperature for 15 minutes. All three standards were used to calibrate the Qubit™ fluorometer. The Qubit fluorometer was calibrated for each measure using the three standard solutions prepared. The reading was recorded in µg/ml and the dilution was considered when calculating the final concentration.

2.2.11. Cleavage of TF from the recombinant pCold TF EDIII protein

To remove imidazole and phosphate buffers from recombinant YFV pColdTF-EDIII protein, diluted eluates were ultrafiltered using a 30 kDa Millipore filter before cleaving the TF. Phosphate and imidazole buffers were removed for optimal digestion with the protease enzyme Factor Xa Protease (Novagen, Wisconsin, United States of American). Proteins were digested for 16 hours at 16ºC as illustrated in Table 2.13. Samples containing the cleaved TF fusion protein, and YFV EDIII protein were collected at the following intervals T=3 hours, T=6 hours and T=16 hours. The protein samples were applied onto a 1000 packed Ni+ column. The column was washed with 100 µl Tris buffer, and the eluate collected. Imidazole was omitted from the elution buffer and His tagged TF protein remained bound to the column.

Table 2.13. Restiction enzyme digestion of TF fusion protein.

Reaction components	**Reaction volume**
Factor Xa Protease (400 units)	1 µl
pColdTF-YFVEDIII eluate two protein (2 940 µg/ml)	90 µl
10 x Factor Xa Cleavage Buffer (50 mM Tris-HCl, pH 8.0, 100 mM; 5 mM $CaCl_2$)	10 µl
Total	**101 µl**

2.2.12. Characterization of expressed proteins by Western blot analysis

Proteins were characterized by Western blot analysis using the Pico Fast Western Blot Kit (Thermo Scientific, Massachusetts, United States of America). The proteins were separated on a 8% SDS-PAGE as described in Materials and Methods 2.2.7. Proteins were transferred from the gel to a polyvinylidene difluoride (PVDF) membrane as follows: gels were soaked in transfer buffer (25 mM Tris-HCl, 192 mM glycine and 20% methanol) for approximately 15 minutes to facilitate equilibration by removing electrophoresis buffer salts and detergents. The PVDF membrane (Roche, Mannheim, Germany) was moistened with methanol to provide affinity, incubated in distilled water for 1-2 minutes to elute the methanol and soaked in transfer buffer for 2-3 minutes, to ensure proper binding. The transfer buffer allows the adsorption of proteins to the membrane during transfer and maintains a chemical environment that maintains the solubility of the proteins. The Trans-Blot® SD semi-dry electrophoretic transfer cell (Bio-rad, California, United States of America) was assembled and proteins were transferred for 60 minutes at 20 V. The membrane was briefly washed in fast western 1 x wash buffer (supplied in the kit) to remove transfer buffer. A mouse anti-His IgG antibody (100mg/ml) (Roche, Mannheim, Germany) diluted 1:200 in the fast western antibody diluents (supplied in the kit) was added to the membrane. The membrane was incubated for 30 minutes at room temperature with shaking. The membrane was removed from the anti-His antibody solution and placed in a clean incubation tray. Anti-mouse IgG horse radish peroxidise (HRPO) reagent (supplied in the kit) diluted 1:10 in antibody diluents was added to the membrane and the membrane incubated for 10 minutes with shaking at room temperature. The membrane was removed from the HRPO solution and washed by suspension in approximately 20 ml of 1 x wash buffer with gentle agitation for five minutes. The wash was repeated three times. The membrane was then exposed to the supersignal west pico solution (supplied in the kit) mixed according to the manufacturer's instructions for one minute at room temperature. The membrane was placed in a clear plastic wrap and exposed to film (CL-Xposure Film, Thermo Scientific, Massachusetts, United States of America) for approximately 60 seconds. The film was incubated for three minutes in film developer and for two minutes in fixer. The film was washed with water to remove excess fixer and visualized.

2.3. Results

2.3.1. One step RT-PCR amplification of the genes encoding the C, NS4a and EDIII of YFV

YFV RNA was extracted from mouse brain suspension and amplified using primer pairs described in Materials and Methods 2.2.1.1 and 2.2.2.2. A 7 µl aliquot of each amplicon was analysed by gel electrophoresis on a 1% agarose gel. In Figure 2.4, the amplicon in lane 2 is approximately 300 bp according to the molecular weight marker which is the predicted size of the EDIII gene. In Figure 2.5 the amplicons in lane 2 and lane 3 are approximately 400 bp and 900 bp respectively in comparison with the molecular weight marker which are consistent with their predicted sizes of 360 bp and 861 bp.

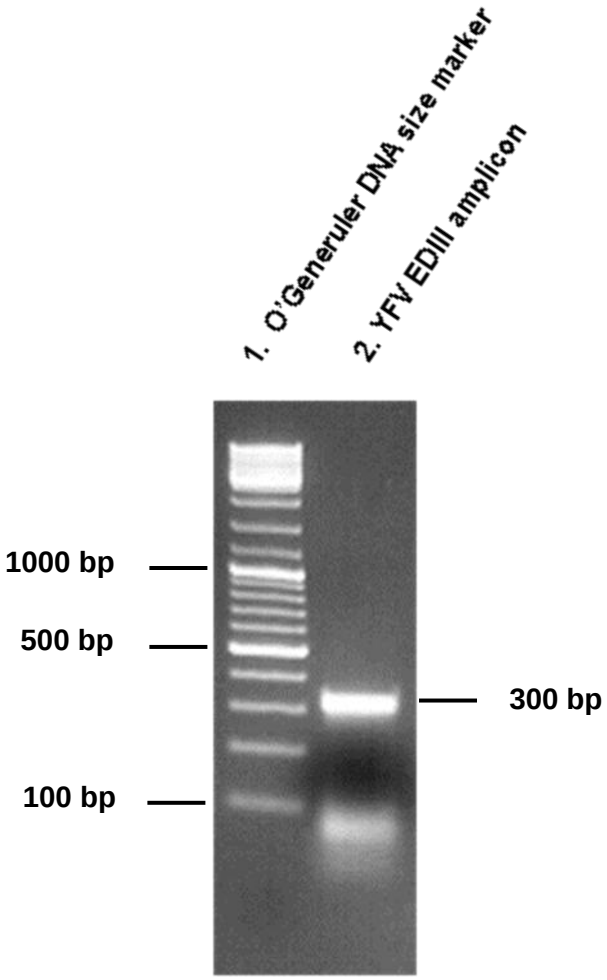


Figure 2.4. Agarose gel electrophoresis analysis of YFV EDIII amplicon. 1) O'Generuler™ DNA size marker (100 bp-10 000 bp); 2) YFV EDIII amplicon.

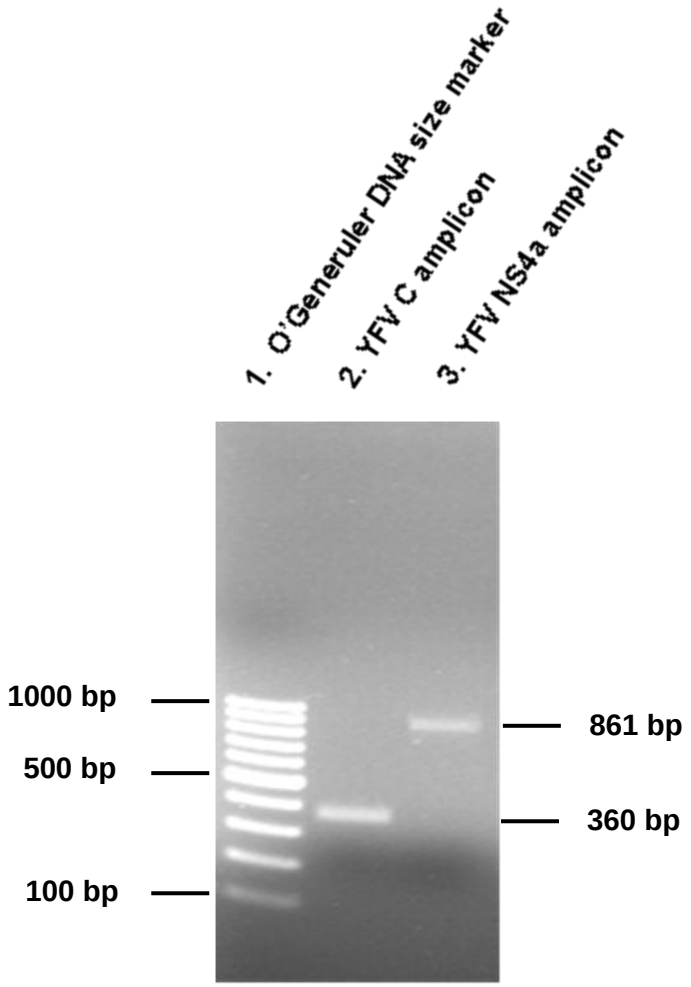


Figure 2.5. Agarose gel electrophoresis analysis of YFV C and NS4a amplicons. 1) O'Generuler™ DNA size marker (100 bp-1000 bp); 2) YFV C amplicon; 3) YFV NS4a amplicon.

Subsequent to purification of each amplicon, a 5 µl aliquot of the purified product was confirmed by gel electrophoresis on a 1% agarose gel. The DNA concentration was determined using the Nanodrop.

2.3.2. A/T cloning of the YFV C, NS4a and EDIII amplicons into the pGEM®-T Easy vector

The YFV amplicons were ligated into pGEM®-T Easy using the overhanging thymidines present on the 3' end of the linearized plasmid. Transformation cultures were incubated overnight on LB/ampicillin/IPTG/X-gal plates and blue and white colonies were obtained from each transformation experiment. Blue colonies indicate that the *lacZ* gene has not been disrupted and that the cloning of the gene into the plasmid and subsequent transformation of the cells was unsuccessful. White colonies represent successful cloning of the gene resulting in disruption of the *lacZ* gene. Positive transformants were confirmed by selection of two or three white colonies from each transformation reaction, propagation of each colony in LB/ampicillin broth overnight, purification of DNA plasmid and restriction analysis of plasmid using *Bam*HI and *Hind*III restriction enzymes (as described in Materials and Methods 2.2.2.2 and 2.2.2.3). Results are shown in Figure 2.6 for cloning the EDIII gene, Figure 2.7 for cloning the C gene and Figure 2.8 for cloning the NS4a gene. In Figure 2.6 undigested and digested plasmids obtained from three colonies were analysed. For each colony, digestion of the plasmid resulted in an approximately 300 bp fragment confirming insertion of the EDIII encoding gene. Similarly, positive transformants were identified in Figures 2.7 and 2.8 confirming insertion of YFV C encoding gene (in two selected colonies) and YFV NS4a encoding gene (in three selected colonies) respectively. As shown in each Figure, supercoiled plasmids migrated further than linearised plasmids. Figure 2.7 shows that the double digestion of the recombinant YFV C plasmid with *Bam*HI and *Hind*III yielded the predicted 360 bp fragment from two colonies. In Figure 2.8, double digestion of the recombinant YFV NS4a plasmid with *Bam*HI and *Hind*III yielded a 861 bp fragment.

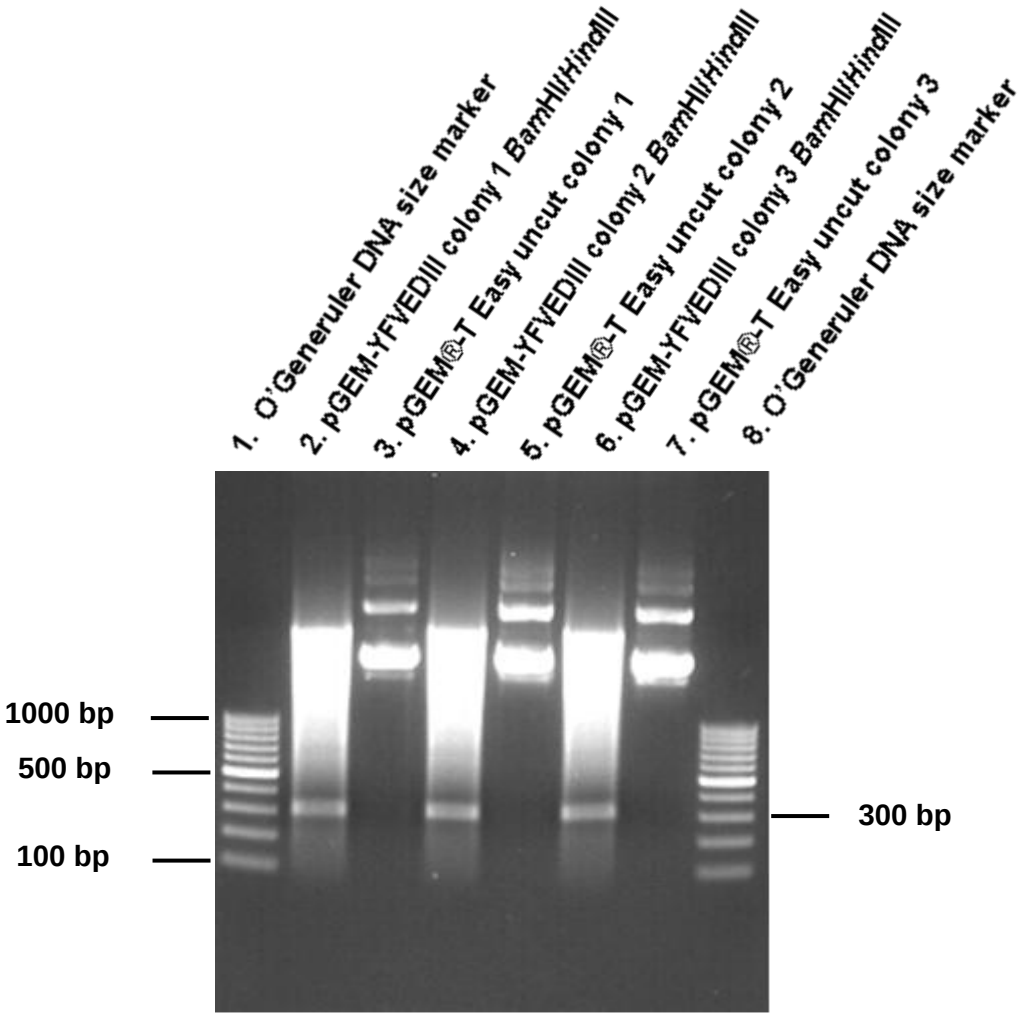


Figure 2.6. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler™ DNA size marker (100bp-1000bp); 2) pGEM-YFVEDIII (colony 1) digested with *Bam*HI and *Hind*III; 3) pGEM®-T Easy plasmid uncut (colony 1); 4) pGEM-YFVEDIII (colony 2) digested with *Bam*HI and *Hind*III; 5) pGEM®-T Easy plasmid uncut (colony 2); 6) pGEM-YFVEDIII (colony 3) digested with *Bam*HI and *Hind*III; 7) pGEM®-T Easy plasmid uncut (colony 3); 8) O'Generuler™ DNA size marker (100 bp-1000 bp).

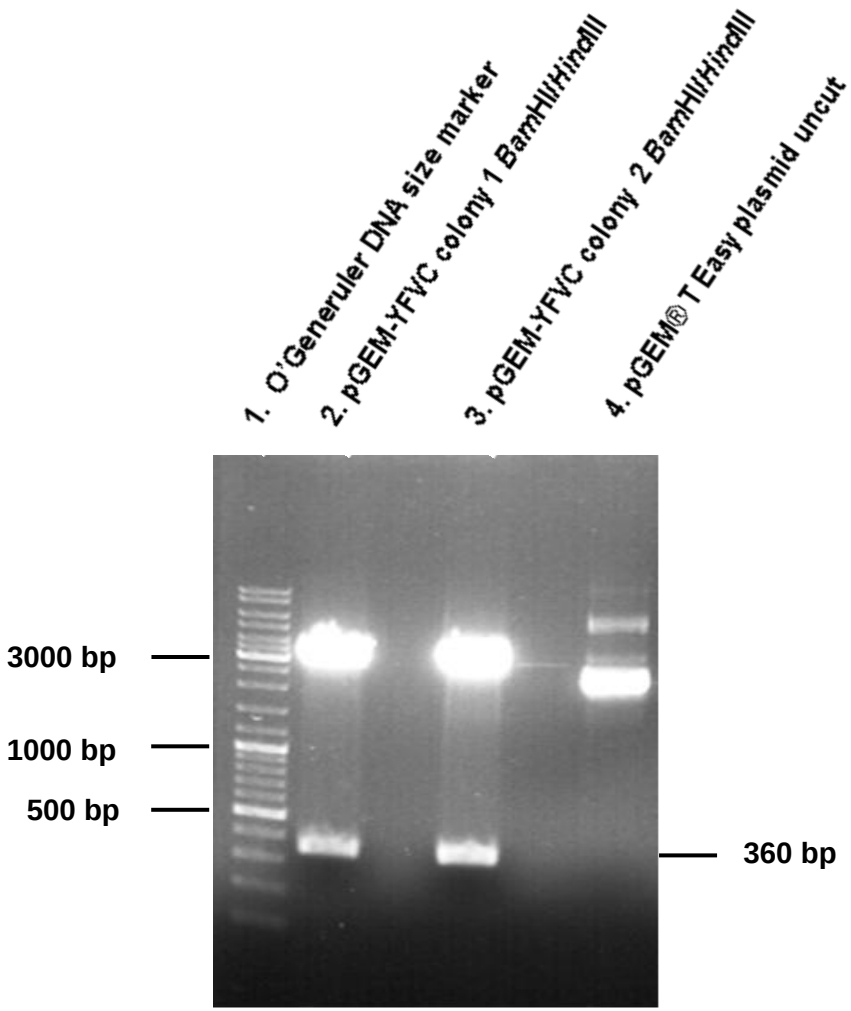


Figure 2.7. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVC construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler™ DNA size marker (100bp-10 000bp); 2) pGEM-YFVC (colony 1) digested with *Bam*HI and *Hind*III; 3) pGEM-YFVC (colony 2) digested with *Bam*HI and *Hind*III; 4) pGEM®-T Easy plasmid uncut (colony 1).

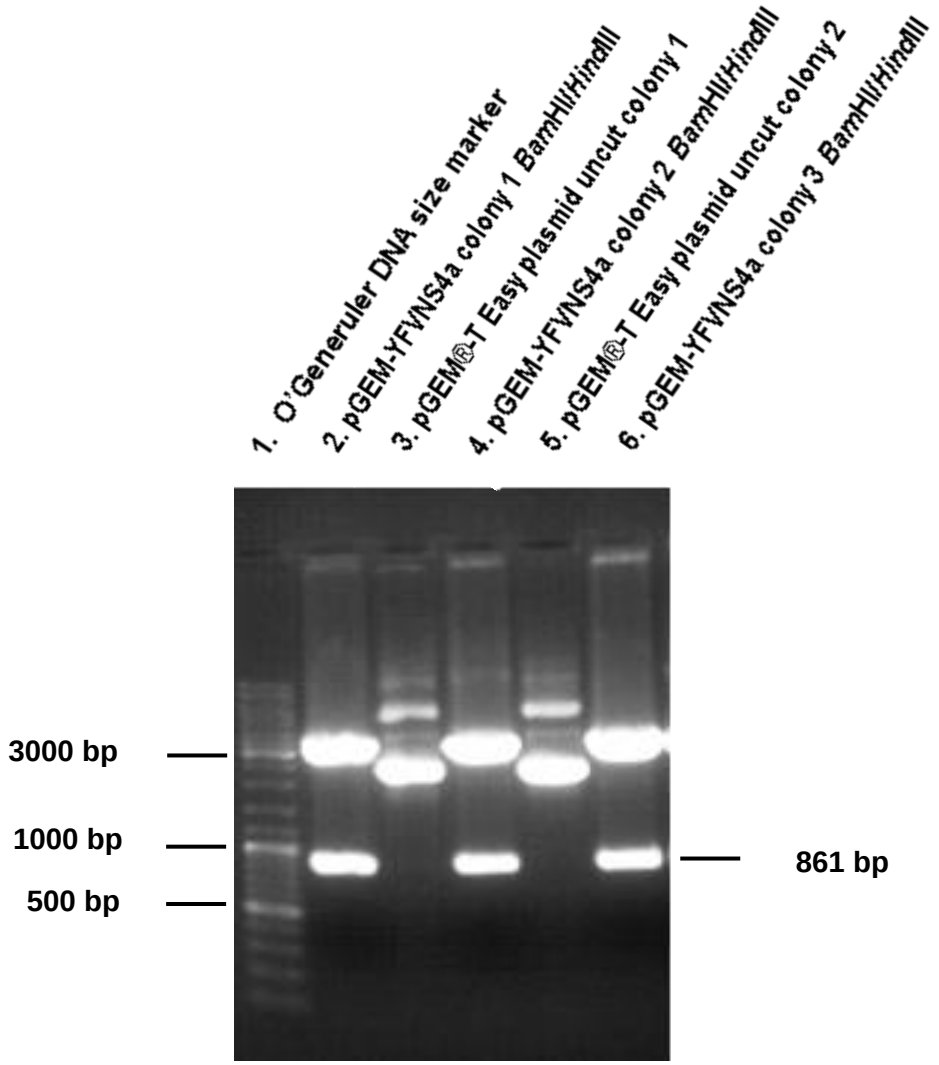


Figure 2.8. Agarose gel electrophoresis analysis of restriction enzyme digestions of pGEM-YFVNS4a construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler™ DNA size marker (100bp-10 000bp); 2) pGEM-YFVNS4a (colony 1) digested with *Bam*HI and *Hind*III; 3) pGEM®-T Easy plasmid uncut (colony 1); 4) pGEM-YFVNS4a (colony 2) digested with *Bam*HI and *Hind*III; 5) pGEM®-T Easy plasmid uncut (colony 2); 6) pGEM-YFVNS4a (colony 3) digested with *Bam*HI and *Hind*III.

For each gene one positively transformed colony was selected for downstream application. The DNA bands were excised from the agarose gels and purified. After purification of the excised bands, a 5 µl aliquot was visualised by gel electrophoresis on a 1% agarose gel as shown in Figures 2.9-2.11. The presence of the following bands depicted in lane 2 of each gel with the correct expected size of 300 bp (EDIII), 360 bp (C) and 861 bp (NS4a), respectively was confirmed. The concentration of the purified DNA was determined for ligation of the rescued genes into expression plasmids.

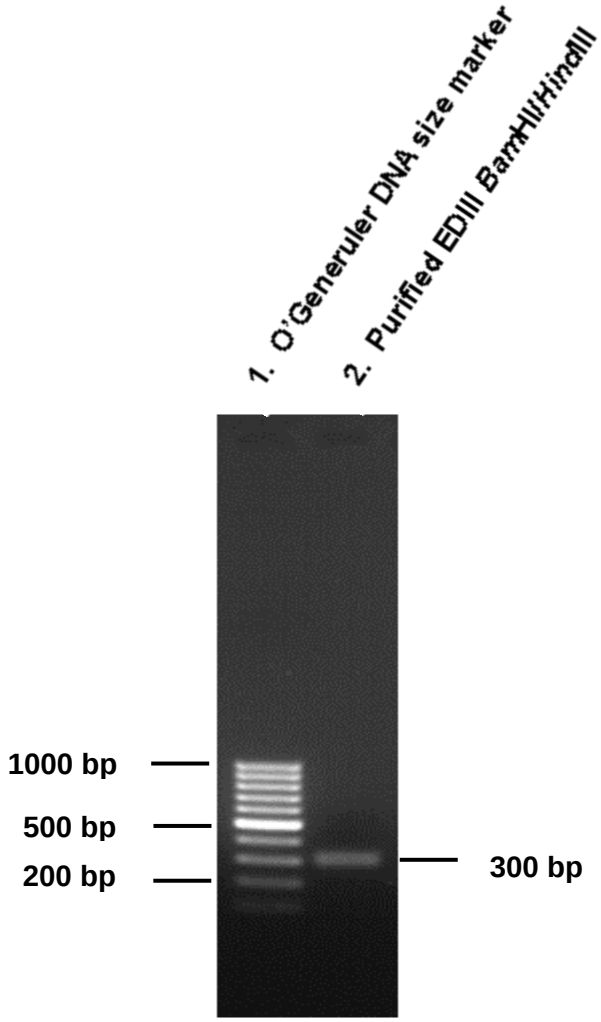


Figure 2.9. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested YFV EDIII gene. 1) O'Generuler DNA size marker (100bp-1000bp); 2) Purified YFV EDIII gene digested with *Bam*HI and *Hind*III.

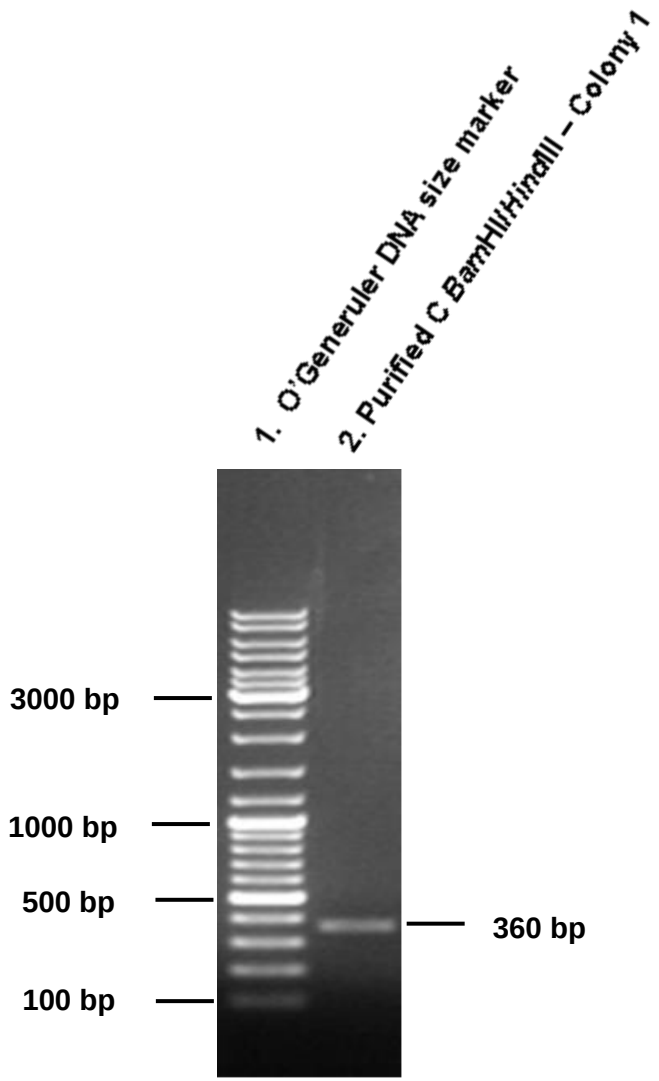


Figure 2.10. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested YFV C gene. 1) O'Generuler DNA size marker (100 bp-10 000 bp); 2) Purified YFV C gene digested with *Bam*HI and *Hind*III.

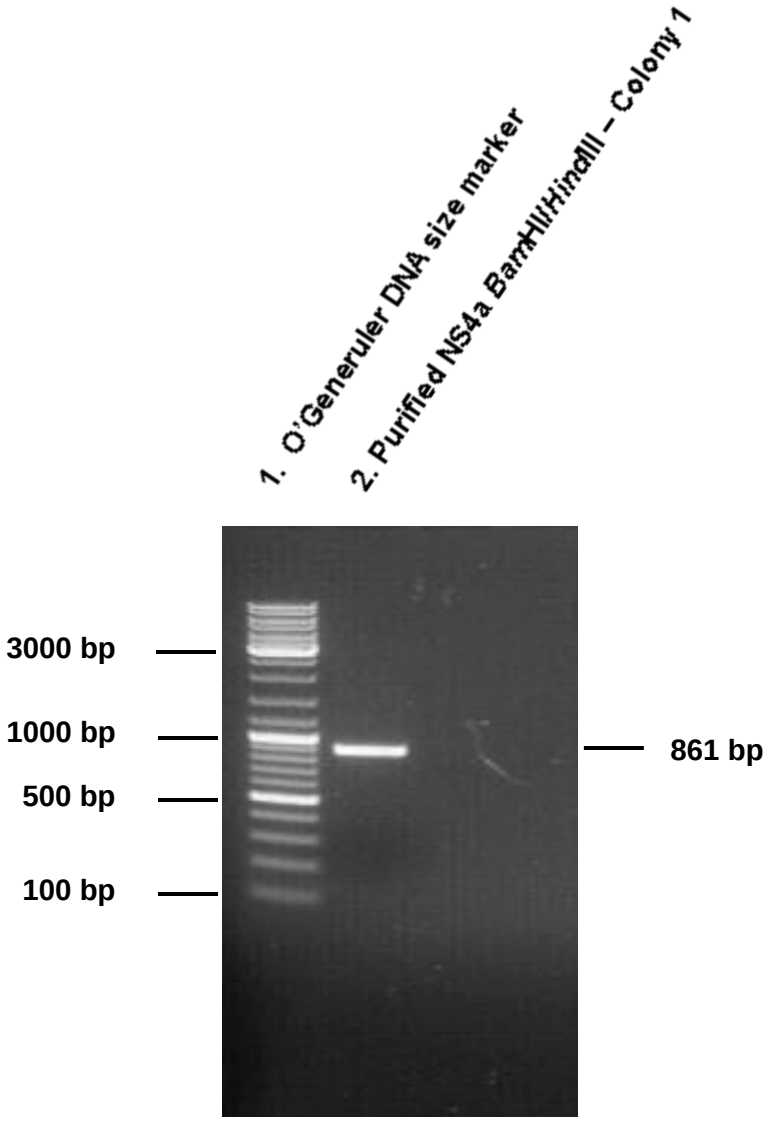


Figure 2.11. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested YFV NS4a gene. 1) O'Generuler DNA size marker (100 bp-10 000 bp); 2) Purified YFV NS4a gene digested with *Bam*HI and *Hind*III.

2.3.3. Cloning of the YFV EDIII into pQE-80L

The YFV EDIII gene was previously cloned into the pQE-80L plasmid vector (in an unrelated project) and was stored as glycerol stocks (designated pQE-80L-YFVEDIII). The presence of the YFV EDIII gene in stored glycerol stocks was confirmed by inoculating each of three 5 ml aliquots of LB/ampicillin broth with 10µl glycerol stock, overnight propagation of the cultures, and restriction analysis of purified DNA plasmids the following day. Separation of the digestion products showed a 300 bp fragment in lanes 2-5 in Figure 2.12 confirming the presence of the 300 bp YFV EDIII gene.

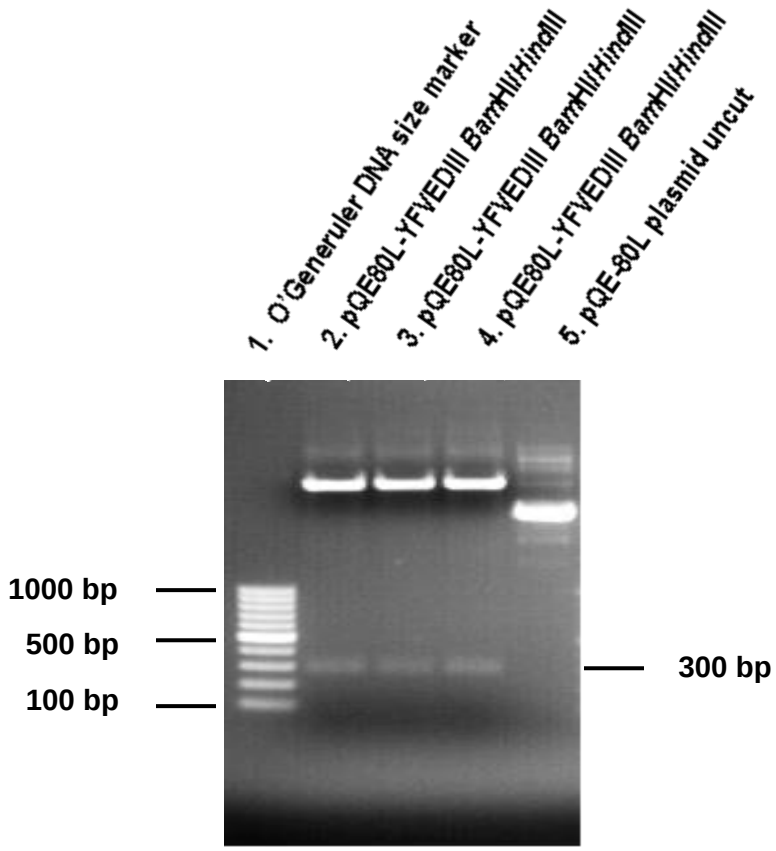


Figure 2.12. Agarose gel electrophoretic analysis of restriction enzyme digestions of the pQE-80L-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler™ DNA size marker (100bp-1000bp); 2) pQE-80L-YFVEDIII construct digested with *Bam*HI and *Hind*III; 3) pQE-80L-YFVEDIII construct digested with *Bam*HI and *Hind*III; 4) pQE-80L-YFVEDIII construct digested with *Bam*HI and *Hind*III; 5) pQE-80L plasmid uncut.

2.3.4. Cloning of the YFV C gene into pQE-80L

A ligation reaction was performed as described in Methods and Materials 2.2.3.2 using 42.16 µg of purified YFV C gene and 14.4 µg linearized pQE-80L vector and T4 DNA ligase. Transformation of OverExpress cells with the ligation mix was performed as described in Materials and Methods 2.2.3.2. The transformation culture was incubated overnight on LB/ampicillin plates and three colonies were selected the following day for identification of positive transformants. A portion of each colony that was selected for PCR was inoculated onto an LB/ampicillin plate so that positive colonies could be further propagated. PCR amplification using YFV specific primers confirmed the presence of a 360 bp gene in each of the three colonies (Figure 2.13 designated colony 1-3 (lane 2-4)). The presence of the YFV C gene in pQE-80L was further confirmed by restriction enzyme analysis which yielded the expected 360 bp fragment (lane 2-4) after double digestion of colony DNA with *Bam*HI and *Hind*III as shown in Figure 2.14.

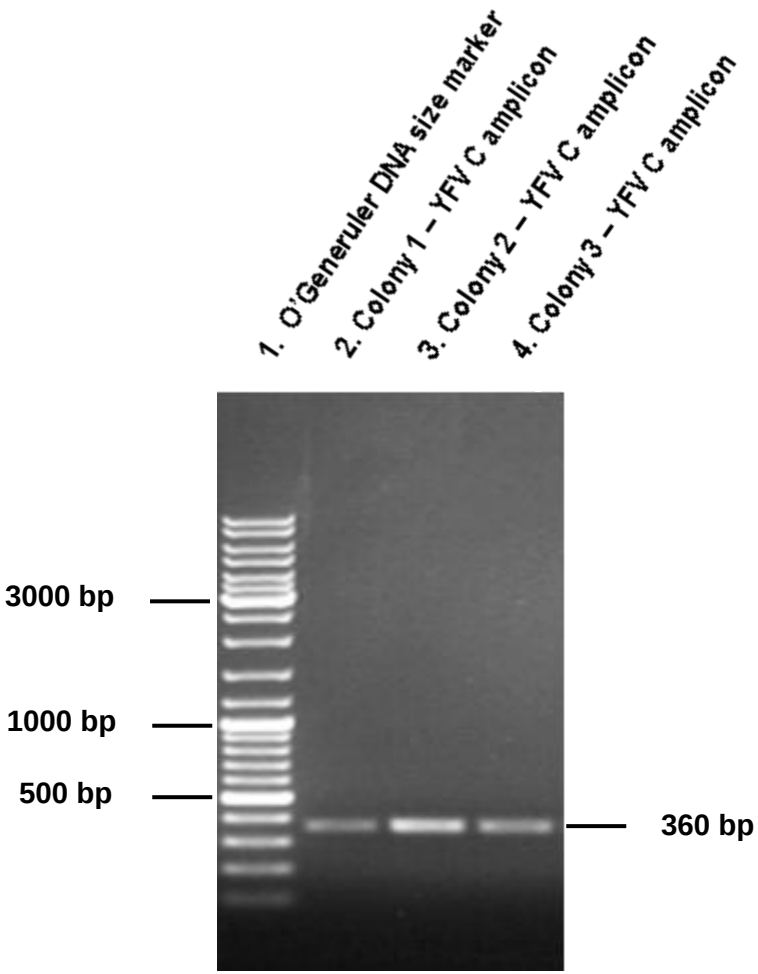


Figure 2.13. Agarose gel electrophoresis of YFV C amplicons performed on three pQE-80L colonies. 1) O'Generuler™ DNA size marker (100 bp-10 000 bp); 2) YFV C amplicon of colony 1; 3) YFV C amplicon of colony 2; 4) YFV C amplicon of colony 3.

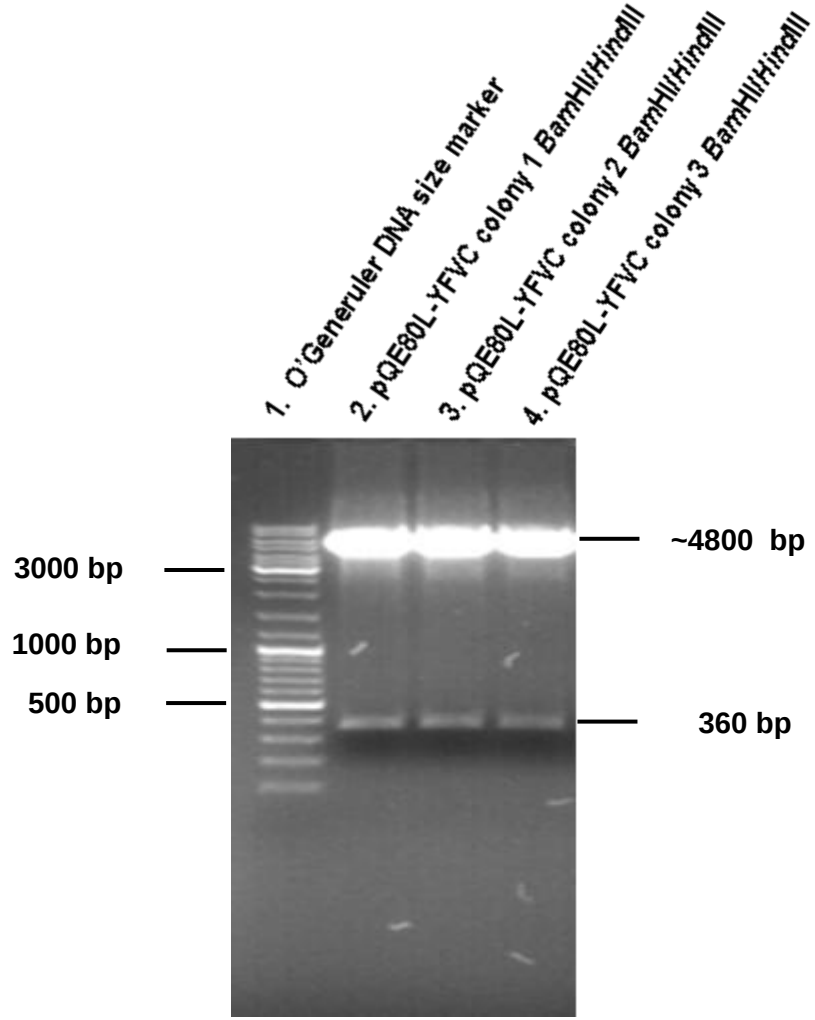


Figure 2.14. Agarose gel electrophoretic analysis of restriction enzyme digestions of the pQE-80L-YFVC construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler™ DNA size marker (100bp-10 000bp); 2) pQE-80L-YFVC construct (colony 1) digested with *Bam*HI and *Hind*III; 3) pQE-80L-YFVC construct (colony 2) digested with *Bam*HI and *Hind*III; 4) pQE-80L-YFVC construct (colony 3) digested with *Bam*HI and *Hind*III.

2.3.5. Sequence analysis of the gene encoding YFV C in pQE-80L vector

The nucleotide sequence data for the construct was determined and used to confirm that the gene was inserted in frame with no mutations occurring as a result of the PCR amplification of the gene. The sequence data of the gene encoding the C protein from the start codon of the pQE-80L vector is shown in Appendix A. 1. Analysis of the sequence data confirmed the correct DNA fragment of 360 bp. The ORF was cloned with the *Bam*HI and *Hind*III restriction enzyme sites and confirmed to be in frame with the start codon, His tag sequence and stop codon of the pQE-80L vector.

2.3.6. Cloning of the YFV NS4a gene into pQE-80L

The confirmation of insertion of the correct DNA into pQE-80L vector is shown in Figure 2.15. A ligation reaction was performed using 124 ng of purified YFV NS4a gene and 14.4 µg pQE-80L vector and T4 DNA ligase. Transformation of OverExpress cells with the YFV NS4a construct was performed as described in Materials and Methods 2.2.3.2. As described for the C gene ligation, transformation culture was incubated overnight on LB/ampicillin plates and one colony was selected the following day. A portion of the colony that was selected for PCR was inoculated onto an LB/ampicillin plate for further propagation. PCR amplification using YFV specific primers confirmed the presence of an 861 bp gene in the colony lane 1, Figure 2.15. Further confirmation was obtained from restriction digest analysis of three selected colonies which yielded an 861 bp fragment (lane 3, lane 4 and lane 5) after double digestion with *Bam*HI and *Hind*III (Figure 2.16).

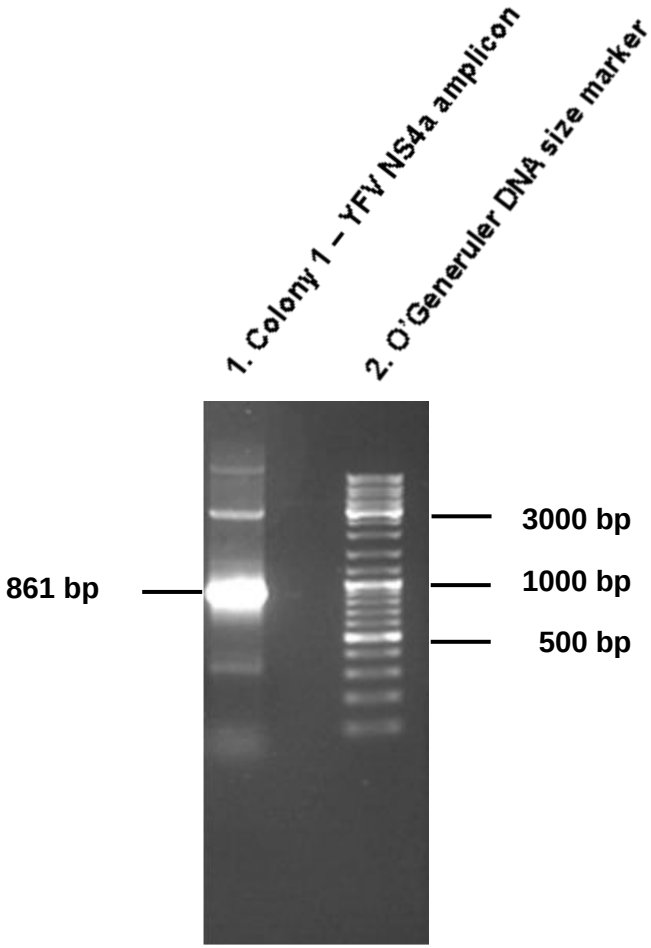


Figure 2.15. Agarose gel electrophoresis analysis of YFV NS4a amplicon performed on a pQE-80L colony. 1) YFV NS4a amplicon - colony 1; 2) O'Generuler DNA size marker (100bp-10 000bp).

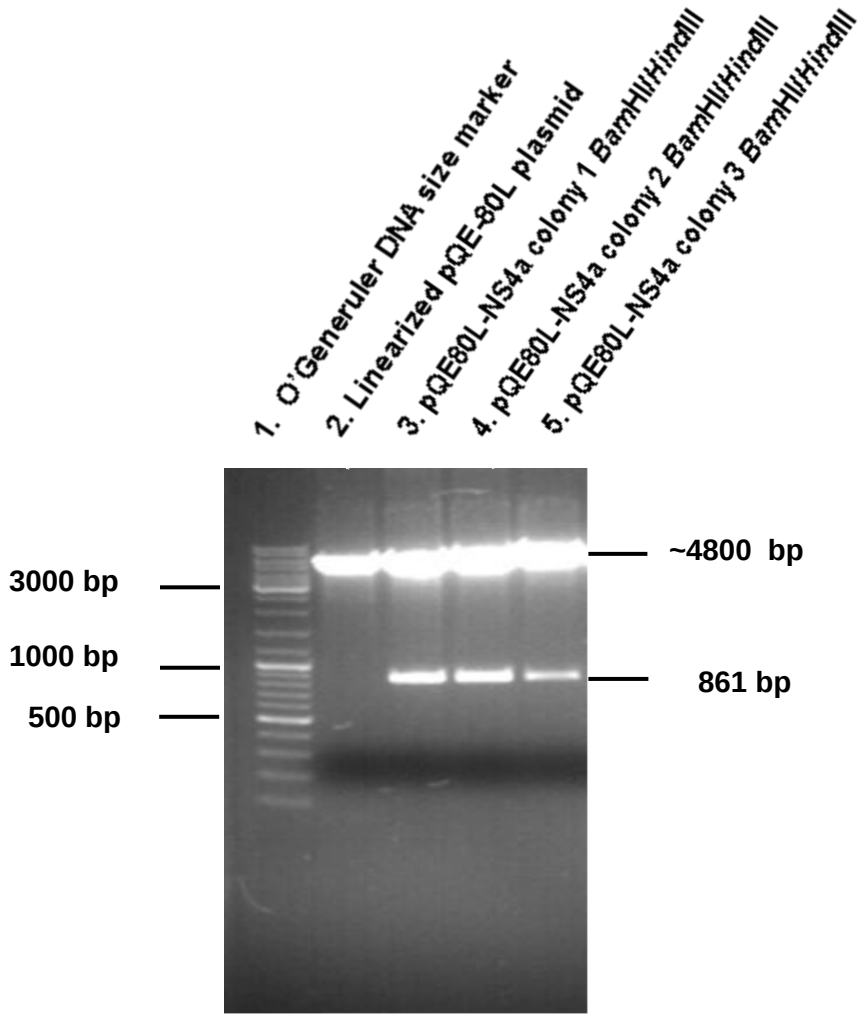


Figure 2.16. Agarose gel electrophoretic analysis of restriction enzyme digestions of the pQE-80L-YFVNS4a construct using *Bam*HI and *Hind*III restriction endonucleases. 1) O'Generuler DNA size marker (100bp-10 000bp); 2) Linearized pQE-80L plasmid; 3) pQE-80L-YFVNS4a construct (colony 1) digested with *Bam*HI and *Hind*III; 4) pQE-80L-YFVNS4a construct (colony 2) digested with *Bam*HI and *Hind*III; 5) pQE-80L-YFVNS4a construct (colony 3) digested with *Bam*HI and *Hind*III.

2.3.7. Sequence analysis of the gene encoding YFV NS4a in pQE-80L vector

The nucleotide sequence data for the construct was determined and used to confirm that the gene was inserted in frame with no mutations. The sequence data of the gene encoding the NS4a protein from the start codon of the pQE-80L vector is shown in Appendix A. 2. Analysis of the sequence data confirmed the correct DNA fragment of 861 bp. The ORF was cloned with the *Bam*HI and *Hind*III restriction enzyme sites and confirmed to be in frame with the start codon, His tag sequence and stop codon of the pQE-80L vector.

2.3.8. Cloning of YFV EDIII gene into pCold TF

pCold TF DNA plasmid propagated in LB/ampicillin broth and purified was prepared for ligation reactions by linearization with *Bam*HI and *Hind*III to obtain plasmid with compatible 5' and 3' cloning ends for ligation. The confirmation of the linearized pCold TF vector (approximately 6000 bp DNA band) is shown in Figure 2.17. The undigested supercoiled pCold TF plasmid DNA (lane 3 and lane 5) migrated further than the linearized pCold TF DNA (lane 2 and lane 6). The linearized DNA was excised and purified as described in Methods and Materials. Subsequent to purification a 5 µl aliquot of purified DNA was visualised on an agarose gel (Figure 2.18). The DNA concentration of the linear purified pCold TF DNA was 9.6 ng/µl.

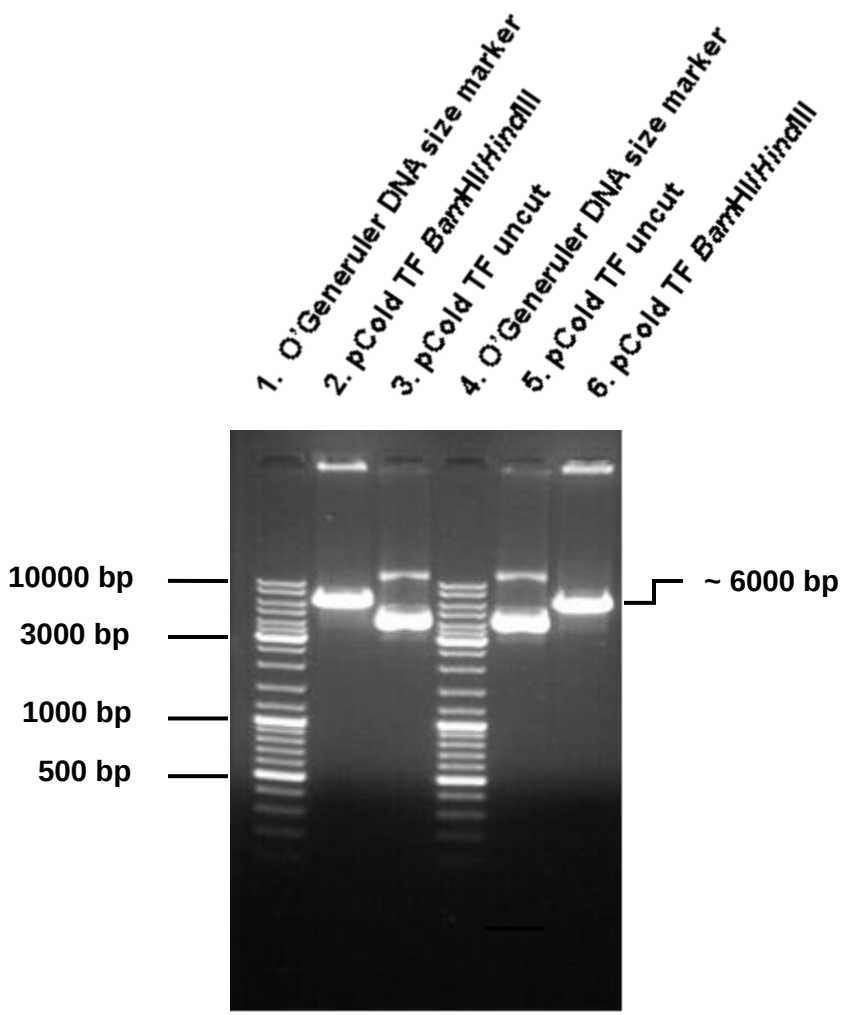


Figure 2.17. Agarose gel electrophoresis of restriction enzyme digestions of pCold TF plasmid DNA. 1) O'Generuler DNA size marker (100bp-10 000bp); 2) pCold TF digested with *Bam*HI and *Hind*III; 3) pCold TF uncut (supercoiled); 4) O'Generuler DNA size marker (100bp-10 000bp); 5) pCold TF uncut (supercoilded); 6) pCold TF digested with *Bam*HI and *Hind*III.

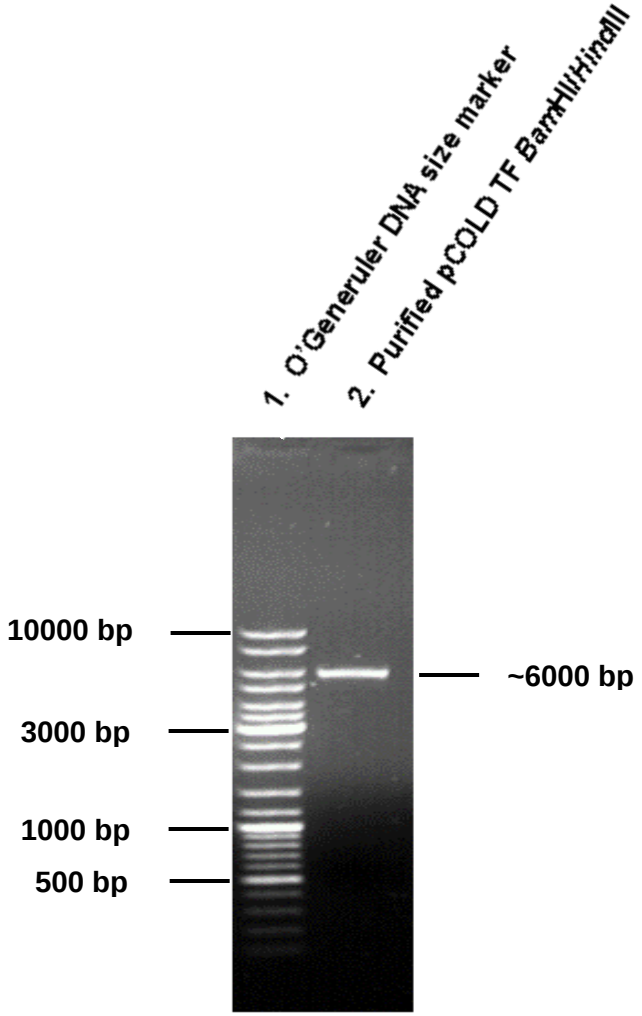


Figure 2.18. Agarose gel electrophoresis analysis of restriction enzyme digestions of purified double digested pCold TF plasmid. 1) O'Generuler DNA size marker (100 bp-10 000 bp); 2) Purified pCold TF plasmid digested with *Bam*HI and *Hind*III.

A total of 17.8 ng of purified YFV EDIII gene was ligated into 48 ng purified double digested pCold TF plasmid. The result of ligation reactions and analysis of colonies selected from subsequent transformation cultures is shown in Figures 2.19 and 2.20. Three colonies, designated colony 1, colony 2 and colony 3, were selected and inoculated into 5 ml of LB/ampicillin media. The plasmid DNA was purified and transformants were identified by PCR and restriction enzyme analysis. PCR performed on the three colonies using YFV EDIII specific primers amplified a 300 bp product from each colony (Figure 2.19, lanes 2-4). Further propagation and restriction digest analysis of the colonies showed the presence of a 300 bp gene after digestion with *Bam*HI and *Hind*III (Figure 2.20, lanes 2-4). Colony 1 was selected for downstream application and designated pColdTF-YFVEDIII.

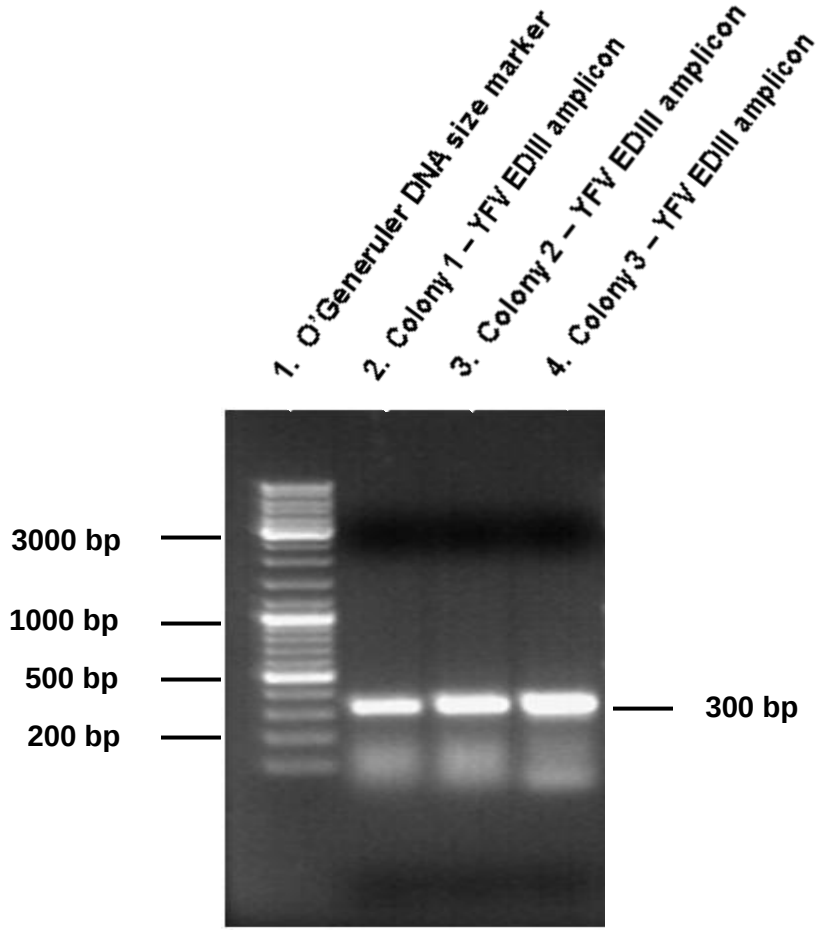


Figure 2.19. Agarose gel electrophoresis analysis of YFV EDIII amplicons on three pCold TF colonies. 1) O'Generuler™ DNA size marker (100bp-10 000bp); 2) YFV EDIII amplicon of colony 1; 3) YFV EDIII amplicon of colony 2; 4) YFV EDIII amplicon of colony 3.

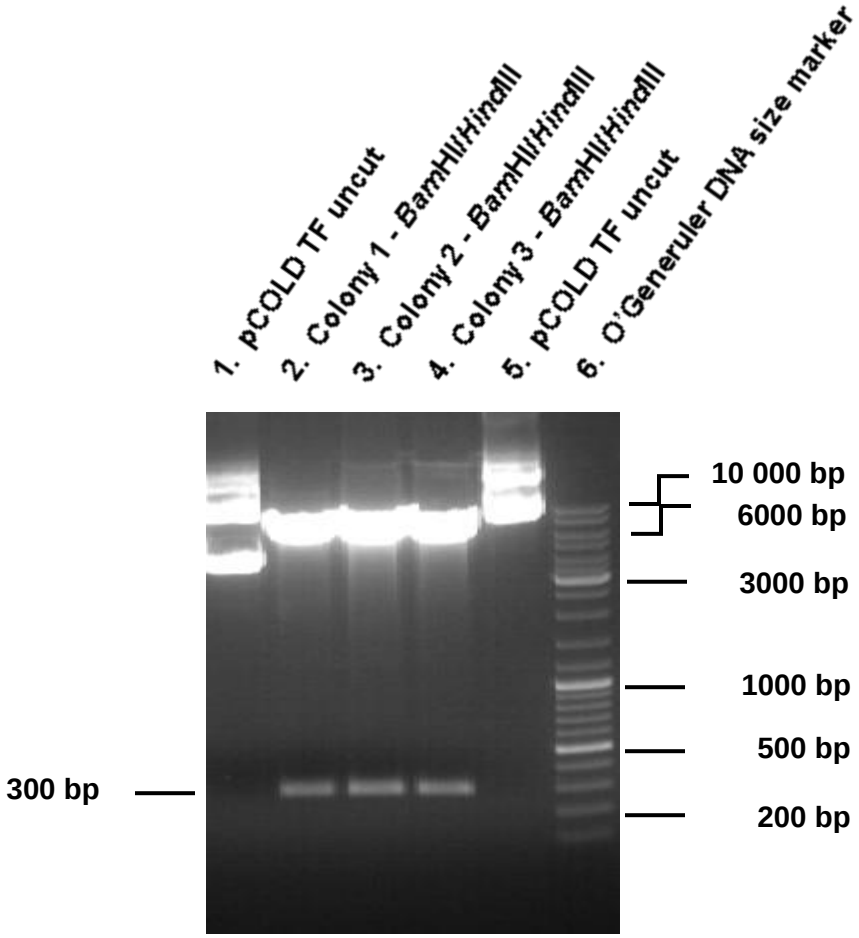


Figure 2.20. Agarose gel electrophoretic analysis of restriction enzyme digestions of the pColdTF-YFVEDIII construct using *Bam*HI and *Hind*III restriction endonucleases. 1) pCold TF plasmid uncut (supercoiled); 2) pColdTF-YFVEDIII construct (colony 1) digested with *Bam*HI and *Hind*III; 3) pColdTF-YFVEDIII construct (colony 2) digested with *Bam*HI and *Hind*III; 4) pColdTF-YFVEDIII construct (colony 3) digested with *Bam*HI and *Hind*III; 5) pCold TF plasmid uncut (supercoiled); 6) O'Generuler DNA size marker (100bp-10 000bp).

2.3.9. Sequence analysis of the gene encoding YFV EDIII in pCold TF

The nucleotide sequence data for pColdTF-YFVEDIII was determined and used to confirm that the gene was inserted in frame with no mutations occurring as a result of the amplification of the gene. The sequence data of the gene encoding the EDIII protein from the start codon of the pCold TF is shown in Appendix A. 3. Analysis of the sequence data confirmed the correct DNA fragment of 300 bp. The ORF was cloned with the *Bam*HI and *Hind*III restriction enzyme sites and confirmed to be in frame with the start codon, His tag sequence, protease cleavage sites and the stop codon of the pCold TF vector.

2.3.10. Expression, solubility, purification and characterization of recombinant EDIII protein frompQE-80L-YFVEDIII construct

In a previous study, the recombinant YFV EDIII protein was cloned and expressed from the pQE-80L-YFVEDIII construct using JM109 *E. coli* cells after IPTG induction with yields insufficient for downstream applications such as ELISA. To increase expression yield and optimize conditions for protein expression, pQE-80L- YFVEDIII construct was used to transform OverExpress cells, and three small-scale cultures were inoculated with pQE-80L-YFVEDIII and protein expression induced using final IPTG concentrations of 0.125 mM at 25ºC, 1 mM at 37ºC and 2.5 mM at 37ºC.

As shown in Figure 2.21 (0.125 mM IPTG), Figure 2.23 (1 mM IPTG), Figures 2.25 and 2.26 (2.5 mM IPTG) respectively, expression of the ~13 kDa recombinant protein at various intervals after induction was confirmed by SDS-PAGE analysis. The recombinant protein size was estimated as ~13 kDa since 1 amino acid = 0.11 kDa, thus 11.1 kDa (EDIII protein) + 1.109 kDa (6 x His tag protein). Expression of a 13 kDa protein was observed in each of the cultures. To determine if the expressed recombinant protein was present in the soluble or insoluble fractions, a protein solubility study was performed.

As shown in Figures 2.22 (0.125 mM), Figure 2.24 (1 mM) and Figure 2.27 (2.5 mM) respectively, the recombinant YFV pQE-80L-EDIII protein (~13 kDa) was detected in the insoluble fractions at T=5 and T=16 hours by SDS-PAGE analysis. Although there was soluble protein present five hours after induction, a higher yield was present in the insoluble fractions at 5 hours and after overnight incubation (16 hours). The levels of expressed protein, observed on the gels, in the soluble fraction were minimal.

From a comparison of the bands on the gels in Figures 2.21, 2.23, 2.25 and 2.26 and considering that a comparable number of cells were estimated using OD values the concentration of IPTG and incubation temperature had little effect on increasing yield. Similarly a comparison of the Figures 2.22, 2.24 and 2.27 suggest that the solubility of the protein was not influenced by the different parameters. As the protein solubility study showed that the recombinant YFV EDIII protein was found predominantly in the insoluble fraction, the protein was purified under denaturing conditions. Due to difference in wells, different amounts of sample were added.

Briefly, the cells were lysed 4 hours post-induction, using buffers containing high concentrations of urea (8M urea), purified on a Ni-TED column in which the His tag binds to the column and is

eluted using imidizole and renaturing of the protein by slow removal of denaturant (urea). Eluates were collected from the Ni-TED columns in 0.5ml fractions and the protein concentrations in each fraction determined and confirmed by SDS-PAGE analysis as shown in Figure 2.28 (0.125 mM), Figure 2.29 (1 mM) and Figure 2.30 (2.5 mM).

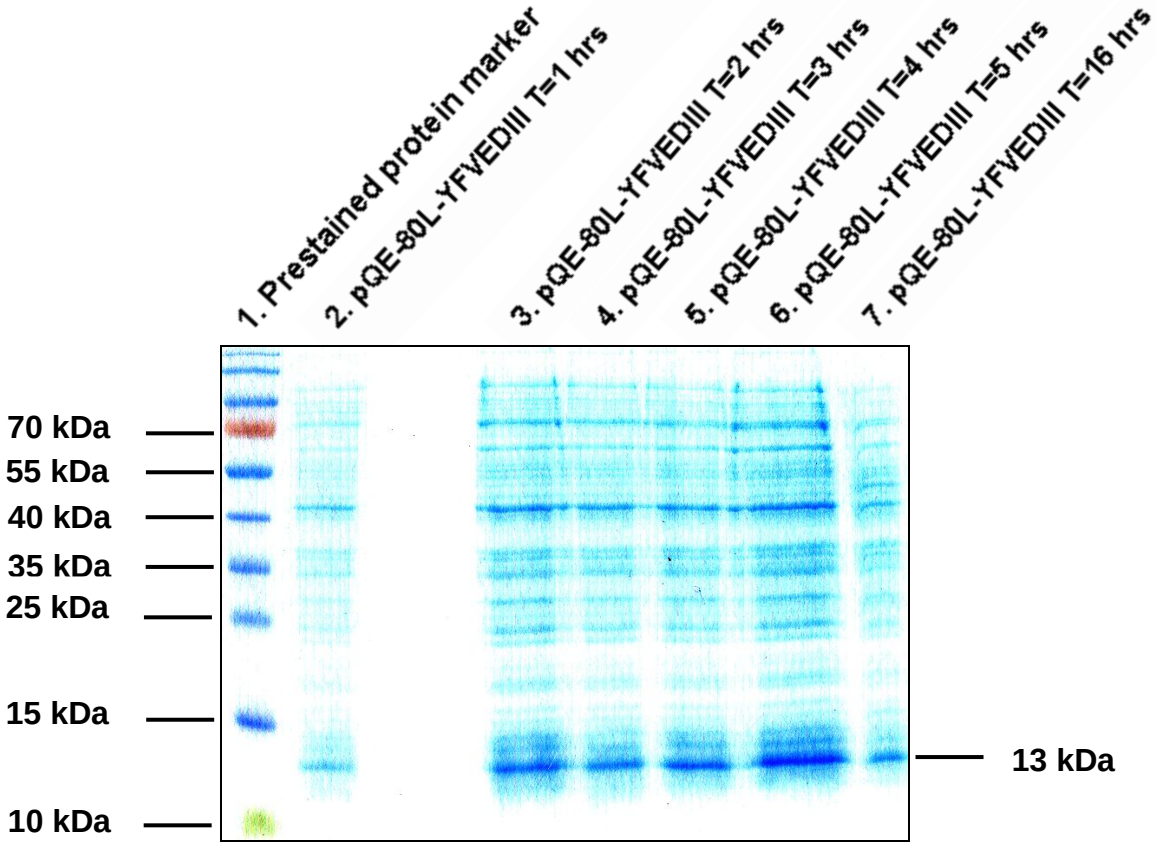


Figure 2.21. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3, T=4, T=5 hours and 16 hours post-induction, with a final IPTG concentration of 0.125 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 0 hours post-induction; 3) Expressed protein after 1 hour post-induction; 4) Expressed protein after 2 hours post-induction; 5) Expressed protein after 3 hours post-induction; 6) Expressed protein after 4 hours post-induction; 7) Expressed protein after 5 hours post-induction; 8) Expressed protein after 16 hours post-induction.

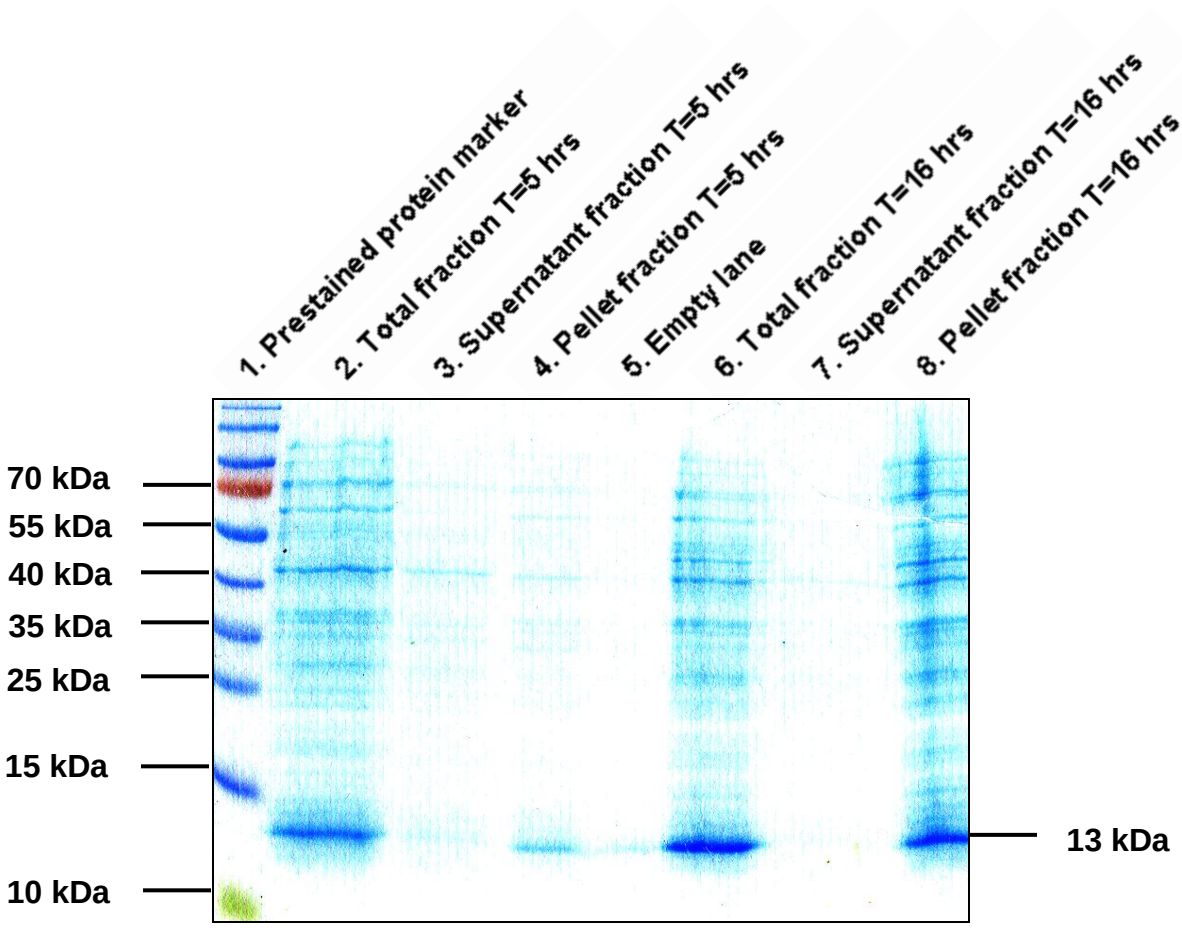


Figure 2.22. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 0.125 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Total fraction proteins expressed after 5 hours induction; 3) Soluble fraction proteins expressed after 5 hours induction; 4) Pellet fraction proteins expressed after 5 hours induction; 5) Empty lane; 6) Total fraction proteins expressed after 16 hours; 7) Soluble fraction proteins expressed after 16 hours; 8) Pellet fractions expressed after 16 hours induction.

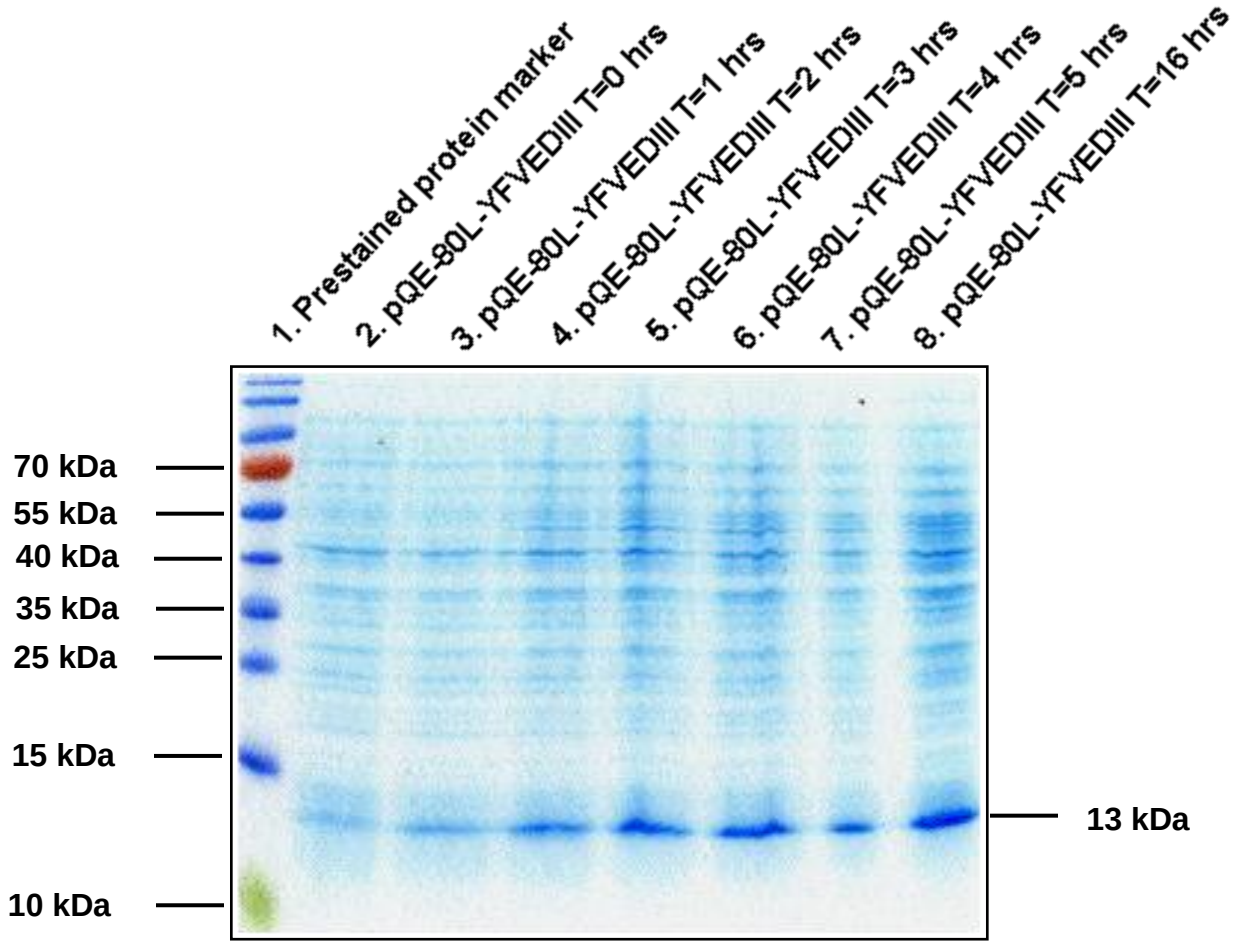


Figure 2.23. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3, T=4, T=5 hours and 16 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 0 hours post-induction; 3) Expressed protein after 1 hour post-induction; 4) Expressed protein after 2 hours post-induction; 5) Expressed protein after 3 hours post-induction; 6) Expressed protein after 4 hours post-induction; 7) Expressed protein after 5 hours post-induction; 8) Expressed protein after 16 hours post-induction.

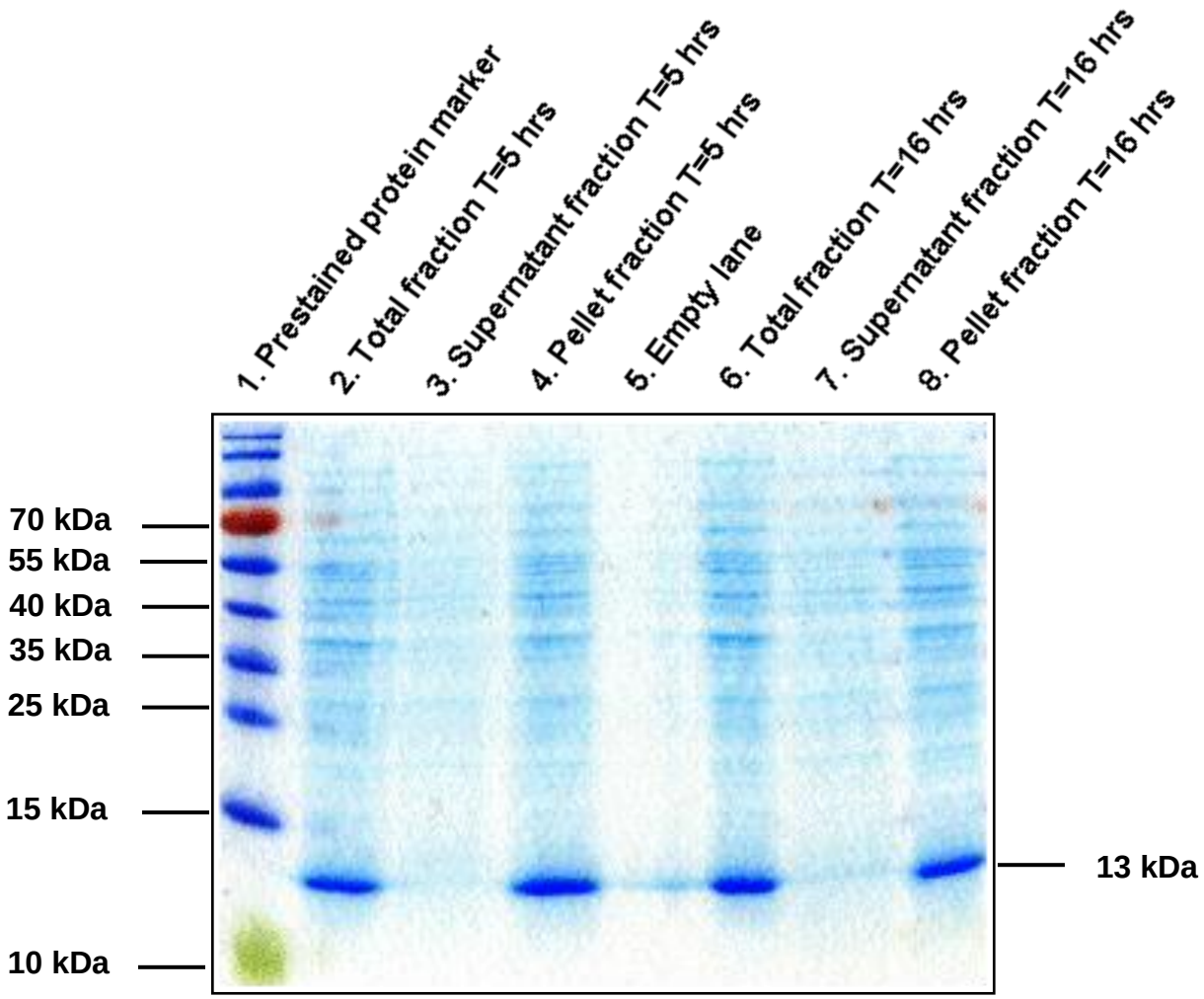


Figure 2.24. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Total fraction proteins expressed after 5 hours induction; 3) Soluble fraction proteins expressed after 5 hours induction; 4) Pellet fraction proteins expressed after 5 hours induction; 5) Empty lane; 6) Total fraction proteins expressed after 16 hours; 7) Soluble fraction proteins expressed after 16 hours; 8) Pellet fractions expressed after 16 hours induction.

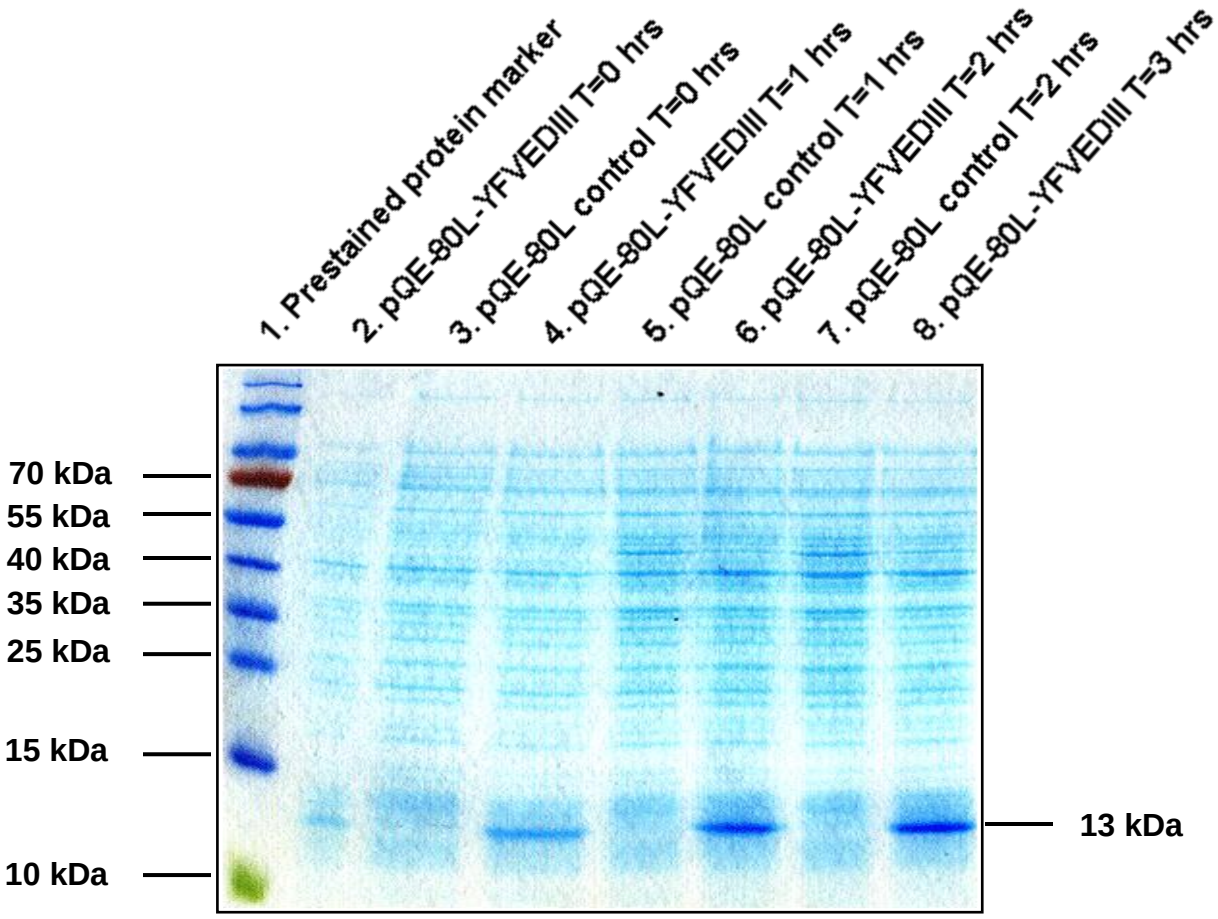


Figure 2.25. SDS-PAGE analysis of proteins expressed at T=0, T=1, T=2, T=3 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 0 hours post-induction; 3) pQE-80L control 0 hour post-induction; 4) Expressed protein after 1 hour post-induction; 5) pQE-80L control 1 hour post-induction; 6) Expressed protein after 2 hours post-induction; 7) pQE-80L control 2 hours post-induction; 8) Expressed protein after 3 hours post-induction.

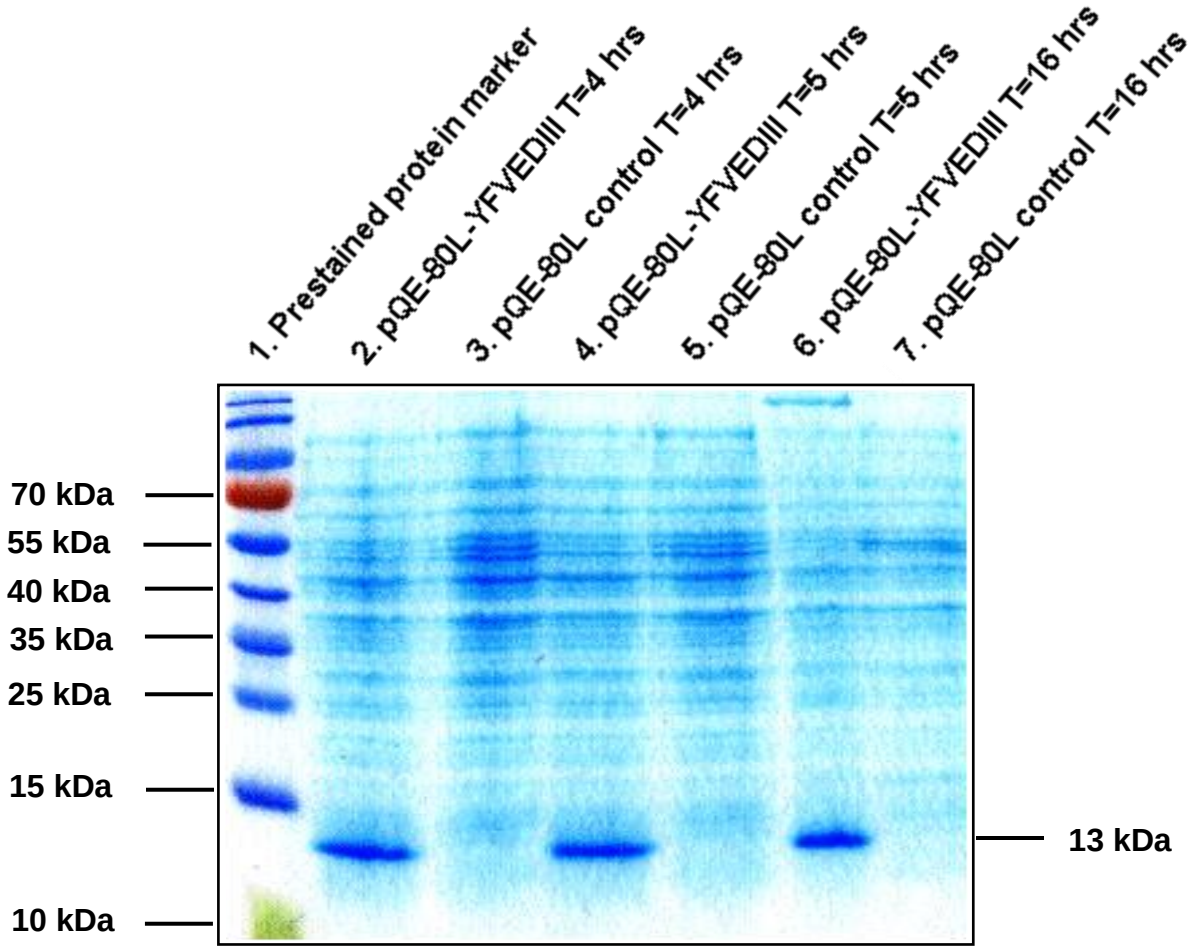


Figure 2.26. SDS-PAGE analysis of proteins expressed at T=4, T=5, T=16 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 4 hours post-induction; 3) pQE-80L control 4 hour post-induction; 4) Expressed protein after 5 hour post-induction; 5) pQE-80L control 5 hour post-induction; 6) Expressed protein after 16 hours post-induction; 7) pQE-80L control 16 hours post-induction.

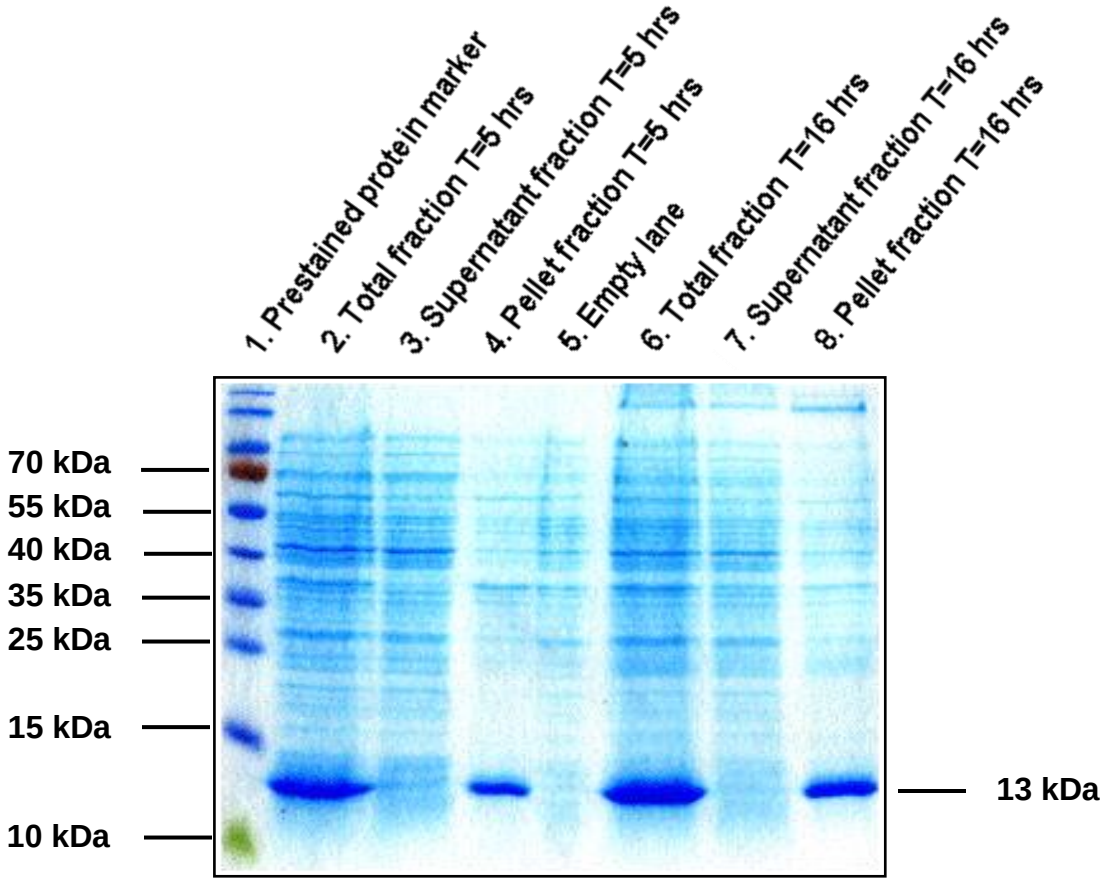


Figure 2.27. SDS-PAGE analysis of proteins at T=5 hours and 16 hours post-induction, with a final IPTG concentration of 2.5 mM, expressed in OverExpress cells containing the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Total fraction proteins expressed after 5 hours induction; 3) Soluble fraction proteins expressed after 5 hours induction; 4) Pellet fraction proteins expressed after 5 hours induction; 5) Empty lane; 6) Total fraction proteins expressed after 16 hours; 7) Soluble fraction proteins expressed after 16 hours; 8) Pellet fractions expressed after 16 hours induction.

Protein concentrations of the four fractions (0.5 ml) collected from the Protino Ni^{+} columns on addition of imidazole (labelled as eluates 1-4) are illustrated in Table 2.14. Protein concentrations were determined from cultures induced with 0.125 mM, 1 mM and 2.5 mM. Confirmation that the protein expressed with an estimated size of 13 kDa was the required EDIII recombinant protein with a His tag was performed using Western blot analysis. An anti-His antibody was used as the primary antibody. Western blot analysis was performed using recombinant protein after purification of the His tagged proteins using Ni-TED columns. A band of ~13 kDa was detected confirming the expression of a His tagged protein as shown in Figure 2.31 (expression induced with 0.125 mM and 1 mM IPTG) and Figure 2.32 (expression induced with 2.5 mM IPTG).

Table 2.14. Protein concentrations of purified recombinant YFV EDIII protein from cultures induced with 0.125 mM, 1mM and 2.5mM IPTG.

Eluates	IPTG 0.125mM	IPTG 1mM	IPTG 2.5mM
1	626 µg/ml	522 µg/ml	992 µg/ml
2	1534 µg/ml	1580 µg/ml	1590 µg/ml
3	548 µg/ml	784 µg/ml	742 µg/ml
4	546 µg/ml	464 µg/ml	688 µg/ml

Additional bands observed at 25 kDa were as a result of non-specific binding. To optimize Western blot analysis additional wash steps were performed to reduce the non-specific binding of the His tag MAb. Due to differences in wells, different amounts were added.

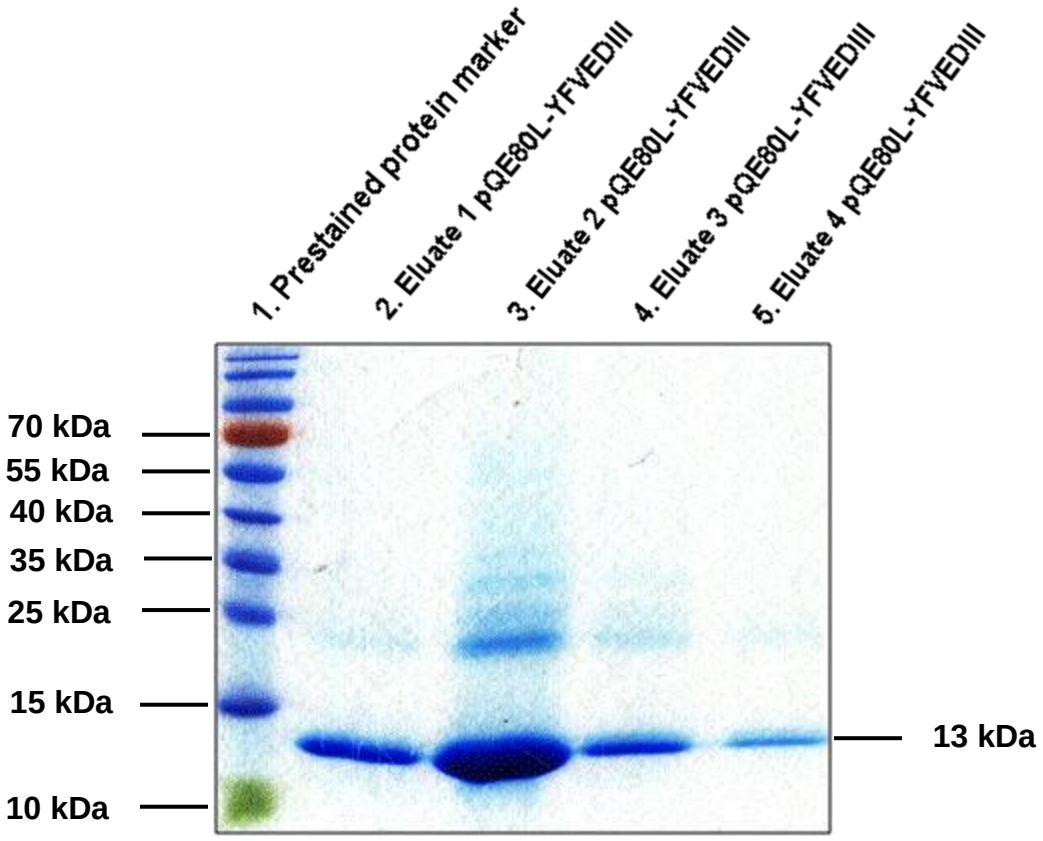


Figure 2.28. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (0.125 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Eluate 1 pQE-80L-YFVEDIII; 3) Eluate 2 pQE-80L-YFVEDIII; 4) Eluate 3 pQE-80L-YFVEDIII; 5) Purified pQE-80L-YFVEDIII protein 4.

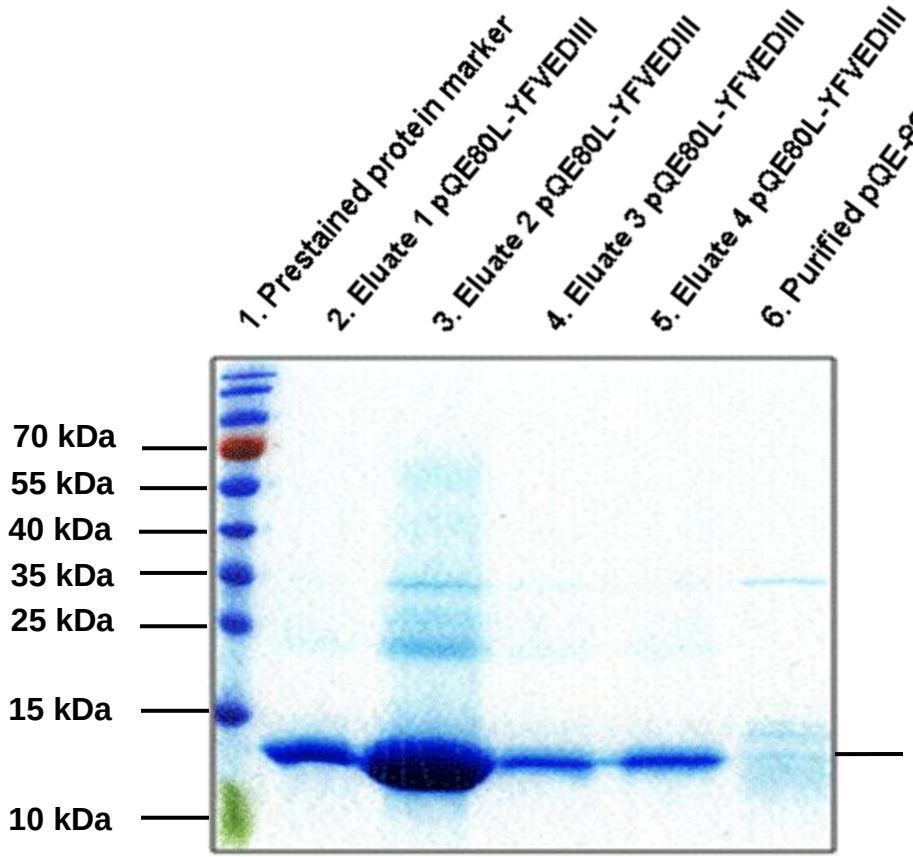


Figure 2.29. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (1 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Eluate 1 pQE-80L-YFVEDIII; 3) Eluate 2 pQE-80L-YFVEDIII; 4) Eluate 3 pQE-80L-YFVEDIII; 5) Eluate 4 pQE-80L-YFVEDIII; 6) Purified pQE-80L control.

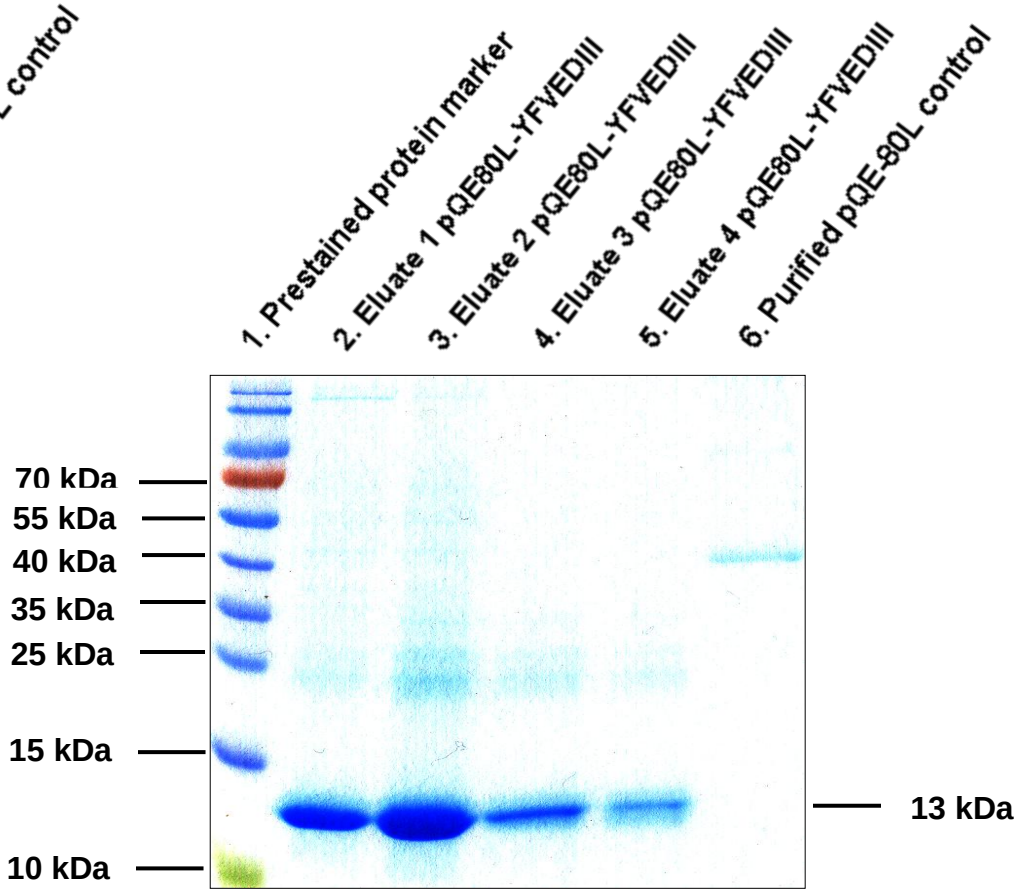


Figure 2.30. SDS-PAGE analysis of purified proteins at T= 16 hours post-induction (2.5 mM IPTG), expressed using the recombinant expression construct, pQE-80L-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Eluate 1 pQE-80L-YFVEDIII; 3) Eluate 2 pQE-80L-YFVEDIII; 4) Eluate 3 pQE-80L-YFVEDIII; 5) Eluate 4 pQE-80L-YFVEDIII; 6) Purified pQE-80L control.

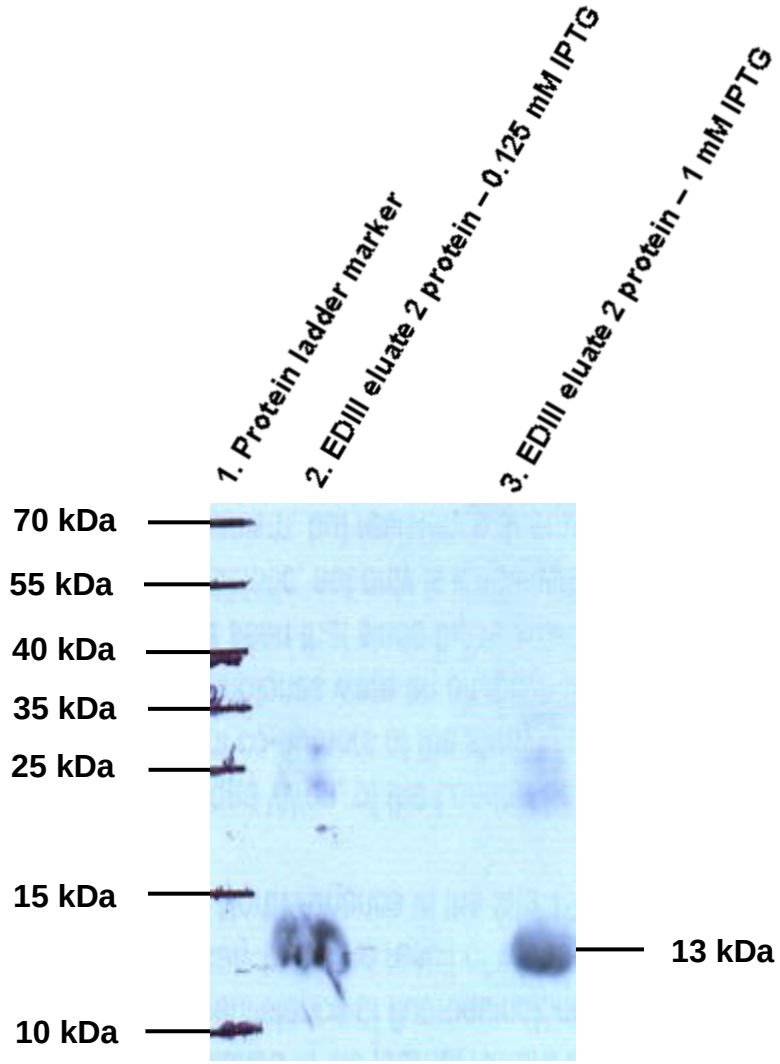


Figure 2.31. Detection of recombinant EDIII protein by Western blot analysis using an anti-His tag antibody. Expressed protein was detected at 16 hours post-induction after purification under denaturing conditions. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 16 hours post-induction with 0.125 mM IPTG; 3) Expressed protein after 16 hours post-induction with 1 mM IPTG.

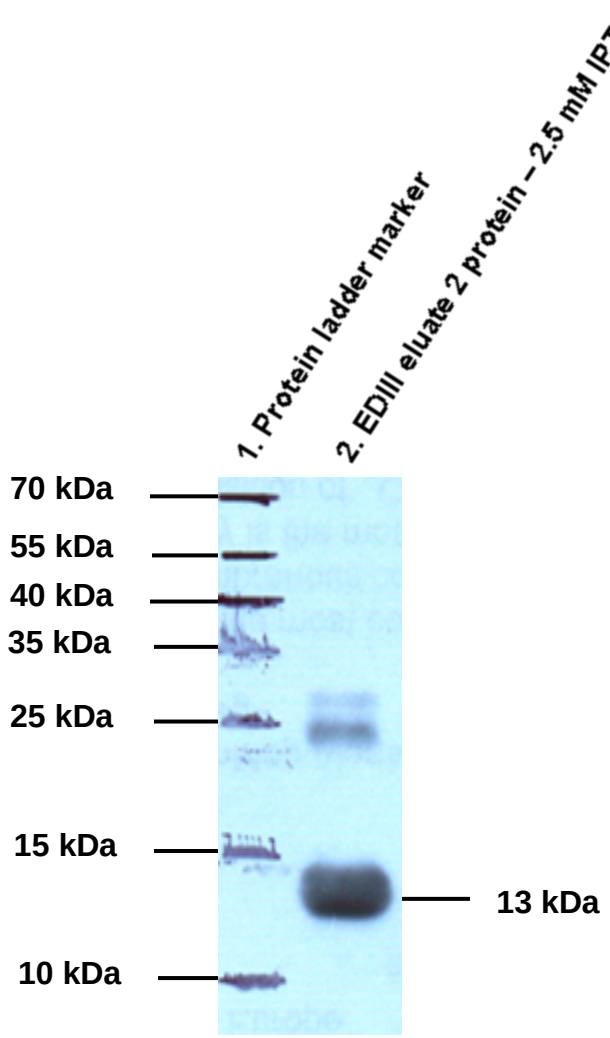


Figure 2.32. Detection of recombinant EDIII protein by Western blot analysis using an anti-His tag antibody. Expressed protein was detected at 16 hours post-induction after purification under denaturing conditions. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Expressed protein after 16 hours post-induction with 2.5 mM IPTG.

2.3.11. Expression, solubility, purification and characterization of expressed EDIII protein using the pCold TF DNA vector

PCR reactions and restriction enzyme digestions were used to identify positive transformants from ligation reactions and transformation cultures using YFV EDIII and pCold TF vector. Recombinant protein expression was induced as described in Materials and Methods 2.2.6 using 1mM IPTG and incubating the cultures at 16ºC. Cells were collected at intervals after induction and analysed for recombinant protein expression as illustrated in Figure 2.33. A ~65 kDa protein was expressed representing the 13 kDa YFV EDIII protein expressed as a fusion protein with the ~52 kDa TF protein.

To determine if the expressed recombinant protein was present in the soluble or insoluble fractions, a protein solubility study was performed. A higher yield of the recombinant YFV EDIII protein (~13kDa) was shown to be expressed in the soluble fraction at 16 and 24 hours after induction (Figure 2.34).

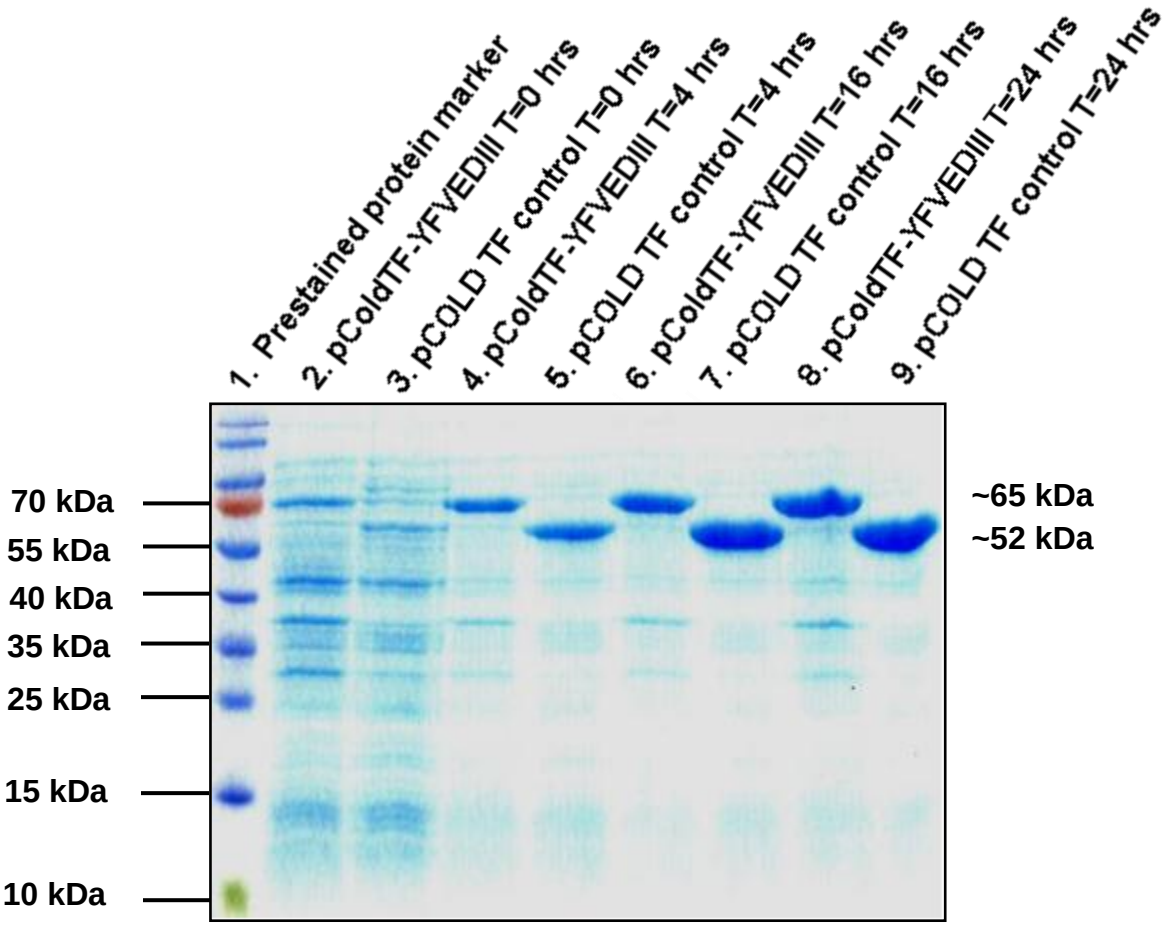


Figure 2.33. SDS-PAGE analysis of proteins expressed at T=0, T=4, T=16 and T=24 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress cells containing the recombinant expression construct, pColdTF-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) pColdTF-TFVEDIII expression 0 hours post-inudction; 3) pCold TF control 0 hour post-induction; 4) pColdTF-YFVEDIII expression 4 hours post-induction; 5) pCold TF control 4 hours post-induction; 6) pColdTF-YFVEDIII expression 16 hours post-induction; 7) pCold TF control 16 hours post-induction; 8) pColdTF-YFVEDIII expression 24 hours post-induction; 9) pCold TF control 24 hours post-induction.

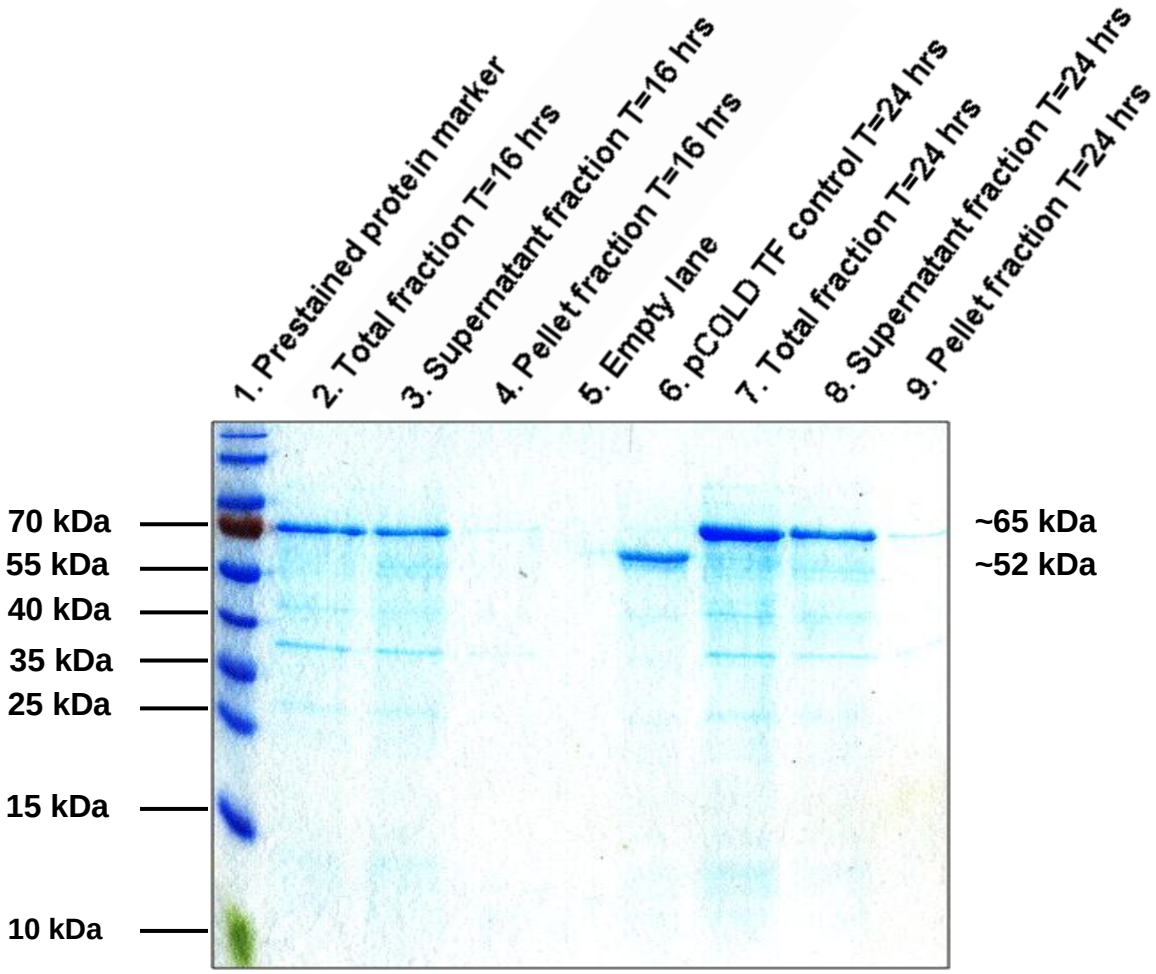


Figure 2.34. SDS-PAGE analysis of proteins at T=16 hours and T=24 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress cells containing the recombinant expression construct, pColdTF-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Total fraction proteins expressed after 16 hours induction; 3) Soluble fraction proteins expressed after 16 hours induction; 4) Pellet fraction proteins expressed after 16 hours induction; 5) Empty lane; 6) pCold TF control after 24 hours induction; 7) Total fraction proteins expressed at 24 hours induction; 8) Soluble fraction proteins expressed after 24 hours; 9) Pellet fractions expressed after 24 hours induction.

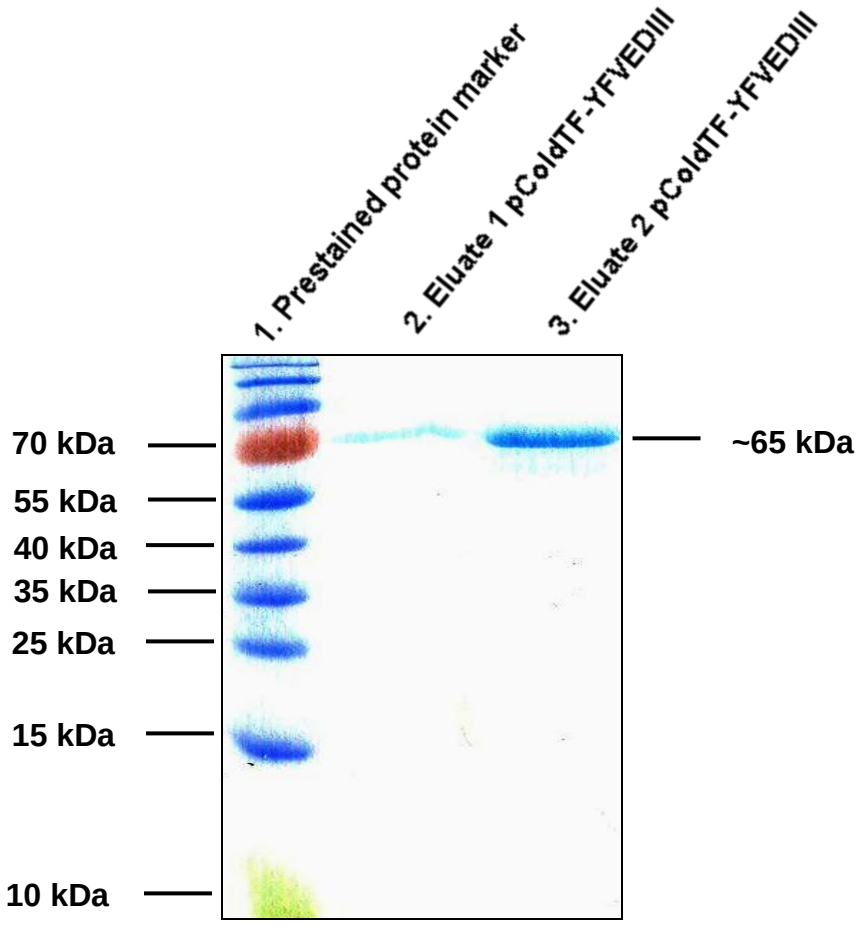


Figure 2.35. SDS-PAGE analysis of purified proteins at T=24 hours post-induction (1 mM IPTG), expressed using the recombinant expression construct, pColdTF-YFVEDIII. 1) Protein molecular weight marker (10kDa – 170kDa); 2) Purified pColdTF-YFVEDIII protein 1; 3) Purified pColdTF-YFVEDIII protein 2.

As the protein solubility study showed that the recombinant YFV pColdTF-EDIII protein was found in the soluble fraction, the protein was purified under native conditions. Briefly, the cells were lysed and purified on a Ni-TED column without using urea as a denaturant. Protein fractions were collected after elution with imidazole (eluates 1-4) and protein concentrations of the four eluates were determined. The protein concentrations were 390 µg/ml (eluate 1), 770 µg/ml (eluate 2), 466 µg/ml (eluate 3) and 528 µg/ml (eluate 4). Purity of the protein was confirmed by SDS-PAGE analysis as shown in Figure 2.35. To confirm the identity of the purified protein a Western blot was performed on eluate 2 for the pColdTF-YFVEDIII protein (data not shown). The anti-His antibody was unable to detect the antigen. After purification the TF fusion tag was cleaved using Factor Xa. Briefly samples were digested with the enzyme for varying incubation times and the digestion products separated using SDS PAGE. Samples

were collected after 3, 6 and 16 hours digestion to monitor optimal digestion of the fusion tag as shown in Figure 2.36

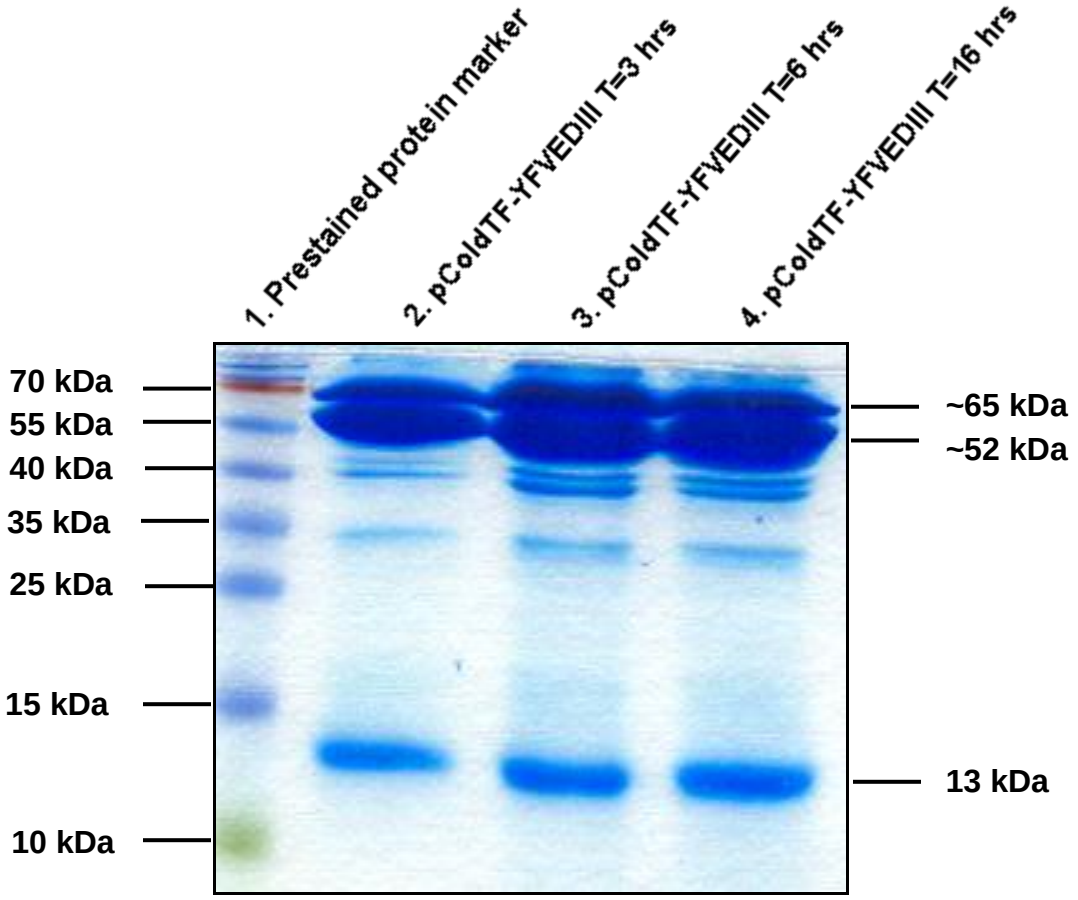


Figure 2.36. SDS-PAGE analysis of purified pColdTF-YFVEDIII protein cleaved using Factor Xa. 1) Prestained protein marker (10kDa – 170kDa); 2) pColdTF-YFVEDIII digested at 3 hours; 3) pColdTF-YFVEDIII digested at 6 hours; 4) pColdTF-YFVEDIII digested at 16 hours.

2.3.12. Expression of the C and NS4a constructs using the pQE-80L vector

The YFV C and YFV NS4a genes were ligated into linearized pQE-80L vector and used to transform OverExpress cells. Various attempts were made to express recombinant YFV C and NS4a proteins. Initially a small scale culture (40 ml) of a positive transformant for each construct was subjected to a final IPTG concentration of 1 mM at 37ºC and the cell lysates collected after various time intervals after induction were analysed by SDS PAGE gel. After a failed attempt expression of each construct was subjected to a final IPTG concentration of 1 mM at 30ºC, 0.05 mM at 37ºC and 0.05 mM at 30ºC.

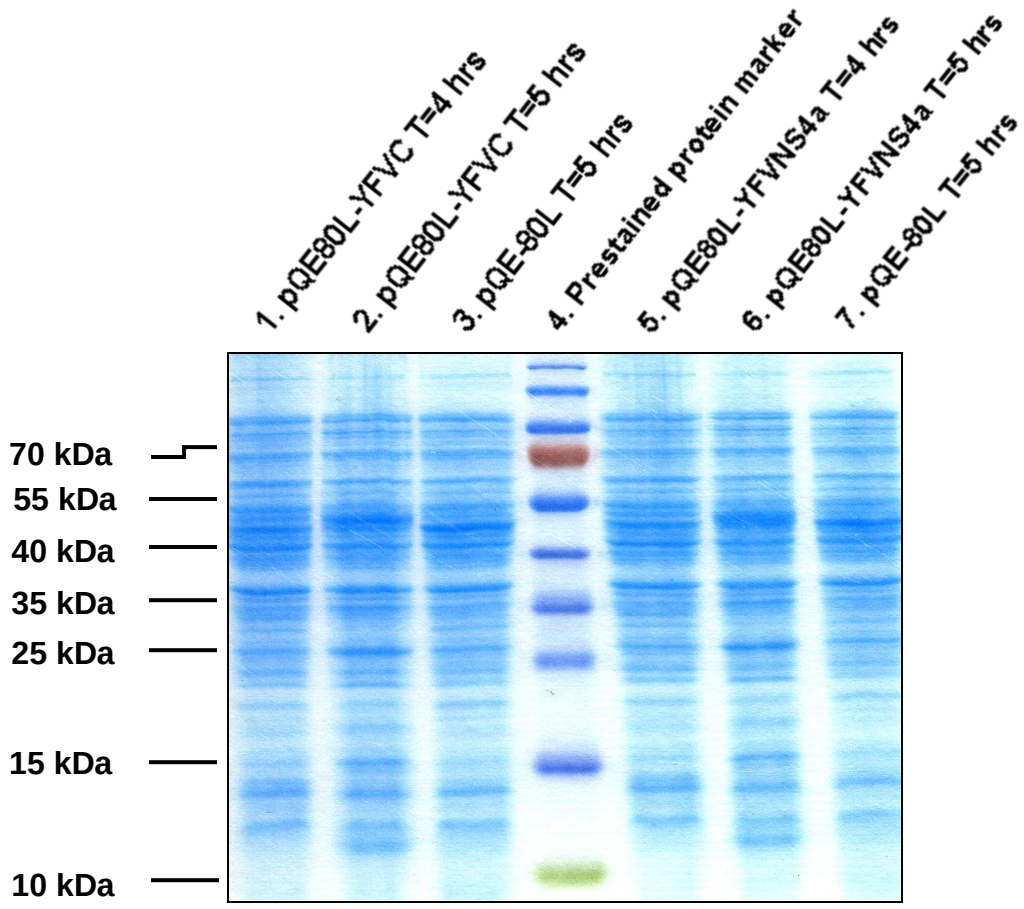


Figure 2.37. SDS-PAGE analysis of proteins expressed at T=4 and T=5 hours post-induction, with a final IPTG concentration of 1 mM, expressed in OverExpress *E.coli* BL21 cells containing the recombinant expression construct, pQE-80L-YFVC and pQE-80L-YFVNS4a. 1) Protein molecular weight marker (10kDa – 170kDa); 2) pQE-80L-YFVC expression 4 hours post-induction; 3) pQE-80L-YFVC expression 5 hours post-induction; 4) pQE-80L-YFVNS4a expression 4 hours post-induction; 5) pQE-80L-YFVNS4a expression 5 hours post-induction.

No detectable expression of recombinant proteins was observed for recombinant YFV pQE-80L-C and recombinant YFV pQE-80L-NS4a protein as shown in Figure 2.37. Western blot would have confirmed if there was low expression however the yield would have been too low for downstream applications. The predicted sizes for the expressed recombinant proteins were 15kDa (pQE-80L-C protein) and 33kDa (pQE-80L-NS4a). It was decided that attempts to express these proteins from native genes using the constructs prepared in this study would be abandoned.

2.4. Summary

The YFV genome encodes for three structural proteins C, prM, E and seven NS proteins, NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5. The E protein is the major structural protein of the virus and is divided into three distinct domains, EDI, EDII and EDIII. EDIII of the E protein is highly antigenic and consists principally of linear-epitopes compared to the C and NS4a proteins that have previously been reported to elicit antibody responses. The complete EDIII, C and NS4a genes were amplified from viral RNA isolated from the YFV 17D vaccine strain. Each amplicon was subsequently cloned into the pGEM®-T Easy vector using T/A cloning and analysed by colony PCR and restriction enzyme digestion for positive transformants. The genes of interest were rescued from the positively transformed constructs for ligation into expression vectors.

Bacterial expression systems are generally preferred for the expression of recombinant proteins because of their ability to grow rapidly and at high density on inexpensive substrates (Terpe, 2006). A large number of cloning vectors and mutant host strains are available for the production of heterologous protein (Terpe, 2006). Researchers have traditionally used *E.coli* expression systems for target protein expression, for high-throughput applications.

The gene encoding the EDIII protein of YFV was cloned into two bacterial expression vectors, namely pQE-80L and pCold TF to compare yield and solubility of recombinant antigens. For high-level expression using the pQE-80L vector, the target gene is cloned downstream of the T5 promoter inducible by IPTG, and a recombinant protein is produced. Altering the experimental conditions such as the temperature and IPTG concentrations did not significantly affect the yield of the recombinant YFV EDIII protein, similarly did not improve the solubility of the protein. The recombinant YFV EDIII protein was expressed more in the insoluble fraction as opposed to the soluble fraction. The recombinant protein was purified by affinity purification using a Protino Ni^{2+} column utilising polyhistidine tagged proteins. A large band representing overexpression of an estimated 13 kDa recombinant purified protein was observed after SDS-PAGE analysis which was confirmed to have a 6 x histidine tag using Western blot analysis.

The pCold™ DNA TF vector uses cold shock technology for expression of the TF chaperone as a soluble fusion tag that should facilitate correct protein folding and improve solubility of expressed proteins. In addition, the expression system uses a cold shock promoter allowing expression to occur at low temperatures (16°C) which slows down cell growth and suppresses the expression of non-specific cellular proteins. To obtain increased amounts of soluble

recombinant protein, co-expression of chaperone proteins during expression has proven to be effective. In this instance a high yield of a soluble recombinant YFV EDIII protein was obtained from the pCold TF construct 16 and 24 hours post induction. After purification, SDS-PAGE analysis confirmed a ~65 kDa recombinant YFV EDIII purified protein. Cleavage with Factor Xa also confirmed that the 65kDa protein was likely a fusion protein comprised of the recombinant EDIII protein and the TF protein.

The genes encoding the C and NS4a proteins of YFV were cloned into the pQE-80L vector. There was no evidence of detectable expression levels observed after attempts to express the proteins. There could be various explanations why these proteins were not expressed including expression of high levels of toxic proteins or possibly use of rare codons that are not recognized by *E. coli* systems. However the purpose of expressing these proteins was to determine if they could be used in detection of YFV antibodies and, if they reacted against polyvalent YFV specific antibodies, then the protein sequences would be used to identify peptide libraries for mapping linear B cell epitopes. It was decided to abandon further attempts to express the C and NS4a proteins and focus on using prediction software to identify peptide libraries for mapping linear B cell epitopes on C, NS4a and EDIII proteins which is discussed in Chapter 3.

In summary, the recombinant YFV EDIII antigens were expressed at high levels from both expression vectors, however the pQE-80L vector expressed the antigen as an insoluble protein compared with the pCold TF vector that expressed it as a soluble protein. The insoluble recombinant protein was purified from insoluble inclusion bodies using a denaturing method followed by refolding of the protein. The soluble TF-EDIII protein was purified without denaturing the protein. The functionality of these proteins as antigens for detection of specific IgG antibody against YFV was investigated and is discussed in the following chapter.

Chapter 3

EVALUATION OF RECOMBINANT YFV EDIII PROTEIN FOR USE IN AN INDIRECT ELISA AND IDENTIFICATION OF LINEAR B CELL EPITOPES USING PEPTIDE LIBREARIES COVERING THE C, EDIII AND NS4a

3.1. Introduction

The prevalence and incidence of YFV cases in Africa and South America is increasing, despite the availability of a safe vaccine (Dash *et al*., 2012). Specific and early detection is important for surveillance and management of disease and outbreaks. Currently diagnosis of YFV is determined routinely using RT-PCR for detection of viral RNA, virus isolation or detection of specific antibody response by immunoassay. Virus isolation readily distinguishes YFV from other flaviviruses, but requires extended incubation in laboratory animals or cell culture that can be labor-intensive and time consuming compared to RT-PCR and serology techniques. Traditional serological techniques for YFV such as HI, IFA and NT can also be labour-intensive and time consuming. As mentioned previously, all flaviviruses are serologically cross-reactive, thus cross-reactivity can make it difficult to distinguish between flaviviruses based on serology particularly in areas where multiple flaviviruses co-circulate i.e. YFV and dengue virus (Monath, 2005). Isolation of YFV and preparation of diagnostic reagents that require culturing of the virus must be performed in BSL three or BSL four facilities due to the biohazardous nature of the virus. Diagnostic kits are not readily available commercially which restricts many laboratories in their ability to rapidly diagnose YFV and to test the immune status of vaccinated individuals.

ELISAs have been applied in a quantitative assay format to monitor immune responses to vaccine (Vazquez *et al*., 2003). Problems associated with laboratory diagnosis such as serological cross reactivity and the interest in developing a safe subunit vaccine increase the importance of identifying immunodominant epitopes. Epitopes are localized regions on the surface of the antigen that elicit an immune response. Antibody defined epitopes are classified as linear (non-conformational) or discontinuous (conformational). A continuous (non-conformational) epitope corresponds to a short continuous stretch of a protein sequence that can induce and bind to an antibody (Van Regenmortel, 2006). Discontinous (conformational) epitopes are composed of different parts of the polypeptide chain that are brought into close proximity by the folding of the protein. Identification and mapping of viral epitopes is important for development of subunit vaccines and improved diagnostics (Wu *et al*, 2003). Prediction of

potentially immunogenic epitopes in a given protein may reduce experimental efforts and costs required for determining epitopes needed for vaccine development and immunodiagnostics. For the detection of antibodies, use of synthetic peptides and recombinant proteins as antigens are alternatives to culturing and purifying native virus eliminating the requirement for sophisticated biosafety facilities and helping to increase diagnostic capacity for biohazardous viruses. In the previous chapter, methods for expressing recombinant proteins were investigated. Recombinant EDIII was expressed although attempts to express the NS4a and C proteins were not successful. In this chapter the functionality of the recombinant EDIII protein in ELISA was investigated and the immunogenicity of the recombinant protein and its ability to induce an immune response that could be detected using native YFV antigen was investigated using an animal model. Secondly, in the absence of recombinant NS4a and C proteins which were not expressed, we used a different approach to identify if these proteins, and the EDIII protein, contain important linear B cell epitopes. To identify epitopic regions we used epitope prediction software and libraries of synthetic peptides.

3.2. Materials and methods

3.2.1. Identification of flavivirus diverse regions

To identify if there are genetically and potentially antigentically diverse regions in the EDIII protein that could be used to develop assays that can serologically differentiate between different flavivirus infections, nucleotide and predicted amino acid sequence data for the EDIII proteins of YFV 17D vaccine strain (GenBank accession number X03700 K02749), WNV lineage 1 strain NY99 (GenBank accession number NC_009942), JEV (GenBank accession number NC_001437) and DENV-2 (GenBank accession number NC_001474) were retrieved from GenBank the data were aligned using Clustal X version 2.1 software.

3.2.2. Human serum samples

A total of 24 serum samples from 24 volunteers with a confirmed history of YFV vaccination between 1985 and 2011 were collected for the study. Informed consent documents were obtained from each volunteer. Estimates of the dates on which the volunteers received their vaccine and the number of boosters received were documented where possible. The complete

cohort of samples were tested for neutralising antibody and IgG antibody against YFV by "in house" ELISA methods developed using recombinant YFV EDIII antigen and YFV whole cell lysate antigen. Twelve of the serum samples were reacted against peptide libraries covering epitope predicted regions in the C, NS4a and EDIII proteins. Serum samples from eight volunteers with no known history of flavivirus vaccination or infection were used as negative controls for the optimization of each ELISA using recombinant YFV EDIII antigen, YFV whole cell lysate antigen and peptide libraries. The panel of negative sera were previously tested by HI at the Centre for Emerging and Zoonotic Diseases, National Institute for Communicable Diseases, Johannesburg and were confirmed to be negative for antibody against YFV, DENV-2 and WNV.

3.2.3. *In-vitro* neutralisation assay

Vero 76 cell cultures (American Tissue Type collection, [ATCC] CRL No. 1587) were grown to confluency and maintained in EMEM with Earle's balance salt solution (Life Technologies, New York, United States of America) supplemented with 2% foetal calf serum (FCS) (Delta Bioproducts, Johannesburg, South Africa), 0.5% gentamycin sulphate (Sigma-Aldrich, Missouri, United States of America) and 0.25% fungizone (amphotericin B).

Determination of tissue culture infectious dose $_{50}$ ($TCID_{50}$)

Virus stocks were prepared from cell cultures infected with YFV strain 17D. Culturing the virus and neutralisation assays were performed within the confines of a BSL two facility. Cells were harvested at the first signs of cytopathic effects (CPE), frozen and thawed and clarified. The supernatant was aliquoted and stored as virus stocks. The $TCID_{50}$ of the virus is the lowest dilution of virus that will cause CPE in 50% of the population of cells. The $TCID_{50}$ of the virus stocks were determined by seeding a 96 well plate with 100 µl of 5 x 10^4 Vero 76 cells per well and titrating a stock of cell culture derived YFV strain 17D in the wells using ten-fold dilutions from 10^0 to 10^6 with EMEM. The test was performed using x6 replicates and 50 µl of each dilution of virus per well. The plate was covered and sealed to prevent drying out and incubated at 37 ºC in CO_2 incubator. The cells were monitored daily for five days to determine the highest dilution of virus that causes CPE. The $TCID_{50}$ was calculated using the Reed-Muench Method.

Neutralisation assay

After determining the $TCID_{50}$, briefly, two fold dilutions of test sera were prepared in duplicate, using serum free media from 1:4 to 1:256, in a 96 well plate. Positive and negative control sera diluted from 1:4 to 1:256 were prepared using serum free media. A 50 µl aliquot of diluted sera was added per well as shown in Table 3.1

Table 3.1. *In vitro* virus neutralisation test using YFV 17D vaccine strain.

Row	Serum dilution
A	1:4
B	1:8
C	1:16
D	1:32
E	1:64
F	1:128
G	1:256
H	No virus

A 50 µl aliquot $100TCID_{50}$ virus was added to each well from rows A-G and 100 µl serum free media was added to the cell control wells in row H. The plates were shaken gently and incubated at 37ºC in a CO_2 incubator for 60 minutes. A 100 µl aliquot of 5×10^4 cells in 6% FCS EMEM media was added to each well. The plates were covered and incubated at 37ºC in CO_2 incubator. The plate was monitored daily and CPE recorded daily for five days. The neutralisation assays were performed by monitoring CPE by observation.

3.2.4. Preparation of YFV cell lysate antigen

A YFV cell lysate antigen was prepared from infected cell cultures. Briefly, eight confluent T150 tissue culture flasks of Vero 76 cells were infected with clarified mouse brain suspension of YFV 17D strain. The 17D strain was selected as it can be handled by vaccinated individuals within the confines of a BSL two laboratory. The cells were incubated at 37ºC for 4-5 days. Percentage infectivity was determined using an IFA assay. Briefly, cells were scraped off the flask with a rubber policeman, the cells were washed in PBS, re-suspended in 10% FCS/PBS, 10 µl aliquots were applied to each well on 8-well slides and the cells dried and fixed by

incubating in cold acetone for a minimum of 20 minutes. Human serum positive for IgG antibody against YFV was diluted 1:10 in PBS and added to each well of cells on the spot slide. The slide was incubated at 37ºC for 20 minute, washed in PBS 3 x 15 seconds, washed in distilled water 1 x 15 seconds, dried and reacted with anti-human IgG fluoroscein isocyanate (FITC) (Zymed Laboratories, Cardiff, United Kingdom) for 30 minutes at 37ºC. The slides were washed as described previously, dried and mounted with glycerol mounting fluid and a coverslip. The percentage of infectivity was estimated visually using a fluorescent microscope (Nikon, Tokyo, Japan). When cells were 80%-90% infected the cells were scraped from the flask and clarified by centrifuging at 10 000 x g for 15 minutes. The supernatant fluid (SNF) was discarded and the cell pellet re-suspended in 5 ml borate buffered saline (BBS) containing 1% Triton X 100. The cell suspension was sonicated for 30 minutes (Branson Sonicator, Smithkline) and clarified as previously described. The SNF was stored at -80ºC for use in ELISA. Mock antigen was prepared by lysis of uninfected Vero cells.

3.2.5. Preparation of the recombinant YFV EDIII antigens using pQE-80L and pCold TF

As described in Chapter 2, the recombinant YFV pQE-80L-EDIII antigen was expressed from pQE-80L-YFVEDIII construct and the recombinant YFV pColdTF-EDIII antigen was expressed from the pColdTF-YFVEDIII construct. Mock pQE-80L and pCold TF antigens were prepared similarly from cultures inoculated with the vector without insertion of the EDIII gene.

3.2.6. ELISA

ELISAs were performed in 96 well microtiter Polysorp plates (Nunc Immunoplate, Roskilde, Denmark), and optimal working dilutions of the reagents were determined by checkerboard titrations. Throughout each assay reagent volumes of 100 µl were used unless stated otherwise, the diluent for reagents was 1 x PBS (pH 7.0) containing 2% skimmed milk powder, incubations were performed for 1 hour at 37ºC, wells were blocked after coating with 200 µl PBS containing 10% skimmed milk powder and plates were washed 3 x 15 seconds with 1 x PBS containing 0.1% Tween 20.

3.2.6.1. ELISA for the detection of IgG antibodies using the recombinant YFV pQE-80L-EDIII antigen

An “in house” ELISA using recombinant YFV pQE-80L-EDIII antigen was developed for the detection of IgG antibodies against YFV. The protein concentration of the recombinant YFV pQE-80L-EDIII antigen was determined as 1140 µg/ml and it was stored for less than a week. Each batch was stored for less than a week. Briefly, a 96 well microtiter PolySorp plate was coated overnight at 4ºC with recombinant YFV pQE-80L-EDIII antigen and mock pQE-80L antigen optimally diluted 1:400 in PBS. After the plates were washed and post-coated (blocked), human serum samples and negative controls, diluted 1:100 in diluent were added to each well in duplicate. The plates were incubated, washed and goat anti-human IgG HRPO (Zymed Laboratories, Cardiff, United Kingdom) diluted 1:2000 was added to each well. After further incubation and washing the positive reactors were visualized using the 2,2’-azino di-ethyl-benzothiazoline-sulfonic acid peroxidase substrate (ABTS) (Kirkegaard and Perry Laboratories, Merryland, United States of America). The plates were incubated at room temperature (22-25ºC) in the dark and the OD values were read at 405 nm. The net OD value of each test and control serum sample was determined. For all ELISAs, freshly prepared recombinant antigen was used because of lack of stability.

3.2.6.2. ELISA for the detection of IgG antibodies using YFV whole viral cell lysate antigen

An “in house” ELISA using YFV whole cell lysate antigen was developed for the detection of IgG antibodies against YFV. The cell lysate antigen is stable under storage over time. Briefly, a 96 well microtiter PolySorp plate coated overnight at 4ºC with YFV whole viral cell lysate antigen diluted 1:500 and mock whole viral cell lysate antigen diluted 1:100 in PBS. After the plates were washed and post-coated (blocked), human serum samples and controls, diluted 1:100 in diluent were added to each well in duplicate. The plates were incubated, washed and goat anti-human IgG HRPO conjugate diluted 1:2000 was added to each well. After further incubation and washing the positive reactors were visualized using ABTS. The plates were incubated at room temperature in the dark and the OD values were read at 405 nm. The net OD value of each test and control serum was determined and the ratio OD of each test serum was determined.

Net OD:

Net OD = OD in wells with virus antigen minus OD in wells with mock antigen

Positive OD/Negative OD ratio for stability analysis: C++ was used as a high positive, C+ as a low positive and C- as a negative control

OD of positive/OD of negative = C++/C- and C+/C-

High positive: C++

Low positive: C+

Negative: C-

Positive/Mean negative OD ratio for peptide analysis:

Positive/Mean negative OD ratio = OD of positive sera in wells with peptides/ mean OD of negative sera in wells with peptide

3.2.7. Determination of ELISA cut-off values

Cut off values which separate positive results from negative results were determined using a panel of eight negative control sera. A cut off value was determined for the YFV pQE-80L-EDIII antigen and the YFV whole viral cell lysate antigen. The ELISA was performed on four separate occasions to obtain values for 64 replicates. The mean net OD and standard deviation (SD) obtained from 64 replicates of the panel of eight negative control sera (two replicates of each sample were tested in each run) was calculated. A cut off value was determined from mean net OD of the negative panel plus 2 SD values.

3.2.8. Characterization of immune response against EDIII

3.2.8.1. Immunisation of mice

Five eight week old NIH female mice were inoculated subcutaneously with 25 µg of purified recombinant YFV pQE-80L-EDIII antigen emulsified with 0.5 ml of TiterMax (R) Gold adjuvant (Sigma-Aldrich, Missouri, United States of America). The mice received a second booster dose two weeks after inoculation and monitored for an additional three weeks before they were anaesthetized using halothane and ex-sanguinated. Blood samples were collected and sera stored at -80ºC until tested. Samples were tested for antibodies using an in vitro neutralisation assay and for antibody in ELISA using using whole cell antigen. The study was approved by the University of the Free State's Animal Ethics Committee (ETOVS NO. 09/06).

Six eight week old NIH female mice were inoculated subcutaneously with YFV vaccine. The mice received a second booster dose two weeks after inoculation and monitored for an additional three weeks before they were anaesthetized using halothane and ex-sanguinated. Blood samples were collected and sera stored at -80ºC until tested. The study was approved by the University of the Free State's Animal Ethics Committee (ETOVS NO. 09/06).

The "in-house" ELISA described in Materials and Methods, 3.2.6 using YFV whole cell lysate antigen was modified to test mouse sera for IgG antibody against YFV. Serum samples from two mice that did not receive YFV antigen were used as negative controls to determine the cut-off.

3.2.9. Stability of bacterially expressed recombinant YFV pQE-80L-EDIII antigen

To determine the stability of the recombinant YFV pQE-80L-EDIII protein over time, ELISA plates were coated with recombinant antigen as described in Materials and Methods 3.2.6.1. Optimal dilutions of protein based on previous checker board titrations and protein concentrations were used to coat the plates. The protein concentration of the recombinant YFV pQE-80L-EDIII protein was 986 µg/ml. The plates were incubated overnight at 4ºC, plates were washed 3 x 15 seconds with 1 x PBS containing 0.1% Tween 20 and wells were blocked with 200 µl PBS containing 10% skimmed milk powder and incubated for 1 hour at 37ºC. Plates were washed and stored in sealed bags at 4ºC. A known high positive serum sample was selected and used as the high positive control (C++), and a known low positive serum sample

was selected and used as the low positive control (C+). A serum sample was selected from a volunteer with no known history of YFV infection or vaccination and used as the negative control (C-). The plates were stored at 4ºC and one plate tested at week 1, week 2, week 5, week 7 and week 19 after preparation.

3.2.10. Mapping linear B cell epitopes

3.2.10.1. Prediction of antigenic and hydrophilic regions

Predicted amino acid sequence data for C, NS4a and EDIII proteins of various YFV vaccine strains were retrieved from GenBank and aligned with Clustal, including YFV vaccine strain 17D (GenBank accession number X03700 k02749), 17DD (GenBank accession number UI7066), 17D-204 (GenBank accession number AF052437) and 17D-213 (GenBank accession number UI7067). Putative epitopic and antigenic regions along the length of the C, NS4a and EDIII proteins of each strain were predicted using the BCPREDS (ailab.cs.iastate.edu/bcpreds/) and ABCpred software (www.imtech.res.in/raghava/abcpred/). Surface hydrophilicity and hydrophobicity were analyzed using the Immune Epitope Database (IEDB) version 2.0 software (www.ImmuneEpitope.org). The Parker Hydrophilicity Prediction software was used to determine regions that were likely to be surface exposed and therefore potentially epitopic.

3.2.10.2. Peptide libraries

Based on outcome from epitope prediction software, a total of 28 overlapping peptides covering predicted antigenic sites on the C, NS4a and EDIII proteins of YFV were synthesized (GenScript, USA). Four overlapping peptides designated C6-C9 represented predicted epitopic regions on the C protein, five overlapping peptides designated NS4a_1-NS4a_5 were derived from the NS4a region and 19 overlapping peptides designated ED10-ED28 were derived from the EDIII region. Each overlapping peptide was nine or eight mers in length and the library offset by three residues.

3.2.10.3. ELISA for epitope identification

An “in house” peptide-based ELISA assay was developed for the detection of IgG antibodies reacting to peptides covering regions on the C, NS4a and EDIII proteins. Twelve serum

samples from vaccinated individuals as described in Materials and Methods, 3.2.2 were tested. Briefly, a 96 well Polysorp plate was coated overnight at 4ºC with 100 µl/well of each peptide (20 µg) diluted in carbonate buffer (pH 9.6). The ELISA was performed as described in Materials and Methods, 3.2.6 except with the addition of 10% bovine serum albumin (BSA)/PBS to replace skimmed milk as a blocking agent and anti-human HRPO was optimally diluted at 1:5000.

3.3. Results

3.3.1. Alignment of coding nucleotide and amino acid sequence data for the EDIII protein of YFV 17D vaccine strain, WNV Lineage 1 strain NY99, DENV-2 and JEV

To identify regions that could potentially be targeted for development of assays to differentiate YFV from other flaviviruses serologically and to show whether nucleotide changes translated into changes in the amino acid sequence, nucleotide and predicted amino acid sequence data for the EDIII protein for WNV Lineage 1 (NY99), DENV-2 and JEV strains were compared to the YFV wild type (Asibi) and YFV 17D vaccine strains. Sequences were aligned using ClustalX as shown in Appendix A. 4. The numbers represent nucleotides from the beginning of the first codon of the gene. Similarly the predicted amino acid sequences were aligned using ClustalX as shown in Appendix A. 5. Few conserved regions were identified along the length of the gene and the translated amino acid sequence. The only region that showed similarity between the different viruses was from amino acid 80 to 90. In this region there were two amino acid differences however the differences were group specific. Sequences were conserved within the YFV group, the JEV group (comprising WNV and JEV) and the DENV group.

3.3.2. Epitope prediction, antigenicity and hydrophilicity

The YFV vaccine strain was selected for the study as it can be handled within the confines of a BSL two laboratory and it was the only strain available in our laboratory. The current vaccine strain that is predominantly used in southern Africa is the YFV 17D-204 vaccine strain. The source of the vaccine received by the vaccinees was unknown hence we were prompted to investigate the differences between the vaccine strains that may impact on the study. To identify differences between the YFV vaccine strains and wild type, nucleotide sequences

retrieved from GenBank were used to predict amino acid sequences of C, EDIII and NS4a proteins. Changes observed at nucleotide level are listed in Table 3.2. Synonymous nucleotide changes observed in the C, EDIII and NS4a proteins are shown in Table 3.2. Appendix A. 6a-6c shows nucleotide sequence alignment of four YFV vaccine strains and the wild type strain. Appendix A. 6a shows the nucleotide sequence alignment of the gene encoding the C protein. Similarly, amino acid sequences of the C, EDIII and NS4a genes were aligned for comparison (Appendix A. 7). No amino acid changes were observed in the C protein. In the EDIII, there were amino acid changes that were present throughout the vaccine strains at positions 7, 13, 39 and 87 relative to the Asibi wild type and one amino acid change at position 33 that was present in the vaccine strains 17D-204, 17D-213 and 17D. In the NS4a protein there was an amino acid change at position 106 in 17D only and amino acid changes at positions 146 and 244 that were conserved in all the vaccine strains relative to Asibi. In summary only one change was identified in 17D strain, which was the strain used to predict the epitopes and used to prepare the recombinant antigen, compared with vaccine strain 17D-204 which is likely the vaccine strain used to immunize the vaccinees in the study. When considering the epitope prediction sites (Table 3.3) position 106 on the NS4a protein was not within the predicted region.

Table 3.2. Nucleotide and amino acid sequence changes among YFV vaccine strains 17D, 17DD, 17D-213, 17D-204 and the YFV Asibi (wild type) strain.

Gene	Nucleotide position	Wild type	17D	17DD	17D-213	17D-204	Amino acid change
C	183	**G**	A	A	A	A	-
	249	**T**	C	T	C	C	-
EDIII	21	**G**	A	A	A	A	M → I*
	38	**C**	T	T	T	T	S → F
	97	**C**	T	**C**	T	T	P → S
	116	**A**	G	G	G	G	K → R
	261	G	G	**A**	G	G	-
	263	**C**	G	G	G	G	T → R
NS4a	9	**G**	T	T	T	T	-
	75	T	T	**C**	T	T	-
	90	T	**C**	T	T	T	-
	186	**A**	**A**	C	C	C	-
	319	A	**G**	A	A	A	I → V
	436	**T**	C	C	C	C	V → A
	732	**A**	G	G	G	G	I → M

*Methionine (M), Isoleucine (I), Serine (S), Phenylalanine (F), Proline (P), Lysine (K), Arginine (R), Threonine (T), Valine (V), Alanine (A).

BCPreds and ABC Linear epitope prediction software was used to identify epitopic regions of the proteins as shown in Table 3.3. From the prediction analyses, propensity values are assigned to each amino acid in the sequence and predicted B cell epitopes were ranked according to their propensity value. These propensity values are based on their physico-chemical properties. The location of continuous or linear epitopes can be predicted using parameters such as hydrophilicity and hydrophobicity, flexibility and accessibility of polypeptide chains. For each parameter a window of seven was selected which assigns a value to each amino acid and includes three amino acids on either side (a total of seven amino acids). This takes into account that epitopes are composed of more than one amino acid. An antibody inducing epitope is usually a minimum of five amino acids in length hence regions less than five amino acids were excluded. The selection of regions for which peptide libraries were identified was based on amino acid length and score. For each protein the regions that scored the highest on both prediction software algorithms were selected to investigate their biological functionality as an epitopic region.

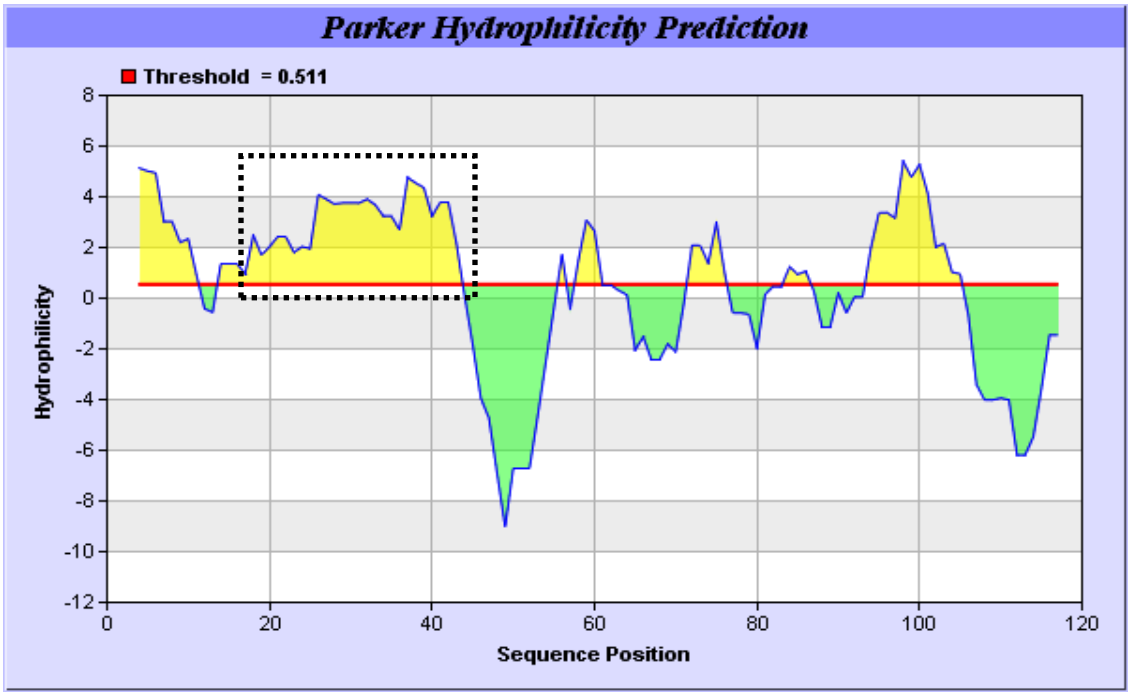
Table 3.3. Common regions identified from linear epitope prediction methods as likely epitopic and antigenic regions on the C, NS4a and EDIII proteins.

	YFV vaccine strains 17D, 17DD, 17D-213, 17D-204	**Amino acid position relative to 17D strain**	**Score range**
C amino acid sequence	IKQKTKQIGNRPGPSRG[a] VQG[b]	26 – 45	0.86-0.999
NS4a amino acid sequence	I PSSASPWSWPDLDLKP GAA	169 – 188	0.92-0.957
EDIII amino acid sequence	VTVNPIASTNDDEVLIEVNPPFGDSYIIV GRGD SRLTY	58 – 95	0.87-0.997
EDIII amino acid sequence	PTDTG HGTVVMQVKVSKGAPC RIPVIVA DDLTAA	18 – 51	0.9-0.969

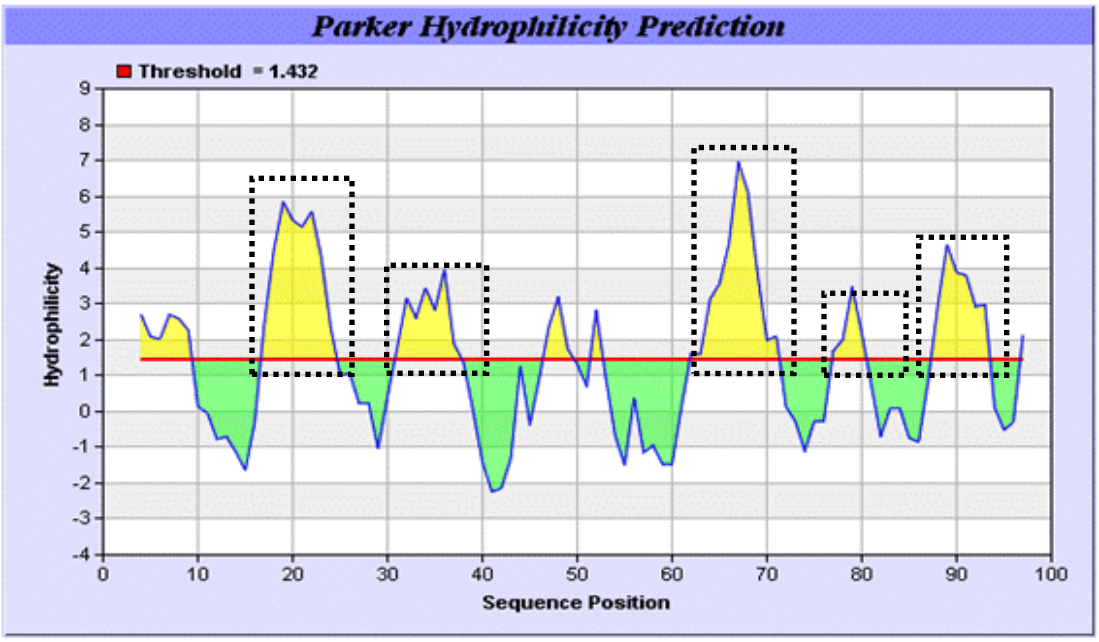
BCPred predictions are shown in boxes. Additional amino acids with high prediction values from ABC prediction software are included in Table 3.3 outside of boxes. [a]Predictions using the BCPreds Linear epitope prediction software. [b]Predictions using the ABC Linear epitope prediction software

Parker Hydrophilicity plots using the amino acid sequence data for NS4a, C and EDIII proteins were generated using the IEDB software as shown in Figure 3.1. Hydrophilic regions are indicated with positive scores shown in yellow. Regions highlighted in boxes were identified as hydrophilic regions that correlated with epitope predictions. A hydrophilic region at the 5' amino (N)-terminal end of the C protein correlated with a predicted epitopic region at amino acid residues 26-45 (Figure 3.1 A.). The NS4a protein had a hydrophilic region (Figure 3.1 C.) correlating with the predicted epitopic region at amino acid residues 169-188. The EDIII protein had five hydrophilic regions on the EDIII protein at amino acid residues 17-25, 30-39, 63-71, 76-81 and 86-95 (Figure 3.1.B). Each of these five regions correlated with the potentially epitopic sites identified using the ABC and BCPreds software from amino acid residues 18-51 and 58-95. Consequently, based on the prediction analyses and hydrophilicity plots, 28 8-mer overlapping peptides offset by three residues were synthesized, corresponding to epitopic and hydrophilic regions of C, NS4a and EDIII proteins.

A.



B.



C.

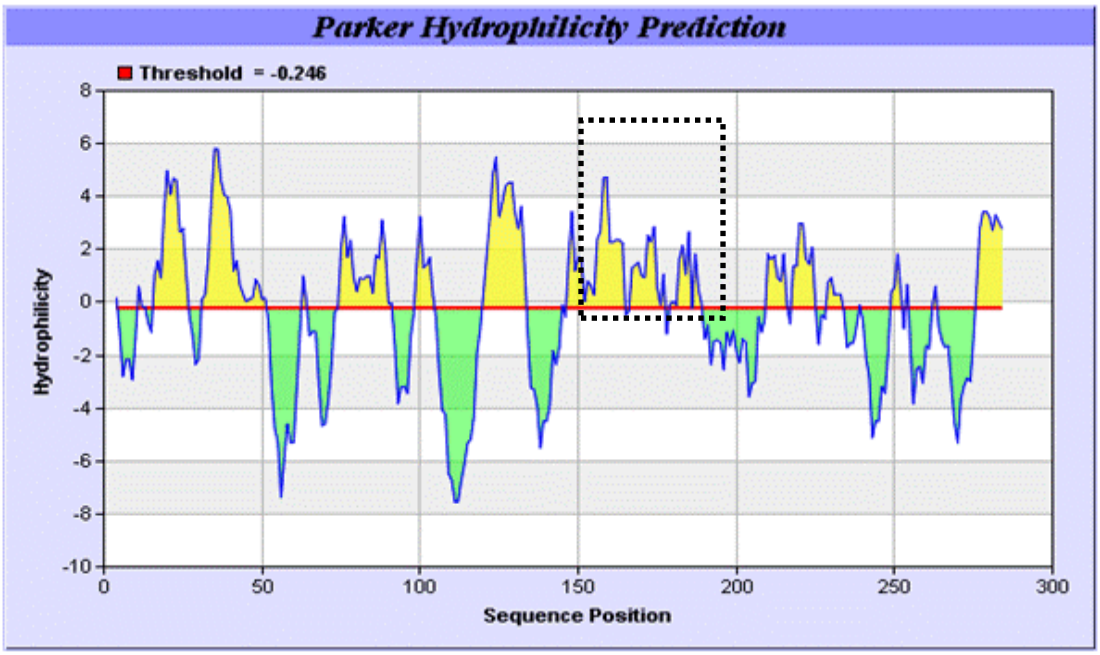


Figure 3.1. Parker Hydrophilicity Prediction plots of the A. C, B. EDIII and C. NS4a proteins (YFV 17D vaccine strain) generated using the IEDB software. Hydrophilic regions shown in boxes correlated with epitopic regions identified using the ABC and BCPreds prediction software. The vertical axis indicates the hydropathy score and the horizontal axis indicates the amino acid residue positions. Hydrophilicity plots are indicated with positive scores and hydrophobicity plots are indicated with negative scores.

3.3.3. Virus neutralisation

Serum samples were tested for neutralisation antibody against YFV to confirm the immune status of each sample. Virus neutralisation assays are the gold standard for serological assays for flaviviruses. A total of 24 sera from YFV vaccinees as shown in Table 3.4 were tested. Four volunteers provided a record that they had received additional boosters at intervals, as indicated in Table 3.4, whereas the remainder received only one vaccination. In some instances the vaccinees were uncertain of exact dates. The VBD number as shown in Table 3.4 is the laboratory number assigned to each serum sample. All the sera had a detectable neutralising antibody response to YFV. Sera from patients with no history of vaccination did not inhibit CPE in the neutralisation assay. The neutralisation antibody titer is the reciprocal of the dilution at which 50% CPE is inhibited. Neutralisation antibodies were determined to confirm that vaccinees had a detectable immune response against vaccinees. Samples were subsequently tested for antibodies using an in house ELISA. The study was not done to investigate protective immunity.

Table 3.4. Serum samples from volunteers with a history of YFV vaccination and antibody results.

VBD number	Year of vaccination	Neutralisation antibody"	IgG ELISA cell lysate antigen	IgG ELISA recombinant EDIII antigen
VBD 30/08	N/A[#]	4	Positive	Positive
VBD 56/08	2008	8	Positive	Positive
VBD 19/09	1995/2000/2009[*]	16	Positive	Positive
VBD 35/09	2008	4	Positive	Positive
VBD 36/09	2008	4	Positive	Positive
VBD 41/09	1965	4	Positive	Positive
VBD 42/09	1991	4	Positive	Negative
VBD 46/09	>10 years ago	4	Positive	Positive
VBD 47/09	N/A	4	Positive	Positive
VBD 48/09	2005	4	Positive	Negative
VBD 49/09	2006	8	Positive	Positive
VBD 50/09	2007	8	Positive	Positive
VBD 51/09	2008	8	Positive	Positive
VBD 54/09	2009	4	Positive	Positive
VBD 55/09	2009	4	Positive	Positive
VBD 65/09	2005	4	Positive	Positive
VBD 66/09	2002/2005	8	Positive	Positive
VBD 7/11	N/A	4	Positive	Negative
VBD 5/12	2011	4	Positive	Positive
VBD 6/12	2006/2011	4	Positive	Negative
VBD 7/12	2003	4	Positive	Positive
VBD 12/12	2007	4	Positive	Positive
VBD 13/12	1985/1998/2005/2007/2008	8	Positive	Positive
VBD 44/12	2007	16	Positive	Positive

[#]N/A not available, [*]years in which YFV vaccine boosters were received, "titer reciprocal of end point dilutions at which CPE is inhibited

3.3.4. Detection of antibody against YFV using ELISA and whole cell lysate antigen

To determine the IgG antibody reactivity against the YFV whole cell lysate antigen, a total of 24 sera from 24 vaccinees were tested. The protein concentrations of the YFV whole cell lysate antigen and mock cell lysate antigen were 4840 µg/ml and 4140 µg/ml respectively. A cut-off value was determined using a panel of eight negative serum samples from volunteers with no known history of flavivirus vaccination or infection. The cut-off was calculated from the mean net OD + 2 SD and was 0.1474. All the sera were positive for IgG antibody in the ELISA using YFV cell lysate antigen. The results are included in Table 3.4 and illustrated in Figure 3.2. All sera tested had a demonstrable IgG antibody response against YFV.

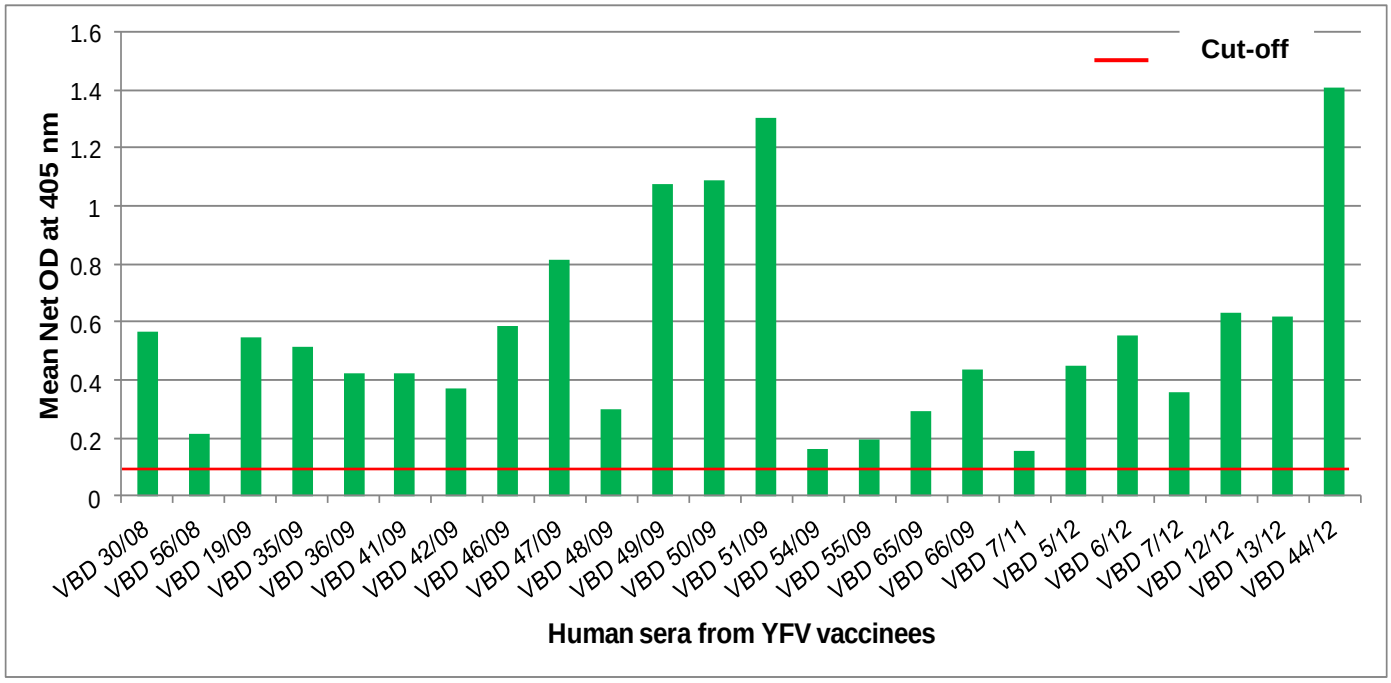


Figure 3.2. Detection of IgG antibodies against YFV using whole cell lysate antigen vaccines.

3.3.5. Detection of antibody against YFV using ELISA with recombinant YFV pQE-80L-EDIII antigen

As described in Chapter 2, a recombinant YFV EDIII antigen was expressed in a bacterial system using pQE-80L expression vector and was subsequently used for the development of an ELISA for detection of IgG antibodies in YFV vaccinees. The protein concentration of the

recombinant YFV pQE-80L-EDIII antigen was determined as 1140 µg/ml. To determine the IgG antibody reactivity against the recombinant YFV pQE-80L-EDIII antigen, a total of six serum samples from volunteers with no known history of flavivirus vaccination or infection were selected to determine the cut-off. The net OD values for the determined using the negative panel were as follows: -0.085, -0.1395, 0.0015, -0.03, -0.032 and 0.014 respectively. The mean was calculated as -0.04517 and the SD set as 0.05756. The cut-off was calculated as the mean + 2 SD and was determined as 0.0699. Figure 3.3 shows a total of 20/24 samples tested were positive in ELISA using the recombinant YFV pQE-80L-EDIII antigen. It is unclear why four sera did not react with the EDIII antigen. There was no correlation between the negative result and date of immunization (that is the duration after vaccination that samples were collected) or the number of boosters received by the individual.

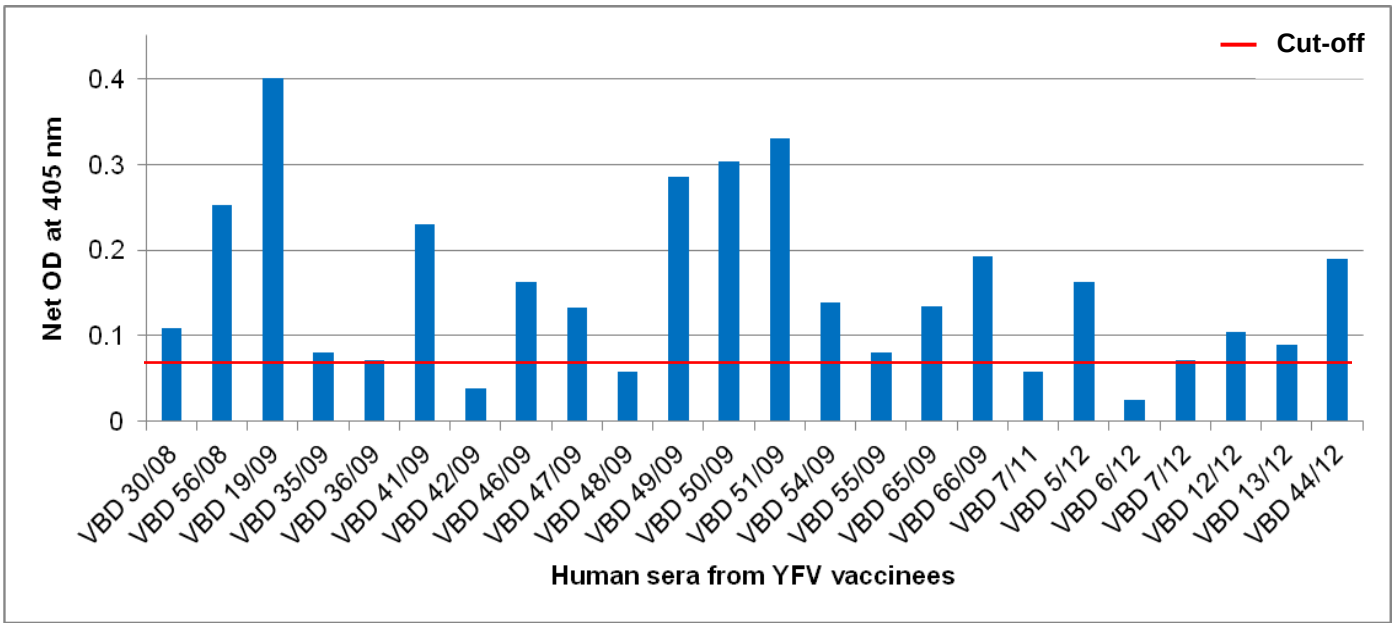


Figure 3.3. Detection of IgG antibodies against recombinant YFV pQE-80L-EDIII antigen using human sera from vaccinees.

3.3.6. Detection of antibody against recombinant YFV pColdTF-EDIII antigen in human sera

As described in Chapter 2, a second bacterial system, pCold TF expression vector was used to express the recombinant YFV EDIII antigen and the antigen was used for the development of an

ELISA for detection of IgG antibodies. The protein concentration of the recombinant YFV pColdTF-EDIII protein was determined as 2000 µg/ml and the pColdTF mock antigen as 1434 µg/ml. To determine the IgG antibody reactivity against the recombinant YFV pColdTF-EDIII antigen, a high positive C++ serum sample (VBD 19/09) and negative serum sample were used to analyze the activity of the antigen in ELISA. Dilutions were performed in duplicate. A negative serum sample from a volunteer with no known history of flavivirus vaccination or infection was selected to determine the cut-off. The cut-off was calculated as the mean + 2 SD and was determined as 0.225. Figure 3.4 shows that the positive sample tested at dilutions 1:320, 1:640 and 1:1280 were negative in ELISA using the recombinant YFV pColdTF-EDIII antigen. Removal of the TF protein did not improve activity of the antigen in ELISA and the pQE-80L antigen was used in subsequent analyses. In addition, since no antibody response was detected stability studies were not performed.

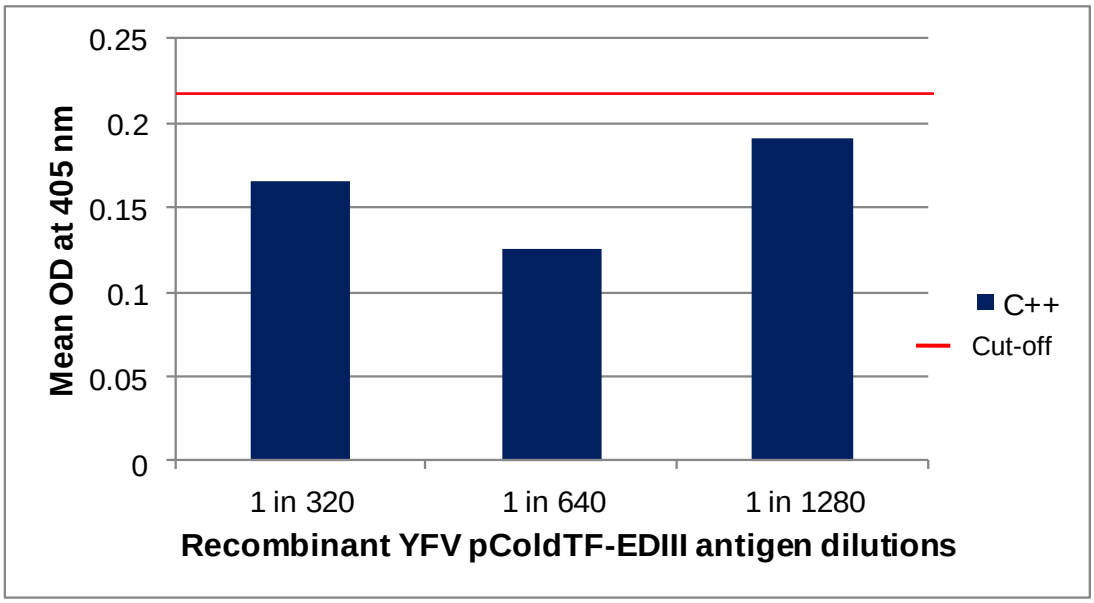


Figure 3.4. Detection of IgG antibodies against recombinant YFV pColdTF-EDIII antigen.

To determine the stability over time of the bacterially expressed recombinant YFV EDIII antigen, pre-coated plates were stored at 4ºC and tested at intervals with C++, C+ and C-. The ratio of net OD value for C++ and C+: net OD value for C- was determined and plotted in Figure 3.5. The antibody reactivity against the recombinant YFV pQE-80L-EDIII antigen decreased over the 4 months. The protein was unstable on precoated plates with a drop in the ratio of the positive

to negative OD values at 5 weeks after preparation of the antigen. A decrease in the ratio over time suggests that it would be difficult to discriminate between positive and negative sera particularly with low titered sera. This assay was repeated (data not shown) and different batches were tested.

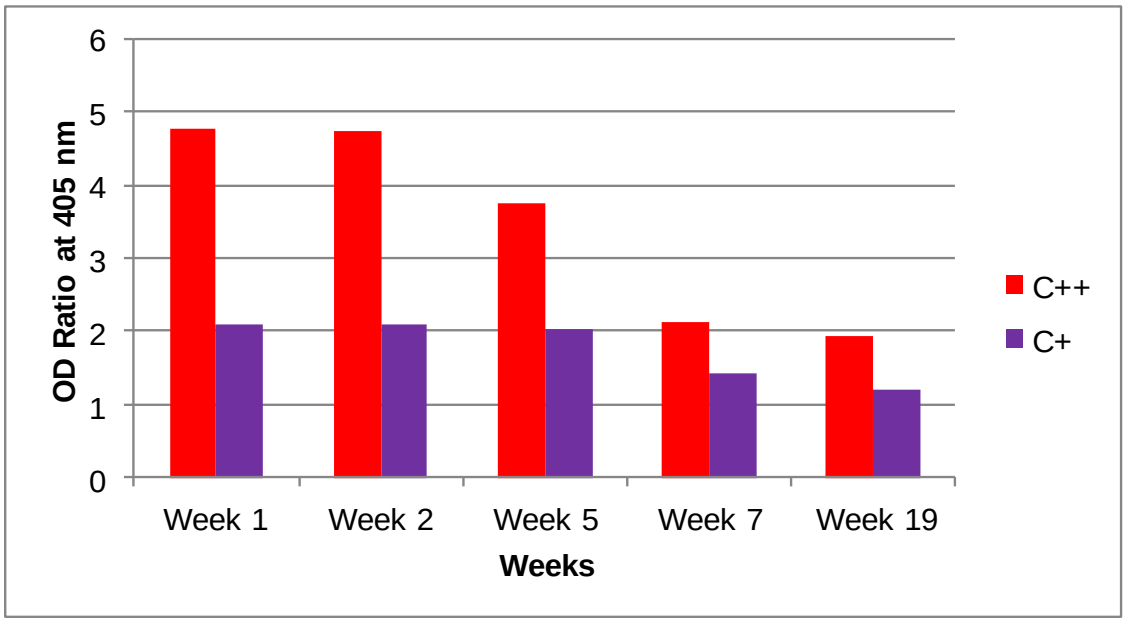


Figure 3.5. Assessment of the biological activity of the recombinant YFV pQE-80L antigen over four months.

3.3.7. ELISA for detection of IgG antibodies in immunized mice

To confirm that the ELISA using whole cell lysate was able to detect an IgG antibody response in mice, sera collected from mice immunized with the vaccine were tested. As described in Materials and Methods 3.2.8.1, six NIH female mice were immunized with the vaccine. The cut-off value was determined as the mean net OD + 3 SD and set as 0.037. The immune responses were assessed after two consecutive boosters and results showed that 6/6 mice had detectable antibody responses (Figure 3.6).

To determine if the recombinant YFV pQE-80L-EDIII antigen induces an antibody response that can be detected using YFV whole cell lysate antigen, an animal model was used. A total of five eight week old NIH female mice were immunized with the recombinant YFV pQE-80L-EDIII antigen as described in Materials and Methods 3.2.8.1 and serum samples tested for IgG

antibody responses using an ELISA and whole cell lysate antigen. The cut-off value was determined as the mean net OD +3SD and set as 0.037. An IgG antibody response was detectable in 5/5 sera (Figure 3.6).

To determine if the recombinant YFV pQE-80L-EDIII antigen induced a neutralising antibody response in mice, mouse sera were tested for neutralising antibodies and 3/5 sera had demonstrable neutralising activity (Appendix B. 1.A and B.). Hence the recombinant YFV pQE-80L-EDIII antigen was able to induce neutralising antibodies in mice, although not all the mice were positive this could likely be related to dose.

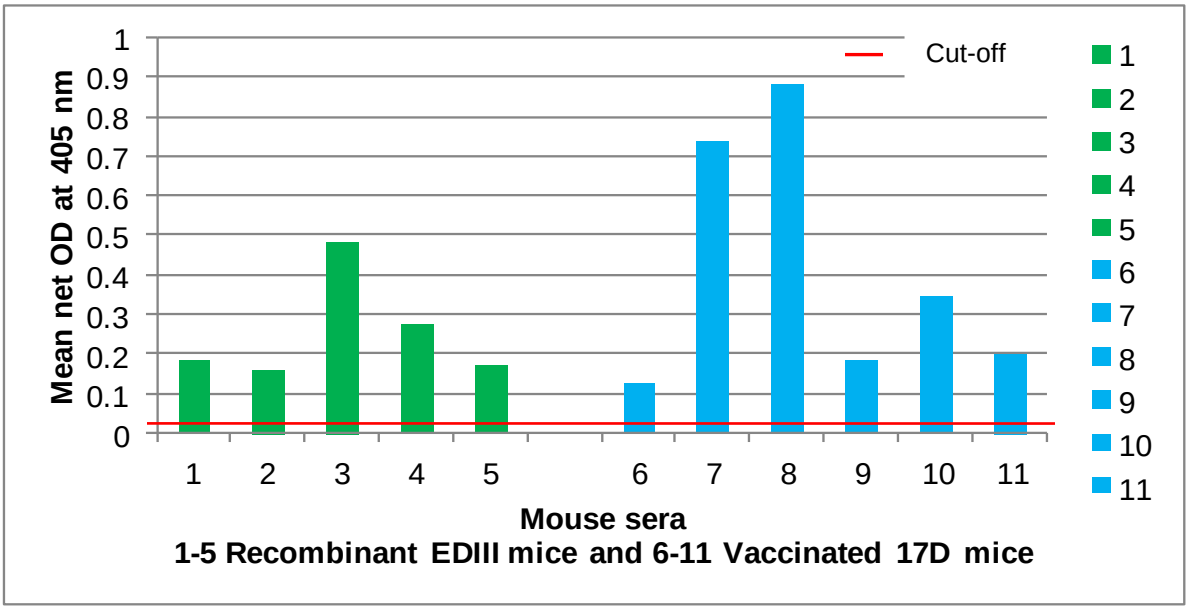


Figure 3.6. Antibody reactivity against YFV whole cell lysate antigen using sera from mice immunized with the YFV 17D vaccine and recombinant YFV pQE-80L-EDIII antigen.

3.3.8. Serological cross-reactivity between YFV and WNV

The serological cross reactivity of flaviviruses is well known. Previous reports have suggested that the recombinant EDIII proteins can be used to distinguish between antibodies against mosquito-borne and tick-borne viruses. To investigate if the recombinant YFV EDIII antigen could distinquish between antibodies against YFV and WNV, another mosquito-borne flavivirus, serum samples from horses that had suspected WNV infections and a known negative were tested. The sera were tested against whole cell lysate antigen and recombinant EDIII antigen.

Four samples were available, one had previously been shown to have had a WNV infection (horse two), horses one and three had suspected infections and horse four was negative. The negative serum sample from a horse with no known history of flavivirus infection was used to determine the cut-off. The cut-off value was determined from the mean net OD + 2 SD and set as 0.2054. As shown in Figure 3.7, horse one and horse two reacted positively against the YFV whole cell lysate antigen. To analyze the cross-reactivity the horse sera were tested using recombinant YFV pQE-80L-EDIII antigen and pQE-80L mock antigen. The cut-off value was determined from the negative serum as the mean + 2 SD and set as 0.2. The line (shown in red) represents the cut-off for YFV whole cell lysate antigen and recombinant YFV pQE-80L-EDIII antigen. As shown in Figure 3.7 horses one and two had no detectable antibody responses suggesting that this antigen has potential to be used to serologically differentiate YFV from WNV.

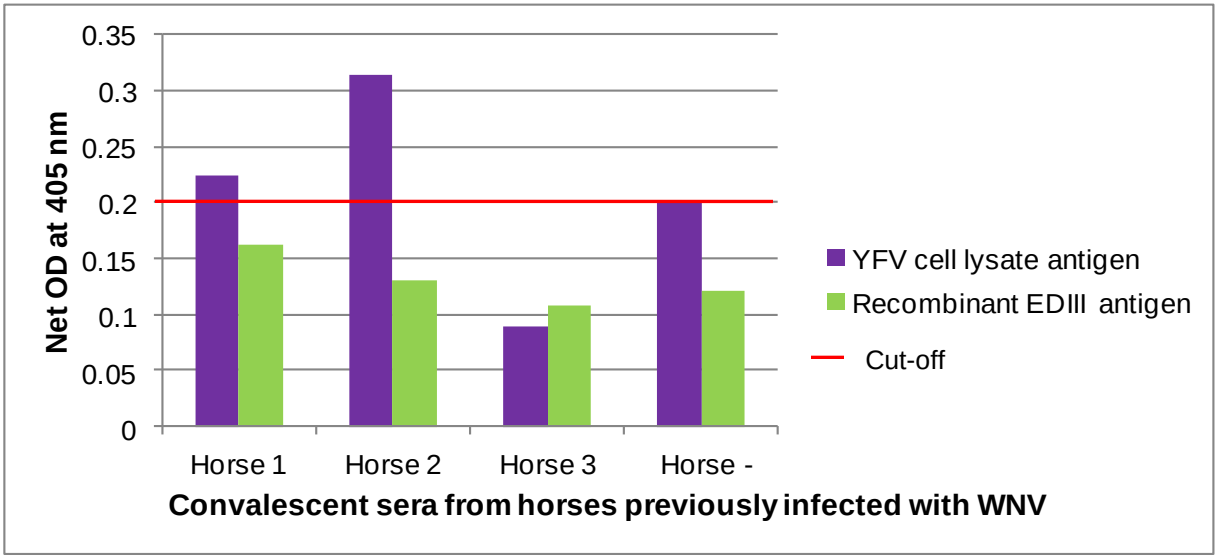


Figure 3.7. Detection of IgG antibodies against YFV cell lysate and recombinant YFV pQE-80L-EDIII antigens using convalescent sera form WNV confirmed horses.

3.3.9. Detection of antibody against C, NS4a and EDIII peptide libraries

A library of 28 overlapping peptides designated NS4a_1 – NS4a_5, C6 – C9 and ED10 – ED28 were synthesized spanning the predicted epitopic regions of the NS4a, C and EDIII proteins (Table 3.5). Peptides were synthesized by GenScript (purity 70%). The peptides were tested in

an ELISA format for reactivity against sera from YFV vaccinees to identify peptides that react which may represent linear B-cell epitopes.

Table 3.5. Peptide libraries of 8 – 9 mers with 3 offset residues were synthesized. The libraries spanned predicted epitopic regions on C, NS4a and EDIII proteins.

Peptide #	Peptide sequence (amino acid residue position relative to the YFV 17D strain)	Peptide length
NS4a_1	IPSSASPWS (169 – 177)	9
NS4a_2	SASPWSWPD (172 – 180)	9
NS4a_3	PWSWPDLDL (175 – 183)	9
NS4a_4	WPDLDLKPG (178 – 186)	9
NS4a_5	LDLKPGAA (181 – 188)	8
C6	KTKQIGNRP (31 – 39)	9
C7	QIGNRPGPS (34 – 42)	9
C8	NRPGPSRGV (37 – 45)	9
C9	GPSRGVQG (40 – 47)	8
ED10	VTVNPIAST (58 – 66)	9
ED11	NPIASTNDD (61 – 69)	9
ED12	ASTNDDEVL (64 – 72)	9
ED13	NDDEVLIEV (67 – 75)	9
ED14	EVLIEVNPP (70 – 78)	9
ED15	IEVNPPFGD (73 – 81)	9
ED16	NPPFGDSYI (84 – 92)	9
ED17	FGDSYIIVG (87 – 95)	9
ED18	SYIIVGRGD (90 – 98)	9
ED19	IVGRGDSRL (93 – 101)	9
ED20	RGDSRLTY (96 – 103)	8
ED21	PTDTGHGTV (18 – 26)	9
ED22	TGHGTVVMQ (21 – 29)	9
ED23	GTVVMQVKV (24 – 32)	9
ED24	VMQVKVPKG (27 – 35)	9
ED25	VKVPKGAPC (30 – 38)	9
ED26	PKGAPCRIP (33 – 41)	9
ED27	APCRIPVIV (36 – 44)	9
ED28	RIPVIVA (39 – 45)	7

The results of antibody reactivity to the synthetic overlapping peptides are summarized in Figure 3.8. A mean negative OD was calculated for all the peptides using negative sera and determined as 0.209. For each sample a ratio of positive:mean negative OD (0.209) was determined and plotted as shown in Figure 3.8. Figure 3.8 shows the IgG specific activities of synthetic peptides using serum samples from vaccinated individuals. Using a ratio of seven as a cut-off, which was selected to exclude background reactivity, the ELISA results showed that the highest reactivity was against peptides 6 and 18-23. However when looking at results from individual sera it is evident that in some instances there were only one or two sera that reacted with the peptide giving the high ratio in Figure 3.8 and hence the majority of these reactions are unlikely to be important. The most significant finding was reactivity of 8/12 sera against peptide 22. Peptide 22 represented amino acids TGHGTVVMQ, position 21 – 29 on the EDIII protein. In addition, mouse sera were tested, however the positive control suggested that the peptides had deteriorated, likely due to repeated freeze thaw cycles. Hence results considered indeterminant and not included.

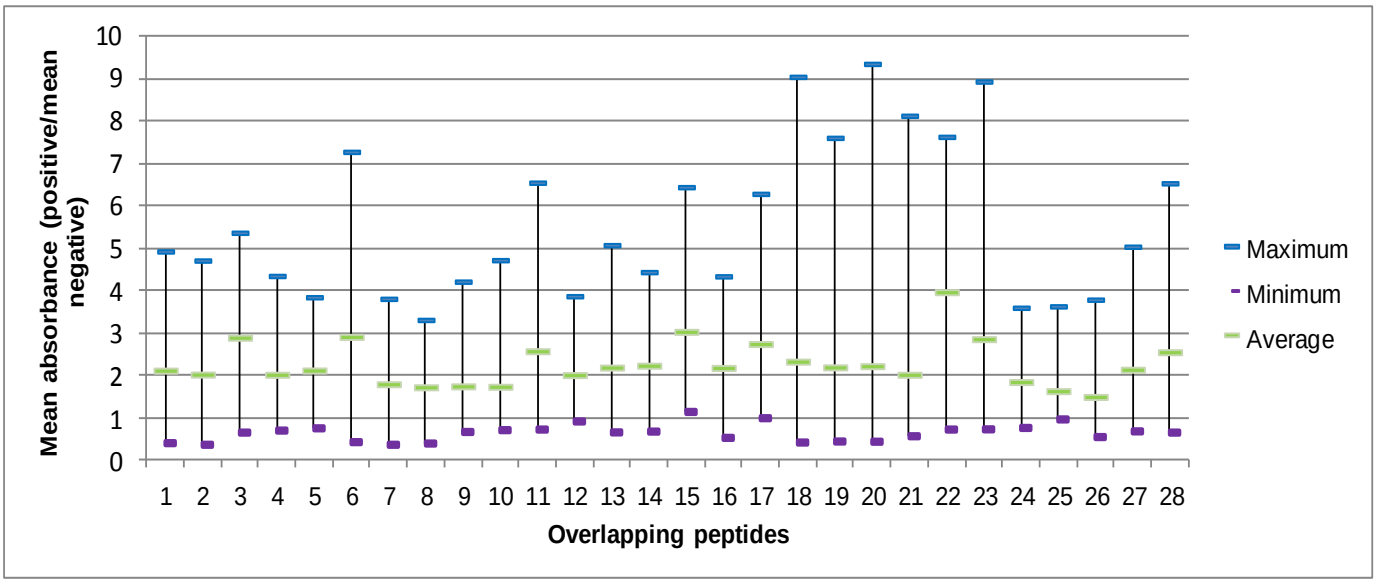


Figure 3.8. Reactivity of sera from YFV vaccinees against synthesized overlapping peptides spanning the NS4a protein (NS4a_1-NS4a5), C protein (C6-C9) and EDIII protein (ED10-ED28) in an indirect ELISA. The y-axis shows the mean reactivity (positive/mean negative) of sera from YFV vaccinees.

3.4. Summary

Surveillance is required to detect spread of YFV from its forest reservoirs to human populations and diagnosis is important for early identification of outbreaks in humans (Vazquez *et al*., 2003). Safe and reliable laboratory tests capable of making a specific serological diagnosis are needed for flaviviruses (Vazquez *et al*., 2003; Tripathi *et al*., 2008). Traditional serological techniques for YFV such as HI, IFA and NT are labour-intensive and time consuming. Isolation of YFV wild type isolates and preparation of diagnostic reagents require BSL three or BSL four facilities, and diagnostic kits are not readily available commercially which restricts many laboratories in their ability to rapidly diagnose YFV and to test the immune status of vaccinated individuals.

In this chapter serum samples from vaccinated individuals were tested using a neutralisation assay to confirm an immune response against YFV in our cohort of samples. Sera from vaccinated persons were used as it can be difficult to obtain samples from natural infections. All sera tested were confirmed to have a detectable immune response against YFV ranging from 4 to 16. The sera were then used to determine the application of ELISA using whole cell lysate antigen and recombinant EDIII antigen for detecting antibody against YFV. A total of 24/24 sera tested positive using whole cell lysate compared with 20/24 that tested positive using recombinant antigen. Similarly, sera from mice immunized with recombinant EDIII were tested by ELISA with cell lysate antigen and neutralisation assays. In total 5/5 mice had a detectable antibody response by ELISA and 3/5 mice had neutralising antibody. It is likely that the absence of neutralising antibody in 2/5 mice is dose related. In a limited number of convalescent samples from horses with suspected WNV infection the YFV EDIII was unable to detect antibody against WNV compared with the whole cell lysate that showed a positive reaction in 2/3 suspected cases (correlating with WNV results from diagnostic laboratory, data not shown). Additional controls such as YFV antibody positive horse sera were not available. A larger cohort of samples would be required to conclusively determine the ability of YFV EDIII to differentiate between flavivirus infection.

Finally the stability of the protein was investigated and precoated plates were stable up to 5 weeks after preparation however by week 7 the OD ratio obtained from the positive and negative control sera was <50% the OD value obtained at week 1. The results of the peptide study identified one region in the EDIII protein that was able to detect antibody in 8/12 sera tested. Other positive reactors were identified but the number of sera reacting were too few to be considered as significant.

Chapter 4

DISCUSSION

In this chapter the main results of the study are summarized. YFV remains one of the major health problems in sub-Saharan Africa and regions of South America and is a risk for travelers to these regions without vaccination (Monath, 2001; Stock *et al*., 2012). Due to high vaccination coverage in South America the rate of transmission of YFV is lower than in Africa (Barnett, 2007). YFV is a mosquito-borne RNA virus maintained in nature by non-human primates and active mosquitoes that breed in tree-holes in the forest canopy (Monath, 2001). Humans are infected by exposure to mosquitoes when they intrude on this cycle during occupational or recreational activities and when there is spillover into urban areas where the virus is maintained in a human to human transmission via domestic *Ae.* spp. mosquitoes (Monath, 2001). Infection in humans ranges from subclinical to systemic disease including jaundice, fatal haemorrhagic fever and renal failure (Barnett, 2007). Neutralising antibodies play an important role in clearance of the virus and protective immunity. Diagnosis of YFV can be achieved by virus isolation, amplification of nucleic acid and detection of viral antigen and antibodies. Viremia is of short duration and serological assays play an important role in diagnosis. However serological cross-reactivity between the members of flaviviruses can complicate interpretation of diagnostic results. In addition diagnostic kits are not readily available commercially which restricts many laboratories in their ability to rapidly diagnose YFV and to test the immune status of vaccinated individuals. Development of novel serological assays that do not require high biocontainment facilities for reagent preparation and could potentially be used to differentiate between members of the flavivirus genus would have important application in diagnosis and surveillance. Identifying immunodominant antigens and epitopes inducing detectable antibodies plays an important role in development of assays. In addition, if the antigens are responsible for inducing antibodies that play a role in protective immunity then they can be used as targets for vaccine development. There is a highly efficacious vaccine available for YFV however as a live attenuated vaccine there are limitations in administering the vaccine to immunocompromised individuals. Hence identification of immunodominant antigens, epitopes and proteins inducing detectable and neutralising antibody responses has application in assay and vaccine development. The development of a recombinant antigen that does not require BSL three facilities for preparation and is safe to use, would have an important role in a laboratory for detecting antibodies in infected individuals and vaccinees. In addition it may be possible to

prepare recombinant antigens or even peptides representing specific epitopes that can differentiate between different *flaviviruses*.

The aim of this study was to identify immunodominant viral proteins that induced detectable antibody responses that could be used for developing diagnostic assays with the potential to differentiate between flavivirus infections and to identify linear B cell epitopes on selected viral proteins. The first objective was to express selected viral recombinant proteins. The EDIII protein of flaviviruses has previously been described as a region that has epitopes which are conserved between different flaviviruses and epitopes which differ suggesting that it could be useful for detection and differentiation assays. In addition the E protein of flaviviruses is important for inducing neutralising antibody responses and protective immunity and is likely a target for vaccine development. Hence the focus of the study was on expression and application of the EDIII protein. The C and NS4a proteins were included as representatives of structural and NS proteins respectively. Viral capsid proteins are likely to induce detectable antibody responses as they are on the outer surface of the virus and well exposed. NS proteins may be useful for detection of early antibody responses as NS proteins are translated early during replication before the structural proteins. EDIII protein was cloned into two expression vectors and various parameters were used to optimize expression of recombinant YFV EDIII protein. However in comparison with previous attempts in our laboratory, the use of XL10 *E.coli* cells instead of JM109 *E.coli* cells allowed expression of protein with adequate yield for downstream application regardless of changes to other expression parameters.

Viral RNA was extracted from an aliquot of 10% mouse brain suspension containing the YFV 17D vaccine strain. The genes encoding the EDIII, NS4a and C proteins were amplified by RT-PCR using specifically designed primers. Using the A/T cloning strategy, the purified PCR products were cloned into pGEM®-T Easy vector. The genes were rescued and subcloned. For bacterial expression the three genes were cloned into pQE-80L vector using specific restriction sites engineered onto the 5' and 3' ends of the PCR products to facilitate cloning. The EDIII gene was also cloned into pCold TF vector. The contructs were used to transform OverExpress cells and protein expression was induced in cell cultures with IPTG. To optimize the expression and yield of the EDIII protein with the pQE-80L vector, culture conditions were varied by lowering the temperature, altering IPTG concentrations and lowering aeration to increase solubility of the product. Expression of the protein was achieved using all variables

with little difference in yield. However, solubility studies confirmed that products expressed at each IPTG concentration were largely insoluble. The recombinant pQE-80L-YFVEDIII protein was purified under denaturing conditions with 8 M urea by affinity chromatography using 6 x Histidine tag. After purification urea was slowly removed by ultrafiltration to allow refolding of the recombinant protein. A ~13 kDa protein was present after purification and confirmed by Western blot analysis.

To obtain a soluble protein the EDIII gene was ligated into pCold TF vector and recombinant protein expressed. A soluble protein was prepared and purified however this protein did not react with positive sera in ELISA and was discontinued. The study was continued with the pQE-80L expressed EDIII protein. The protein was expressed in an insoluble form and purified by denaturing and refolding the protein. Attempts to express the C and NS4a proteins were not successful. Future work will likely focus on using codon optimization of genes for expression studies which has been useful in our laboratory for expressing proteins (Samudzi *et al*., 2012).

The second objective was to determine the biological activity of the recombinant proteins and application in ELISA for detection of IgG antibody against YFV. Serum samples from patients with natural infections of YFV are difficult to obtain in South Africa. Outbreaks occur in central regions of Africa largely in rural areas that make collection of samples difficult. Hence it was decided to use serum samples from people with a history of vaccination. The vaccine for YFV is a live attenuated vaccine that mimics natural infections. Bioinformatics was used to explore the genetic and encoded amino acid differences between vaccine strains and wild type Asibi strain. For the proteins that were investigated in this study, amino acid differences were conserved throughout the vaccine strain with the exception of one change in the NS4a protein of strain 17D that was not present in other vaccine strains but was also not included in the predicted epitopic regions. The serum samples were initially tested for the presence of neutralising antibodies, the serological gold standard and all samples had a detectable neutralising antibody response ranging from 4 to 16. “In house” ELISA were subsequently developed initially using both recombinant EDIII antigens expressed in the study and a whole cell lysate antigen. The EDIII protein expressed using the pCold TF system was not reactive compared with the recombinant protein expressed using pQE-80L. Removal of the TF by enzyme cleavage did not improve the reactivity and hence the study was continued using the recombinant YFV pQE-80L-EDIII antigen. The whole cell lysate antigen was able to detect IgG antibody against YFV in 24/24 samples tested. The EDIII protein detected antibody in 20/24 samples that were tested and

4/24 were non reactors. There was no correlation between the negative samples and the OD obtained from the whole cell lysate, neutralisation antibody titer or from time after vaccination and numbers of boosters. In a limited study using convalescent sera from horses with suspected WNV, the YFV cell lysate antigen detected IgG antibody against WNV in 2/3 sera (horse one and two previously confirmed as antibody positive for WNV using HI). These sera did not react in the ELISA using EDIII suggesting that the recombinant antigen has potential to differentiate between WNV and YFV. The limitations of this study are the number of serum samples and controls. A conclusive decision will require a larger cohort of WNV antibody positive samples and possibly other flavivirus antibody samples. The ability to distinguish may make the recombinant antigen a useful assay to supplement a test using whole cell antigens. The whole cell antigens could provide the sensitivity required for diagnosis and detection and the recombinant antigen could provide the specificity. The stability of the protein would however be an issue for routine use of EDIII protein as precoated plates were only stable for approximately one month. There are other methods such as lyophilisation of the protein that could be investigated to stabilize the protein for longer periods of time.

The EDIII protein was further characterized to determine if the protein induced an immune response in an animal model that could be detected using a native antigen, whole cell lysate antigen, and a neutralising antibody response. Sera from mice immunized with recombinant EDIII were tested by ELISA and neutralisation assays. In total 5/5 mice had a detectable antibody response by ELISA using whole cell lysate antigen and 3/5 mice had neutralising antibody. It is likely that the absence of neutralising antibody in 2/5 mice is dose related. However the detection of a neutralising antibody in 3/5 mice suggests that this region of the E protein is worth investigating further for vaccine purposes. The EDIII region of WNV has been investigated previously and shown to induce a specific and protective immune response against WNV (Chu et al., 2007) but reports investigating the immune response against YFV EDIII have not been published. The results reported here, although limited, indicate that the YFV EDIII region has potential as a subunit vaccine candidate.

Finally the third objective was to use epitope prediction software to predict epitopic regions on selected viral proteins and use peptide libraries covering epitopic regions to identify possible linear B cell epitopes. Bioinformatics was used to identify epitopic regions to reduce the cost of synthesizing peptide libraries that would span the entire length of the three proteins that were selected for the study. The use of whole virus antigen requires biocontainment for preparation

and the antigen cannot differentiate between closely related flaviviruses. Identifying immunogenic epitopes encoded by pathogens is a pre-requisite for developing diagnostic peptides (AnandaRao *et al*., 2005). Therefore synthetic peptide libraries covering potentially epitopic regions of the C, NS4a and EDIII were used to identify linear B cell epitopes by reacting peptides against immune sera, from people with a history of yellow fever vaccination. The results of the peptide study identified a region in the EDIII protein that was able to detect antibody in 8/12 sera tested. Interpretation of this result (8/12) could be lack of sensitivity or that the peptide is not immunodominant and therefore would not be useful in further applications. The lack of sensitivity may be related to the purity of the peptides (70%) used to screen the sera. Lower purity peptides are frequently used for screening purposes to reduce costs. Further investigation, using peptides with >95% purity and that are truncated to identify exact epitopic site and employing a competition assay to confirm the specificity of the reaction against the peptide, will need to be completed to determine the significance of the result and the application of the peptide.

In summary the EDIII protein prepared in this study has some potential for differentiation of flavivirus antibodies although it lacks sensitivity for routine diagnosis. A potential epitope, TGHGTVVMQ, from amino acid 21 to 29 on the EDIII protein was identified using bioinformatics and was shown to have some reactivity against immune sera. The significance of this epitope needs further investigation. Finally the EDIII region of the YFV protein shows potential as a target region for vaccine development which has not previously been published.

REFERENCES

Acheson NH. (2007). Fundamentals of Molecular Virology. 183-4;188.

Agathos SN. Production scale insect cell culture. Biotechnol Adv. 1991;9:51-68.

Ahn M-H, Song M, Oh E-Y, Jamal A, Kim H, Ko K, Choo Y-K, Kim B-J, Ko K. Production of therapeutic proteins with baculovirus expression system in insect cell. Entomol Res. 2008;38:S71-S8.

Amexis G and Yong NS. Multiple antigenic peptides as vaccine platform for the induction of humoral responses against dengue-2 virus. Viral Immunol. 2007;20:657-63.

AnandaRao R, Swaminathan S, Khanna N. The identification of immunodominant linear epitopes of dengue type 2 virus capsid and NS4a proteins using pin-bound peptides. Virus Res. 2005;112:60-8.

Baneyx F. Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol. 1999;10:411-21.

Barnett ED. Yellow Fever: Epidemiology and Prevention. Clin Infect Dis. 2007;44:850-6.

Barrett ADT. Molecular epidemiology of yellow fever virus. Rev Biomed. 2010;21:213-20.

Barrett ADT and Teuwen DE. Yellow fever vaccine – how does it work and why do rare cases of serious adverse events take place? Curr Opin Immunol. 2009;21:308-13.

Baronti C, Goitia NJV, Cook S, Roca Y, Revollo J, Flores JV, de Lamballerie X. Molecular epidemiology of yellow fever in bolivia from 1999 to 2008. Vector Borne Zoonot Dis. 2011;11:277-84.

Beasley DWC and Barrett AD. Identification of neutralising epitopes within structural domain III of the West Nile Virus envelope protein. J Virol. 2002;76:13097-100.

Beasley D, Holbrook M, Travassos Da Rosa A, Coffey L, Carrara A, Phillippi-Falkenstein K, Bohm RP, Ratterree MS, Lillibridge KM, Ludwig GV, Estrada-Franco J, Weaver SC, Tesh RB, Shope RE, Barrett ADT. Use of a recombinant envelope protein subunit antigen for specific serological diagnosis of west nile virus infection. J Clin Microbiol. 2004;42:2759-65.

Bredenbeek PJ, Kooi EA, Lindenbach B, Huijkman N, Rice CM, Spaan WJM. A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. J Gen Virol. 2003;84:1261-8.

Carrington CVF and Auguste AJ. Evolutionary and ecological factors underlying the tempo and distribution of yellow fever virus activity. Infect Genet Evol. 2012;13:1-13.

ChambersTJ, McCart DW, Rice CM. Production of yellow fever virus proteins in infected cells: identification of diverse polyprotein species and analysis of cleavage kinetics using region-specific polycolonal antisera. Virology. 1990;177:159-74.

Chávez JH, Silva JR, Amarilla AA, Figueiredo LTM. Domain III peptides from flavivirus envelope protein are useful antigens for serologic diagnosis and targets for immunization. Biologicals. 2010;38:613-8.

Chin JFL, Chu JJH, Ng ML. The envelope glycoprotein domain III of dengue virus serotypes 1 and 2 inhibit virus entry. Microbes and Infect. 2007;9:1-9.

Chu J-HJ, Chiang C-CS, Ng M. Immunization of flavivirus west nile recombinant envelope domain 3 protein induced specific immune response and protection against west nile virus infection. J Immunol. 2007;178:2699-705.

Chu J-HJ, Rajamanonmani R, Li J, Bhuvanakantham R, Lescar J, Ng M. Inhibition of west nile virus entry by using a recombinant domain III from the envelope glycoprotein. J Gen Virol. 2005;86:405-12.

Clarke E. Refolding of recombinant proteins. Curr Opin Biotechnol. 1998;9:157-63.

Crill WD and Roehrig JT. Monoclonal antibodies that bind to domain III of the dengue virus E glycoprotein are the most efficient blockers of virus absorption to Vero cells. J Virol. 2001;75:7769-73.

Dash PK, Boutonnier A, Prina E, Sharma S, Reiter P. Development of a SYBR green I based RT-PCR assay for yellow fever virus: application in assessment of YFV infection in Aedes aegypti. Virol J. 2012;9:1-8.

De Groot AS and Martin W. From immunome to vaccine : epitope mapping and vaccine design tools. Novartis Found Symp. 2003;254:57-252.

Demain AL and Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv. 2009;27:297-306.

Deubel V, Digoutte J-P, Monath TP, Girard M. Genetic Heterogeneity of yellow fever virus strains from Africa and the Americas. J Gen Virol. 1986;67:209-13.

Dos Santos FB, Miagostovich MP, Nogueira RMR, Schatzmayr HG, Riley RW, Harris E. Analysis of recombinant dengue virus polypeptides for dengue diagnosis and evaluation of the humoral immune response. Am J Trop Med Hyg. 2004;71:144-52.

Fischer B, Summer I, Goodenough P. Isolation, renaturation and formation of disulfide bonds of eurkaryotic proteins expressed in Escherichia coli as inclusion. Biotechnol Bioeng. 1993; 41:3-13.

Frierson GJ. The yellow fever vaccine: A history. Yale J Biol Med. 2010;83:77-85.

Gardner CL and Ryman KD. Yellow fever: A reemerging threat. Clin Lab Med. 2010;30:237-60.

Gómara MJ, Riedemann S, Vega I, Ibarra H, Ercilla G, Haro I. Use of linear and multiple antigenic peptides in the immunodiagnosis of acute hepatitis A virus infection. J Immunol Methods. 2000;234:23-34.

González L, Boyle RW, Zhang M, Castillo J, Whittier S, Della-Latta P, Clarke LM, George JR, Fang X, Wang JG, Hosein B, Wang CY. Synthetic-peptide-based enzyme-linked immunosorbent assay for screening human serum or plasma for antibodies to human immunodeficiency virus type 1 and type 2. Clin Diagn Lab Immunol. 1997;4:598.

Holbrook MR, Shope RE, Barrett ADT. Use of recombinant E protein domain III based enzyme-linked immunosorbent assays for differentiation of tick-borne encephalitis serocomplex flaviviruses form mosquito-borne flaviviruses. J Clin Microbiol. 2004;42:4101-10.

Holzmann H, Kundi M, Stiasny K, Clement J, Mckenna P, Kunz C, Heinz FX. Correlation between ELISA, hemagglutination inhibition and neutralisation tests after vaccination against tick-borne encephalitis. J Med Virol. 1996;48:102-7.

Kuno G, Chang GJ, Tsuchiya R, Karabastos N, Cropp CB. Phylogeny of the genus flaviviruses. J Virol. 1998;72:73-83.

Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:1-7.

Lee E and Lobigs M. E protein domain III determinants of yellow fever virus 17D vaccine strain enhance binding to glycosaminoglycans, impede virus spread, and attenuate virulence. J Virol. 2008;82:6024-33.

Lepineic L, Dalgarno L, Huong VTQ, Monath TP, Digoutte JP, Deubel V. Geographic distribution and evolution of yellow fever viruses based on direct sequencing of genomic cDNA fragments. J Gen Virol. 1994;75:417-23.

Li L, Lok SM, Yu IM. The flavivirus precursor membrane envelope protein complex: structure and maturation. Science. 2008;319:1830-4.

Li L, Barrett ADT, Beasley DWC. Differential expression of domain III neutralising epitopes on the envelope proteins of west nile virus strains. 2005;335:99-105.

Lindenbach BD and Rice CM. Molecular biology of the flaviviruses. In: Chambers, pp. 23-61. Edited by T. A. & T. P. Monath. The Flaviviruses: structure, replication, and evolution. Adv Virus Res. 2003;59. London: Elsevier

Lorenz IC, Kartenbeck J, Mezzacassa A. Intracellular assembly and secretion of recombination subviral particles from tick-borne encephalitis virus. J Virol. 2003;77:4370-82.

Mackenzie JM, Westaway EG. Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. J Virol. 2001;75:10787-99.

Mandl CW, Guirakhoo F, Holzmann H, Heinz FX, Kunz C. Antigenic structure of the flavivirus envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. J Virol. 1989;63:564-71.

Marfin AA, Barwick Eidex R, Monath TP. Yellow fever. Pathogens. 2005;71:797-812.

Mason PW, Dalrymple JM, Gentry MK, McCown JM, Hoke CH, Burke DS, Fournier MJ, Mason TL. Molecular characterization of a neutralising domain of the japanese encephalitis virus structural glycoprotein. J Gen Virol. 1989;70:2037-49.

Misawa S and Kumugai I. Refolding of therapeutic proteins produced in Escherichia coli as inclusion bodies. Biopoly. 1999;51:297-307.

Monath TP. (1989). Yellow fever. In: Monath, T.P. (eds) The Arboviruses: epidemiology and ecology. 139-231.

Monath TP. Yellow fever: an update. Lancet. 2001;1:11-20.

Monath TP. Yellow fever. Medicine. 2005;33:21-3.

Monath TP and Cetron MS. Prevention of yellow fever in persons traveling to the tropics. Vaccines. 2002;34:1369-78.

Monath TP, Fowler E, Johnson CT, Balser J, Morin MJ, Sisti M, Trent DW. An inactivated cell-culture vaccine against yellow fever. N Engl J Med. 2011;364:1326-33.

Monath TP, Hill LJ, Brown NV, Cropp CB, Schlesinger JJ, Saluzzo JF, Wands JR. Sensitive and specific monoclonal immunoassay for detecting yellow fever virus in laboratory and clinical specimens. J Clin Microbiol. 1986;23:129-34.

Mutebi JP and Barrett ADT. The epidemiology of yellow fever in Africa. Microbes and Infect. 2002;4:1459-68.

Mutebi JP, Rijnbrand RCA, Wang H, Ryman KD, Wang E, Fulop LD, Titball R, Barrett ADT. Genetic relationships and evolution of genotypes of yellow fever virus and other members of the yellow fever virus group within the flavivirus genus based on the 3' noncoding region. J Virol. 2004:78:9652-65.

Mutebi J-P, Wang H, Li L, Bryant JE, Barrett ADT. Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa. J Virol. 2001;75:6999-7008.

Patkar CG, Jones CT, Chang Y, Warrier R, Kuhn RJ. Functional requirements of the yellow fever virus capsid protein. J Virol. 2007;81:6471-81.

Ponomarenko JV and Regenmortel MHV. B-cell epitope prediction. Structural Bioinformatics, Second Edition. 2009;35:849-79.

Pulmanausahakul R, Khakpoor A, Smith DR. The development of flavivirus vaccines. Afr J Biotechnol. 2010;9:409-15.

Rossiter D., Pharm D and Pharm M. (2010). South African medicines formulary: Yellow fever vaccine. (ninth edition). The Division of Clinical Pharmacology, Faculty of Health Sciences, University of Cape Town: Health and Medical Publishing Group.

Ryman KD, Ledger TN, Weir RC, Schlesinger JJ, Barrett ADT. Yellow fever virus envelope protein has two discrete type-specific neutralising epitopes. J Gen Virol. 1997;78:1353-6.

Ryman KD, Ledger TN, Campbell GA, Watowich SJ, Barrett ADT. Mutation in a 17D-204 vaccine substrain-specific envelope protein epitope alters the pathogenesis of yellow fever virus in mice. Virology. 1998;244:59-65.

Sambrook and Russel (2001). Molecular Cloning: A Laboratory Manual (3rd ed.). Cold Spring Harbor Laboratory Press.

Samudzi RR, Leman PA, Paweska JT, Swanepoel R, Burt FJ. Bacterial expression of Crimean-Congo hemorrhagic fever virus nucleoprotein and its evaluation as a diagnositic reagent in an indirect ELISA. J Virol Methods. 2012;179:70-6.

Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B. Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. Immunome Res. 2008;4:1-17.

Stock NK, Boschetti N, Herzog C, Appelhans MS, Niedrig M. The phylogeny of yellow fever virus 17D vaccines. Vaccine. 2012;30:989-94.

Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2006;72:211-23.

Theiler M and Smith HH. The use of yellow fever virus modified by in vitro cultivation for human immunization. J Exper Med. 1937;65:787-800.

Tomori O. Yellow Fever: The recurring plague. Crit Rev Clin Lab Sci. 2004;41:391-427.

Tripathi NG, Babu JP, Shrivastva A, Parida M, Jana AM, Rao PVL. Production and characterization of recombinant dengue virus type 4 envelope domain III protein. J Biotechnol. 2008;134:278-86.

Van Regenmortel MHV. Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines. J Mol Recognit. 2006. 19;183-7.

Vaughan K, Greenbaum J, Blythe M, Peters B, Sette A. Meta-analysis of all immune epitope data in the flavivirus genus: Inventory of current immune epitope data status in the context of virus immunity and immunopathology. Viral Immunol. 2010;23:259-84.

Vázquez S, Valdes O, Pupo M, Delgado I, Alvarez M, Pelegrino JL, Guzmán MG. MAC-ELISA and ELISA inhibition methods for detection of antibodies after yellow fever vaccination. J Virol Methds. 2003;110:179-84.

Verma R, Boleti E, George AJT. Antibody engineering: Comparison of bacterial, yeast, insect and mammalian expression systems. J Immunol Methods. 1998;216:165-81.

Von Lindern JJ, Aroner S, Barrett ND, Wicker JA, Davis CT, Barrett ADT. Genome analysis and phylogenetic relationships between east, central and west African isolates of yellow fever virus. J Gen Virol. 2006;87:895-907.

Wamala JF, Malimbo M, Okot CL, Atai-Omoruto AD, Tenywa E, Miller JR, Balinandi S, Shoemaker T, Oyoo C, Omony EO, Kagirita A, Musenero MM, Makumbi I, Nanyunja M, Lutwana JJ, Downing R, Mbonye AK. Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010-January 2011. Internat J Infect Dis. 2012;16:e536-e42.

Wu H-C, Jung M-Y, Chiu C-Y, Chao T-T, Lai S-C, Jan J-T, Shaio M-F. Identification of a dengue virus type 2 (DEN-2) serotype-specific B-cell epitope and detection of DEN-2-immunized animal serum samples using an epitope-based peptide antigen. J Gen Virol. 2003;84:2771-9.

Yu IM, Zhang W, Holdaway HA. Structure of the immature dengue virus low pH primer proteolytic maturation. Science. 2008;319:1834-7.

http://www.cdc.gov/yellowfever/transmission/index.html

ailab.cs.iastate.edu/bcpreds/

www.imtech.res.in/raghava/abcpred/

www.ImmuneEpitope.org

http://www.technelysium.com.au/ChromasPro.html

http://www.clustal.org/clustal2/

http://www.mbio.ncsu.edu/bioedit/bioedit.html

http://www.infectionlandscapes.org/2011/07/yellow-fever.html

APPENDIX A

Nucleotide and amino acid sequence data

A. 1. Nucleotide and amino acid sequence data of the pQE-80L and the YFV C gene. The sequence data is shown from the start codon of the vector illustrating the YFV C gene in frame with the start codon, 6 x Histidine tag and stop codon of the vector.

A	T	G	A	G	A	G	G	A	T	C	G	C	A	T	C	A	C	C	A	T	C	A	C
M			R			G			S			H			H			H			H		
START												**6 X Histidine tag**											

C	A	T	C	A	C	G	G	A	T	C	C	T	C	T	G	G	T	C	G	T	A	A	A
H			H			G			S			S			G			R			K		
6 X Histidine tag						***Bam*HI**						**C encoding region**											

G	C	T	C	A	G	G	G	A	A	A	A	A	C	C	C	T	G	G	G	C	G	T	C
A			Q			G			K			T			L			G			V		
C encoding region																							

A	A	T	A	T	G	G	T	A	C	G	A	C	G	A	G	G	A	G	T	T	C	G	C
N			M			V			R			R			G			V			R		
C encoding region																							

T	C	C	T	T	G	T	C	A	A	A	C	A	A	A	A	T	A	A	A	A	C	A	A
S			L			S			N			K			I			K			Q		
C encoding region																							

A	A	A	A	C	A	A	A	A	C	A	A	A	T	T	G	G	A	A	A	C	A	G	A
K			T			K			Q			I			G			N			R		
C encoding region																							

C	C	T	G	G	A	C	C	T	T	C	A	A	G	A	G	G	T	G	T	T	C	A	A
P			G			P			S			R			G			V			Q		
C encoding region																							

G	G	A	T	T	T	A	T	C	T	T	T	T	T	C	T	T	T	T	T	G	T	T	C
G			F			I			F			F			F			L			F		
C encoding region																							

A	A	C	A	T	T	T	T	G	A	C	T	G	G	A	A	A	A	A	G	A	T	C
N			I			L			T			G			K			K			I		
C encoding region																							

A	C	A	G	C	C	C	A	C	C	T	A	A	A	G	A	G	G	T	T	G	T	G	G
T			A			H			L			K			R			L			W		
C encoding region																							

A	A	A	A	T	G	C	T	G	G	A	C	C	C	A	A	G	A	C	A	A	G	G	C
K			M			L			D			P			R			Q			G		
C encoding region																							

T	T	G	G	C	T	G	T	T	C	T	A	A	G	G	A	A	A	G	T	C	A	A	G
L			A			V			L			R			K			V			K		
C encoding region																							

A	G	A	G	T	G	G	T	G	G	C	C	A	G	T	T	T	G	A	T	G	A	G	A
R			V			V			A			S			L			M			R		
C encoding region																							

G	G	A	T	T	G	T	C	C	T	C	A	A	G	G	A	A	A	C	G	C	C	G	T
G			L			S			S			R			K			R			R		
C encoding region																							

T	C	C	C	A	T	G	A	T	G	T	T	C	T	G	A	C	T	G	T	G	C	A	A
S			H			D			V			L			T			V			Q		
C encoding region																							

T	T	C	C	T	A	A	T	T	T	T	G	G	G	A	A	T	G	C	T	G	T	T	G
F			L			I			L			G			M			L			L		
C encoding region																							

A	T	G	A	C	G	G	G	T	G	G	A	A	A	G	C	T	T	T	A	A			
M			T			G			G			K			L			-					
C encoding region												**_Hind_III**						**STOP**					

A. 2. Nucleotide and amino acid sequence data of the pQE-80L and the YFV NS4a gene. The sequence data is shown from the start codon of the vector illustrating the YFV NS4a gene in frame with the start codon, 6 x Histidine tag and stop codon of the vector.

A	T	G	A	G	A	G	G	A	T	C	G	C	A	T	C	A	C	C	A	T	C	A	C
M			R			G			S			H			H			H			H		
START												**6 X Histidine tag**											

C	A	T	C	A	C	G	G	A	T	C	C	G	G	A	G	C	T	G	C	T	G	A	A
H			H			G			S			G			A			A			E		
6 X Histidine tag						**_Bam_HI**						**NS4a encoding region**											

G	T	G	C	T	A	G	T	T	G	T	G	C	T	G	A	T	G	G	A	A	C	T	C
V			L			V			V			L			S			E			L		
NS4a encoding region																							

C	C	T	G	A	T	T	T	C	C	T	G	G	C	T	A	A	A	A	A	A	G	G	T
P			D			F			L			A			K			K			G		
NS4a encoding region																							

G	G	A	G	A	G	G	C	A	A	T	G	G	A	T	A	C	C	A	T	C	A	G	T
G			E			A			M			D			T			I			S		
NS4a encoding region																							

G	T	G	T	T	C	C	T	C	C	A	C	T	C	T	G	A	G	G	A	A	G	G	C
V			F			L			H			S			E			E			G		
NS4a encoding region																							

T	C	T	A	G	G	G	C	T	T	A	C	C	G	C	A	A	T	G	C	A	C	T	A
S			R			A			Y			R			N			A			L		
NS4a encoding region																							

T	C	A	A	T	G	A	T	G	C	C	T	G	A	G	G	C	A	A	T	G	A	C	A
S			M			M			P			E			A			M			T		
NS4a encoding region																							

A	T	A	G	T	C	A	T	G	C	T	G	T	T	T	A	T	A	C	T	G	G	C	T
I			V			M			L			F			I			L			A		
NS4a encoding region																							

G	G	A	C	T	A	C	T	G	A	C	A	T	C	G	G	G	A	A	T	G	G	T	C
G			L			L			T			S			G			M			V		
NS4a encoding region																							

A	T	C	T	T	T	T	T	C	A	T	G	T	C	T	C	C	C	A	A	A	G	G	C
I			F			F			M			S			P			K			G		
NS4a encoding region																							

A	T	C	A	G	T	A	G	A	A	T	G	T	C	T	A	T	G	G	C	G	A	T	G
I			S			R			M			S			M			A			M		
NS4a encoding region																							

G	G	C	A	C	A	A	T	G	G	C	C	G	G	C	T	G	T	G	G	A	T	A	T
G			T			M			A			G			C			G			Y		
NS4a encoding region																							

C	T	C	A	T	G	T	T	C	C	T	T	G	G	A	G	G	C	G	T	C	A	A	A
L			M			F			L			G			G			V			K		
NS4a encoding region																							

C	C	C	A	C	T	C	A	C	A	T	C	T	C	C	T	A	T	G	T	C	A	T	G
P			T			H			I			S			Y			V			M		
NS4a encoding region																							

C	T	C	A	T	A	T	T	C	T	T	T	G	T	C	C	T	G	A	T	G	G	T	G
L			I			F			F			V			L			M			V		
NS4a encoding region																							

G	T	T	G	T	G	A	T	C	C	C	C	G	A	G	C	C	A	G	G	G	C	A	A
V			V			I			P			E			P			G			Q		
NS4a encoding region																							

C	A	A	C	A	A	A	G	G	T	C	C	A	T	C	C	A	A	G	A	C	A	A	C
Q			R			S			I			Q			D			N			Q		
NS4a encoding region																							

C	A	A	G	T	G	G	C	A	T	A	C	C	T	C	A	T	T	A	T	T	G	G	C
V			A			Y			L			I			I			G			I		
NS4a encoding region																							

A	T	C	C	T	G	A	C	G	C	T	G	G	T	T	T	C	A	G	C	G	G	T	G
L			T			L			V			S			A			V			A		
NS4a encoding region																							

G	C	A	G	C	C	A	A	C	G	A	G	C	T	A	G	G	C	A	T	G	C	T	G
A			N			E			L			G			M			L			E		
NS4a encoding region																							

G	A	G	A	A	A	A	C	C	A	A	A	G	A	G	G	A	C	C	T	C	T	T	T
K			T			K			E			D			L			F			G		
NS4a encoding region																							

G	G	G	A	A	G	A	A	G	A	A	C	T	T	A	A	T	T	C	C	A	T	C	T
K			K			N			L			I			P			S			S		
NS4a encoding region																							

A	G	T	G	C	T	T	C	A	C	C	C	T	G	G	A	G	T	T	G	G	C	C	G
A			S			P			W			S			W			P			D		
NS4a encoding region																							

G	A	T	C	T	T	G	A	C	C	T	G	A	A	G	C	C	A	G	G	A	G	C	T
L			D			L			K			P			G			A			A		
NS4a encoding region																							

G	C	C	T	G	G	A	C	A	G	T	G	T	A	C	G	T	T	G	G	C	A	T	T
W			T			V			Y			V			G			I			V		
NS4a encoding region																							

G	T	T	A	C	A	A	T	G	C	T	C	T	C	T	C	C	A	A	T	G	T	T	G
T			M			L			S			P			M			L			H		
NS4a encoding region																							

C	A	C	C	A	C	T	G	G	A	T	C	A	A	A	G	T	C	G	A	A	T	A	T
H			W			I			K			V			E			Y			G		
NS4a encoding region																							

G	G	C	A	A	C	C	T	G	T	C	T	C	T	G	T	C	T	G	G	A	A	T	A
N			L			S			L			S			G			I			A		
NS4a encoding region																							

G	C	C	C	A	G	T	C	A	G	C	C	T	C	A	G	T	C	C	T	T	T	C	T
Q			S			A			S			V			L			S			F		
NS4a encoding region																							

T	T	C	A	T	G	G	A	C	A	A	G	G	G	G	A	T	A	C	C	A	T	T	C
M			D			K			G			I			P			F			M		
NS4a encoding region																							

A	T	G	A	A	G	A	T	G	A	A	T	A	T	C	T	C	G	G	T	C	A	T	A
K			M			N			I			S			V			I			M		
NS4a encoding region																							

A	T	G	C	T	G	C	T	G	G	T	C	A	G	T	G	G	C	T	G	G	A	A	T
L			L			V			S			G			W			N			S		
NS4a encoding region																							

T	C	A	A	T	A	A	C	A	G	T	G	A	T	G	C	C	T	C	T	G	C	T	C
I			T			V			M			P			L			L			C		
NS4a encoding region																							

T	G	T	G	G	C	A	T	A	G	G	G	T	G	C	G	C	C	A	T	G	C	T	C
G			I			G			C			A			M			L			H		
NS4a encoding region																							

C	A	C	T	G	G	T	C	T	C	T	C	A	T	T	T	T	A	C	C	T	G	G	A
W			S			L			I			L			P			G			I		
NS4a encoding region																							

A	T	C	A	A	A	G	C	G	C	A	G	C	A	G	T	C	A	A	A	G	C	T	T
K			A			Q			Q			S			K			L			A		
NS4a encoding region																							

G	C	A	C	A	G	A	G	A	A	G	G	C	T	T	T	A	A						
Q			R			R			K			L			-								
									***Hind*III**						**STOP**							

A. 3. Nucleotide and amino acid sequence data of the pCold TF and the YFV EDIII gene. The sequence data is shown from the Thrombin site of the vector illustrating the YFV EDIII gene in frame with the Thrombin and Factor Xa sites including the NdeI, SacI, KpnI, XhoI restriction enzyme sites, 6 x Histidine tag, start and end codons of the YFV EDIII gene.

C	T	G	G	T	G	C	C	A	C	G	C	G	G	T	A	G	T	G	G	T	G	G	T
L			V			P			R			G			S			G			G		
Thrombin																							

A	T	C	G	A	A	G	G	T	A	G	G	C	A	T	A	T	G	G	A	G	C	T	C
I			E			G			R			H			M			G			L		
Factor Xa												***Nde*I**						***Sac*I**					

G	G	T	A	C	C	C	T	C	G	A	G	G	G	A	T	C	C	A	A	G	G	G	G
G			T			L			E			G			S			K			G		
***Kpn*I**						***Xho*I**						***Bam*HI**						**EDIII**					

A	C	A	T	C	C	T	A	C	A	A	A	A	T	A	T	G	C	A	C	T	G	A	C
T			S			Y			K			I			C			T			D		
EDIII encoding region																							

A	A	A	A	T	G	T	T	T	T	T	T	G	T	C	A	A	G	A	A	C	C	C	A
K			M			F			F			V			K			N			P		
EDIII encoding region																							

A	C	T	G	A	C	A	C	T	G	G	C	C	A	T	G	G	C	A	C	T	G	T	T
T			D			T			G			H			G			T			V		
EDIII encoding region																							

G	T	G	A	T	G	C	A	G	G	T	G	A	A	A	G	T	G	T	C	A	A	A	A
V			M			Q			V			K			V			S			K		
EDIII encoding region																							

G	G	A	G	C	C	C	C	C	T	G	C	A	G	G	A	T	T	C	C	A	G	T	G
G			A			P			C			R			I			P			V		
EDIII encoding region																							

A	T	A	G	T	A	G	C	T	G	A	T	G	A	T	C	T	T	A	C	A	G	C	G
I			V			A			D			D			L			T			A		
EDIII encoding region																							

G	C	A	A	T	C	A	A	T	A	A	A	G	G	C	A	T	T	T	T	G	G	T	T
A			I			N			K			G			I			L			V		
EDIII encoding region																							

A	C	A	G	T	T	A	A	C	C	C	C	A	T	C	G	C	C	T	C	A	A	C	C
T			V			N			P			I			A			S			T		
EDIII encoding region																							

A	A	T	G	A	T	G	A	T	G	A	A	G	T	G	C	T	G	A	T	T	G	A	G
N			D			D			E			V			L			I			E		
EDIII encoding region																							

G	T	G	A	A	C	C	C	A	C	C	T	T	T	T	G	G	A	G	A	C	A	G	C
V			N			P			P			F			G			D			S		
EDIII encoding region																							

T	A	C	A	T	T	A	T	C	G	T	T	G	G	G	A	G	A	G	G	A	G	A	T
Y			I			I			V			G			R			G			D		
EDIII encoding region																							

T	C	A	C	G	T	C	C	C	A	C	T	T	A	C	C	A	G	T	G	G	C	A	C
S			R			P			T			Y			Q			W			H		
EDIII encoding region																							

A	A	A	G	A	G	A	A	G	C	T	T	G	T	C	G	A	C						
K			E			K			L			V			D								
						***Hind*III**																	

A. 4. Nucleotide sequence alignment of gene encoding EDIII protein regions for YFV 17D, WNV Lineage 1 (NY99), JEV and DENV-2 strains compared to the YFV 17D strain. Conserved nucleotide bases are indicated by an asterix (" * ").

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50         60         70         80
17DEDIII        AAGGGGACAT CCTACAAAAT ATGCACTGAC AAAATGTTTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
AsibiEDIII      AAGGGGACAT CCTACAAAAT GTGCACTGAC AAAATGTCTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
JEVEDIII        AAAGGCACAA CCTATGGCAT GTGTACAGAA AAATTCTCGT TCGCGAAAAA TCCGGCGGAC ACTGGCCACG GAACAGTTGT
WNVEDIII        AAGGGAACAA CCTATGGCGT CTGTTCAAAG GCTTTCAAGT TTCTTGGGAC TCCCGCAGAC ACAGGTCACG GCACTGTGGT
DENVEDIII       AAAGGAATGT CATACTCTAT GTGCACAGGA AAGTTTAAAG TTGTGAAGGA AATAGCAGAA ACACAACATG GAACAATAGT
Clustal Co      ** ** *    * **     *  **  *         *      *              * **  **    ** * * **  * **

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         90        100        110        120        130        140        150        160
17DEDIII        GATGCAGGTG AA--AGTGTC AAAAGGA-GC CCCCTGCAGG ATTCCAGTGA TAGTAGCTGA TGATCTTACA GCGGCAATCA
AsibiEDIII      GATGCAGGTG AA--AGTGCC AAAAGGA-GC CCCCTGCAAG ATTCCAGTGA TAGTAGCTGA TGATCTTACA GCGGCAATCA
JEVEDIII        CATTGAACTA TCCTACTCTG GGAGTGATGG CCCCTGCAAA ATTCCGATTG TCTCCGTTGC GAGCCTCAAT GACATGACCC
WNVEDIII        GTTGGAATTG CAGTACACTG GCACGGATGG ACCTTGCAAA GTTCCTATCT CGTCAGTGGC TTCATTGAAC GACCTAACGC
DENVEDIII       TATCAGAGTG CAATATGAAG GGGACGGCTC TCCATGCAAG ATCCCTTTTG AGATAATGGA T---TTGGAA AAAAGACATG
Clustal Co        *     *      *           *      ** ****    * **  *                   *

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         170        180        190        200        210        220        230        240
17DEDIII        ATAAAGGCAT TTTGGTTACA GTTAACCCCA TCGCCTCAA C---CAATGAT GAT---GAAG TGCTGATTGA GGTGAACCCA
AsibiEDIII      ATAAAGGCAT TTTGGTTACA GTTAACCCCA TCGCCTCAA C---CAATGAT GAT---GAAG TGCTGATTGA GGTGAACCCA
JEVEDIII        CCGTTGGGCG GCTGGTGACA GTGAACCCTT TCGTCGCGA CTTCCAGTGCC AACTCAAAGG TGCTGGTCGA GATGGAACCC
WNVEDIII        CAGTGGGCAG ATTGGTCACT GTCAACCCTT TTGTTTCAG TGGCCACGGCC AACGCTAAGG TCCTGATTGA ATTGGAACCA
DENVEDIII       TCTTAGGTCG CCTGATTACA GTCAACCCAA TTGTGACAG A---AAAAGAT AGC---CCAG TCAACATAGA AGCAGAACCT
Clustal Co           **      ** * **  ** *****   *     *   *    *  *           * *     * **      * **

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         250        260        270        280        290        300        310
17DEDIII        CCTTTTGGAG ACAGCTACAT TATCGTTGGG AGAGGAGATT CACGTCTCAC TTACCAGTGG CACAAAGAG
AsibiEDIII      CCTTTTGGAG ACAGCTACAT TATCGTTGGG ACAGGAGATT CACGTCTCAC TTACCAGTGG CACAAAGAG
JEVEDIII        CCCTTCGGAG ACTCCTACAT CGTGGTTGGG AGGGGAGACA AGCAGATCAA CCACCATTGG CACAAAGCT
WNVEDIII        CCCTTTGGAG ACTCATACAT AGTGGTGGGC AGAGGAGAAC AACAGATCAA TCACCATTGG CACAAGTCT
DENVEDIII       CCATTCGGAG ACAGCTACAT CATCATAGGA GTAGAGCCGG GACAACTGAA GCTCAACTGG TTTAAGAAA
Clustal Co      ** ** **** **   *****   *  * **     *         *   * *     * * ***    **
```

A. 5. Predicted amino acid sequence alignment of EDIII protein for YFV 17D, WNV Lineage 1 (NY99), JEV and DENV-2 strains compared to the YFV 17D strain. Conserved amino acids are indicated by an asterix (" * ").

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50         60         70         80
WNVEDIII        KGTTYGVCSK AFKFLGTPAD TGHGTVVLEL QYTGTDGPCK VPISSVASLN DLTPVGRLVT VNPFVSVATA NAKVLIELEP
JEVEDIII        KGTTYGMCTE KFSFAKNPAD TGHGTVVIEL SYSGSDGPCK IPIVSVASLN DMTPVGRLVT VNPFVATSSA NSKVLVEMEP
17DEDIII        KGTSYKICTD KMFFVKNPTD TGHGTVVMQV KVS-KGAPCR IPVIVADDLT AAINKGILVT VNPIASTN-- DDEVLIEVNP
AsibiEDIII      KGTSYKMCTD KMSFVKNPTD TGHGTVVMQV KVP-KGAPCK IPVIVADDLT AAINKGILVT VNPIASTN-- DDEVLIEVNP
DENV2EDIII      KGMSYSMCTG KFKVVKEIAE TQHGTIVIRV QYEGDGSPCK IPFEIMD-LE KRHVLGRLIT VNPIVTEK-- DSPVNIEAEP
Clustal Co      ** :* :*:  : .     :: * ***:*:.: .    ..**: :*.     *       * *:* ***:.   :    :* :*  *

                ....|....| ....|....| ....|....|
                        90        100        110
WNVEDIII        PFGDSYIVVG RGEQQINHHW HKS
JEVEDIII        PFGDSYIVVG RGDKQINHHW HKA
17DEDIII        PFGDSYIIVG RGDSRLTYQW HKE
AsibiEDIII      PFGDSYIIVG TGDSRLTYQW HKE
DENEDIII        PFGDSYIIIG VEPGQLKLNW FKK
Clustal Co      *******::*    : :. :* .*
```

A. 6. Nucleotide sequence alignment of (A. 6a) C, (A.6b) EDIII and (A. 6c) NS4a for YFV vaccine strains 17D-213, 17DD, 17D-204, 17D and YFV Asibi (wild type) strain. Conserved nucleotide bases are indicated by an asterix (" * "). Nucleotide changes are shaded.

A. 6a.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50         60         70         80
CAP17D-204      TCTGGTCGTA AAGCTCAGGG AAAAACCCTG GGCGTCAATA TGGTACGACG AGGAGTTCGC TCCTTGTCAA ACAAAATAAA
CAP17D-213      TCTGGTCGTA AAGCTCAGGG AAAAACCCTG GGCGTCAATA TGGTACGACG AGGAGTTCGC TCCTTGTCAA ACAAAATAAA
CAP17D          TCTGGTCGTA AAGCTCAGGG AAAAACCCTG GGCGTCAATA TGGTACGACG AGGAGTTCGC TCCTTGTCAA ACAAAATAAA
CAP17DD         TCTGGTCGTA AAGCTCAGGG AAAAACCCTG GGCGTCAATA TGGTACGACG AGGAGTTCGC TCCTTGTCAA ACAAAATAAA
CAPASIBI        TCTGGTCGTA AAGCTCAGGG AAAAACCCTG GGCGTCAATA TGGTACGACG AGGAGTTCGC TCCTTGTCAA ACAAAATAAA
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          90        100        110        120        130        140        150        160
CAP17D-204       ACAAAAAACA AAACAAATTG GAAACAGACC TGGACCTTCA AGAGGTGTTC AAGGATTTAT CTTTTTCTTT TTGTTCAACA
CAP17D-213       ACAAAAAACA AAACAAATTG GAAACAGACC TGGACCTTCA AGAGGTGTTC AAGGATTTAT CTTTTTCTTT TTGTTCAACA
CAP17D           ACAAAAAACA AAACAAATTG GAAACAGACC TGGACCTTCA AGAGGTGTTC AAGGATTTAT CTTTTTCTTT TTGTTCAACA
CAP17DD          ACAAAAAACA AAACAAATTG GAAACAGACC TGGACCTTCA AGAGGTGTTC AAGGATTTAT CTTTTTCTTT TTGTTCAACA
CAPASIBI         ACAAAAAACA AAACAAATTG GAAACAGACC TGGACCTTCA AGAGGTGTTC AAGGATTTAT CTTTTTCTTT TTGTTCAACA
Clustal Co       ********** ********** ********** ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         170        180        190        200        210        220        230        240
CAP17D-204       TTTTGACTGG AAAAAAGATC ACAGCCCACC TAAAGAGGTT GTGGAAAATG CTGGACCCAA GACAAGGCTT GGCTGTTCTA
CAP17D-213       TTTTGACTGG AAAAAAGATC ACAGCCCACC TAAAGAGGTT GTGGAAAATG CTGGACCCAA GACAAGGCTT GGCTGTTCTA
CAP17D           TTTTGACTGG AAAAAAGATC ACAGCCCACC TAAAGAGGTT GTGGAAAATG CTGGACCCAA GACAAGGCTT GGCTGTTCTA
CAP17DD          TTTTGACTGG AAAAAAGATC ACAGCCCACC TAAAGAGGTT GTGGAAAATG CTGGACCCAA GACAAGGCTT GGCTGTTCTA
CAPASIBI         TTTTGACTGG AAAAAAGATC ACGGCCCACC TAAAGAGGTT GTGGAAAATG CTGGACCCAA GACAAGGCTT GGCTGTTCTA
Clustal Co       ********** ********** ** ******* ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         250        260        270        280        290        300        310        320
CAP17D-204       AGGAAAGTCA AGAGAGTGGT GGCCAGTTTG ATGAGAGGAT TGTCCTCAAG GAAACGCCGT TCCCATGATG TTCTGACTGT
CAP17D-213       AGGAAAGTCA AGAGAGTGGT GGCCAGTTTG ATGAGAGGAT TGTCCTCAAG GAAACGCCGT TCCCATGATG TTCTGACTGT
CAP17D           AGGAAAGTCA AGAGAGTGGT GGCCAGTTTG ATGAGAGGAT TGTCCTCAAG GAAACGCCGT TCCCATGATG TTCTGACTGT
CAP17DD          AGGAAAGTTA AGAGAGTGGT GGCCAGTTTG ATGAGAGGAT TGTCCTCAAG GAAACGCCGT TCCCATGATG TTCTGACTGT
CAPASIBI         AGGAAAGTTA AGAGAGTGGT GGCCAGTTTG ATGAGAGGAT TGTCCTCAAG GAAACGCCGT TCCCATGATG TTCTGACTGT
Clustal Co       ******** * ********** ********** ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....|
                         330        340        350        360
CAP17D-204       GCAATTCCTA ATTTTGGGAA TGCTGTTGAT GACGGGTGGA
CAP17D-213       GCAATTCCTA ATTTTGGGAA TGCTGTTGAT GACGGGTGGA
CAP17D           GCAATTCCTA ATTTTGGGAA TGCTGTTGAT GACGGGTGGA
CAP17DD          GCAATTCCTA ATTTTGGGAA TGCTGTTGAT GACGGGTGGA
CAPASIBI         GCAATTCCTA ATTTTGGGAA TGCTGTTGAT GACGGGTGGA
Clustal Co       ********** ********** ********** **********
```

A. 6b.

```
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50      60         70         80
17D213EDII       AAGGGGACAT CCTACAAAAT ATGCACTGAC AAAATGTTTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
17DEDIII         AAGGGGACAT CCTACAAAAT ATGCACTGAC AAAATGTTTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
17D204EDII       AAGGGGACAT CCTACAAAAT ATGCACTGAC AAAATGTTTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
17DDEDIII        AAGGGGACAT CCTACAAAAT ATGCACTGAC AAAATGTTTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
ASIBI            AAGGGGACAT CCTACAAAAT GTGCACTGAC AAAATGTCTT TTGTCAAGAA CCCAACTGAC ACTGGCCATG GCACTGTTGT
Clustal Co       ********** **********  ********* ******* ** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           90        100        110        120        130        140        150        160
17D213EDII       GATGCAGGTG AAAGTGTCAA AAGGAGCCCC CTGCAGGATT CCAGTGATAG TAGCTGATGA TCTTACAGCG GCAATCAATA
17DEDIII         GATGCAGGTG AAAGTGTCAA AAGGAGCCCC CTGCAGGATT CCAGTGATAG TAGCTGATGA TCTTACAGCG GCAATCAATA
17D204EDII       GATGCAGGTG AAAGTGTCAA AAGGAGCCCC CTGCAGGATT CCAGTGATAG TAGCTGATGA TCTTACAGCG GCAATCAATA
17DDEDIII        GATGCAGGTG AAAGTGCCAA AAGGAGCCCC CTGCAGGATT CCAGTGATAG TAGCTGATGA TCTTACAGCG GCAATCAATA
ASIBI            GATGCAGGTG AAAGTGCCAA AAGGAGCCCC CTGCAAGATT CCAGTGATAG TAGCTGATGA TCTTACAGCG GCAATCAATA
Clustal Co       ********** ****** *** ********** ***** **** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          170        180        190        200        210        220        230        240
17D213EDII       AAGGCATTTT GGTTACAGTT AACCCCATCG CCTCAACCAA TGATGATGAA GTGCTGATTG AGGTGAACCC ACCTTTTGGA
17DEDIII         AAGGCATTTT GGTTACAGTT AACCCCATCG CCTCAACCAA TGATGATGAA GTGCTGATTG AGGTGAACCC ACCTTTTGGA
17D204EDII       AAGGCATTTT GGTTACAGTT AACCCCATCG CCTCAACCAA TGATGATGAA GTGCTGATTG AGGTGAACCC ACCTTTTGGA
17DDEDIII        AAGGCATTTT GGTTACAGTT AACCCCATCG CCTCAACCAA TGATGATGAA GTGCTGATTG AGGTGAACCC ACCTTTTGGA
ASIBI            AAGGCATTTT GGTTACAGTT AACCCCATCG CCTCAACCAA TGATGATGAA GTGCTGATTG AGGTGAACCC ACCTTTTGGA
Clustal Co       ********** ********** ********** ********** ********** ********** ********** **********

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          250        260        270        280        290        300
17D213EDII       GACAGCTACA TTATCGTTGG GAGAGGAGAT TCACGTCTCA CTTACCAGTG GCACAAAGAG
17DEDIII         GACAGCTACA TTATCGTTGG GAGAGGAGAT TCACGTCTCA CTTACCAGTG GCACAAAGAG
17D204EDII       GACAGCTACA TTATCGTTGG GAGAGGAGAT TCACGTCTCA CTTACCAGTG GCACAAAGAG
17DDEDIII        GACAGCTACA TTATCGTTGG AAGAGGAGAT TCACGTCTCA CTTACCAGTG GCACAAAGAG
ASIBI            GACAGCTACA TTATCGTTGG GACAGGAGAT TCACGTCTCA CTTACCAGTG GCACAAAGAG
Clustal Co       ********** **********  * ******* ********** ********** **********
```

A. 6c.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50         60         70         80
17DD            GGAGCTGCTG AAGTGCTAGT TGTGCTGAGT GAACTCCCTG ATTTCCTGGC TAAAAAAGGT GGAGAGGCAA TGGACACCAT
17D204          GGAGCTGCTG AAGTGCTAGT TGTGCTGAGT GAACTCCCTG ATTTCCTGGC TAAAAAAGGT GGAGAGGCAA TGGATACCAT
17D213          GGAGCTGCTG AAGTGCTAGT TGTGCTGAGT GAACTCCCTG ATTTCCTGGC TAAAAAAGGT GGAGAGGCAA TGGATACCAT
17D             GGAGCTGCTG AAGTGCTAGT TGTGCTGAGT GAACTCCCTG ATTTCCTGGC TAAAAAAGGT GGAGAGGCAA TGGATACCAT
NS4AASIBI       GGAGCTGCGG AAGTGCTAGT TGTGCTGAGT GAACTCCCTG ATTTCCTGGC TAAAAAAGGT GGAGAGGCAA TGGATACCAT
Clustal Co      ******** * ********** ********** ********** ********** ********** ********** **** *****

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        90        100        110        120        130        140        150        160
17DD            CAGTGTGTTT CTCCACTCTG AGGAAGGCTC TAGGGCTTAC CGCAATGCAC TATCAATGAT GCCTGAGGCA ATGACAATAG
17D204          CAGTGTGTTT CTCCACTCTG AGGAAGGCTC TAGGGCTTAC CGCAATGCAC TATCAATGAT GCCTGAGGCA ATGACAATAG
17D213          CAGTGTGTTT CTCCACTCTG AGGAAGGCTC TAGGGCTTAC CGCAATGCAC TATCAATGAT GCCTGAGGCA ATGACAATAG
17D             CAGTGTGTTC CTCCACTCTG AGGAAGGCTC TAGGGCTTAC CGCAATGCAC TATCAATGAT GCCTGAGGCA ATGACAATAG
NS4AASIBI       CAGTGTGTTT CTCCACTCTG AGGAAGGCTC TAGGGCTTAC CGCAATGCAC TATCAATGAT GCCTGAGGCA ATGACAATAG
Clustal Co      *********  ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       170        180        190        200        210        220        230        240
17DD            TCATGCTGTT TATACTGGCT GGACTCCTGA CATCGGGAAT GGTCATCTTT TTCATGTCTC CCAAAGGCAT CAGTAGAATG
17D204          TCATGCTGTT TATACTGGCT GGACTCCTGA CATCGGGAAT GGTCATCTTT TTCATGTCTC CCAAAGGCAT CAGTAGAATG
17D213          TCATGCTGTT TATACTGGCT GGACTCCTGA CATCGGGAAT GGTCATCTTT TTCATGTCTC CCAAAGGCAT CAGTAGAATG
17D             TCATGCTGTT TATACTGGCT GGACTACTGA CATCGGGAAT GGTCATCTTT TTCATGTCTC CCAAAGGCAT CAGTAGAATG
NS4AASIBI       TCATGCTGTT TATACTGGCT GGACTACTGA CATCGGGAAT GGTCATCTTT TTCATGTCTC CCAAAGGCAT CAGTAGAATG
Clustal Co      ********** ********** ***** **** ********** ********** ********** ********** **********
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       250        260        270        280        290        300        310        320
17DD            TCTATGGCGA TGGGCACAAT GGCCGGCTGT GGATATCTCA TGTTCCTTGG AGGCGTCAAA CCCACTCACA TCTCCTATAT
17D204          TCTATGGCGA TGGGCACAAT GGCCGGCTGT GGATATCTCA TGTTCCTTGG AGGCGTCAAA CCCACTCACA TCTCCTATAT
17D213          TCTATGGCGA TGGGCACAAT GGCCGGCTGT GGATATCTCA TGTTCCTTGG AGGCGTCAAA CCCACTCACA TCTCCTATAT
17D             TCTATGGCGA TGGGCACAAT GGCCGGCTGT GGATATCTCA TGTTCCTTGG AGGCGTCAAA CCCACTCACA TCTCCTATGT
NS4AASIBI       TCTATGGCGA TGGGCACAAT GGCCGGCTGT GGATATCTCA TGTTCCTTGG AGGCGTCAAA CCCACTCACA TCTCCTATAT
Clustal Co      ********** ********** ********** ********** ********** ********** ********** ******** *
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       330        340        350        360        370        380        390        400
17DD            CATGCTCATA TTCTTTGTCC TGATGGTGGT TGTGATCCCC GAGCCAGGGC AACAAAGGTC CATCCAAGAC AACCAAGTGG
17D204          CATGCTCATA TTCTTTGTCC TGATGGTGGT TGTGATCCCC GAGCCAGGGC AACAAAGGTC CATCCAAGAC AACCAAGTGG
17D213          CATGCTCATA TTCTTTGTCC TGATGGTGGT TGTGATCCCC GAGCCAGGGC AACAAAGGTC CATCCAAGAC AACCAAGTGG
17D             CATGCTCATA TTCTTTGTCC TGATGGTGGT TGTGATCCCC GAGCCAGGGC AACAAAGGTC CATCCAAGAC AACCAAGTGG
NS4AASIBI       CATGCTCATA TTCTTTGTCC TGATGGTGGT TGTGATCCCC GAGCCAGGGC AACAAAGGTC CATCCAAGAC AACCAAGTGG
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450        460        470        480
17DD            CATACCTCAT TATTGGCATC CTGACGCTGG TTTCAGCGGT GGCAGCCAAC GAGCTAGGCA TGCTGGAGAA AACCAAAGAG
17D204          CATACCTCAT TATTGGCATC CTGACGCTGG TTTCAGCGGT GGCAGCCAAC GAGCTAGGCA TGCTGGAGAA AACCAAAGAG
17D213          CATACCTCAT TATTGGCATC CTGACGCTGG TTTCAGCGGT GGCAGCCAAC GAGCTAGGCA TGCTGGAGAA AACCAAAGAG
17D             CATACCTCAT TATTGGCATC CTGACGCTGG TTTCAGCGGT GGCAGCCAAC GAGCTAGGCA TGCTGGAGAA AACCAAAGAG
NS4AASIBI       CATACCTCAT TATTGGCATC CTGACGCTGG TTTCAGTGGT GGCAGCCAAC GAGCTAGGCA TGCTGGAGAA AACCAAAGAG
Clustal Co      ********** ********** ********** ****** *** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       490        500        510        520        530        540        550        560
17DD            GACCTCTTTG GGAAGAAGAA CTTAATTCCA TCTAGTGCTT CACCCTGGAG TTGGCCGGAT CTTGACCTGA AGCCAGGAGC
17D204          GACCTCTTTG GGAAGAAGAA CTTAATTCCA TCTAGTGCTT CACCCTGGAG TTGGCCGGAT CTTGACCTGA AGCCAGGAGC
17D213          GACCTCTTTG GGAAGAAGAA CTTAATTCCA TCTAGTGCTT CACCCTGGAG TTGGCCGGAT CTTGACCTGA AGCCAGGAGC
17D             GACCTCTTTG GGAAGAAGAA CTTAATTCCA TCTAGTGCTT CACCCTGGAG TTGGCCGGAT CTTGACCTGA AGCCAGGAGC
NS4AASIBI       GACCTCTTTG GGAAGAAGAA CTTAATTCCA TCTAGTGCTT CACCCTGGAG TTGGCCGGAT CTTGACCTGA AGCCAGGAGC
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       570        580        590        600        610        620        630        640
17DD            TGCCTGGACA GTGTACGTTG GCATTGTTAC AATGCTCTCT CCAATGTTGC ACCACTGGAT CAAAGTCGAA TATGGCAACC
17D204          TGCCTGGACA GTGTACGTTG GCATTGTTAC AATGCTCTCT CCAATGTTGC ACCACTGGAT CAAAGTCGAA TATGGCAACC
17D213          TGCCTGGACA GTGTACGTTG GCATTGTTAC AATGCTCTCT CCAATGTTGC ACCACTGGAT CAAAGTCGAA TATGGCAACC
17D             TGCCTGGACA GTGTACGTTG GCATTGTTAC AATGCTCTCT CCAATGTTGC ACCACTGGAT CAAAGTCGAA TATGGCAACC
NS4AASIBI       TGCCTGGACA GTGTACGTTG GCATTGTTAC AATGCTCTCT CCAATGTTGC ACCACTGGAT CAAAGTCGAA TATGGCAACC
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       650        660        670        680        690        700        710        720
17DD            TGTCTCTGTC TGGAATAGCC CAGTCAGCCT CAGTCCTTTC TTTCATGGAC AAGGGGATAC CATTCATGAA GATGAATATC
17D204          TGTCTCTGTC TGGAATAGCC CAGTCAGCCT CAGTCCTTTC TTTCATGGAC AAGGGGATAC CATTCATGAA GATGAATATC
17D213          TGTCTCTGTC TGGAATAGCC CAGTCAGCCT CAGTCCTTTC TTTCATGGAC AAGGGGATAC CATTCATGAA GATGAATATC
17D             TGTCTCTGTC TGGAATAGCC CAGTCAGCCT CAGTCCTTTC TTTCATGGAC AAGGGGATAC CATTCATGAA GATGAATATC
NS4AASIBI       TGTCTCTGTC TGGAATAGCC CAGTCAGCCT CAGTCCTTTC TTTCATGGAC AAGGGGATAC CATTCATGAA GATGAATATC
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       730        740        750        760        770        780        790        800
17DD            TCGGTCATAA TGCTGCTGGT CAGTGGCTGG AATTCAATAA CAGTGATGCC TCTGCTCTGT GGCATAGGGT GCGCCATGCT
17D204          TCGGTCATAA TGCTGCTGGT CAGTGGCTGG AATTCAATAA CAGTGATGCC TCTGCTCTGT GGCATAGGGT GCGCCATGCT
17D213          TCGGTCATAA TGCTGCTGGT CAGTGGCTGG AATTCAATAA CAGTGATGCC TCTGCTCTGT GGCATAGGGT GCGCCATGCT
17D             TCGGTCATAA TGCTGCTGGT CAGTGGCTGG AATTCAATAA CAGTGATGCC TCTGCTCTGT GGCATAGGGT GCGCCATGCT
NS4AASIBI       TCGGTCATAA TACTGCTGGT CAGTGGCTGG AATTCAATAA CAGTGATGCC TCTGCTCTGT GGCATAGGGT GCGCCATGCT
Clustal Co      ********** * ******** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| .
                       810        820        830        840        850        860
17DD            CCACTGGTCT CTCATTTTAC CTGGAATCAA AGCGCAGCAG TCAAAGCTTG CACAGAGAAG G
17D204          CCACTGGTCT CTCATTTTAC CTGGAATCAA AGCGCAGCAG TCAAAGCTTG CACAGAGAAG G
17D213          CCACTGGTCT CTCATTTTAC CTGGAATCAA AGCGCAGCAG TCAAAGCTTG CACAGAGAAG G
17D             CCACTGGTCT CTCATTTTAC CTGGAATCAA AGCGCAGCAG TCAAAGCTTG CACAGAGAAG G
NS4AASIBI       CCACTGGTCT CTCATTTTAC CTGGAATCAA AGCGCAGCAG TCAAAGCTTG CACAGAGAAG G
Clustal Co      ********** ********** ********** ********** ********** ********** *
```

A. 7. Predicted amino acid sequence alignment of (A. 7a) C, (A. 7b) EDIII and (A. 7c) NS4a for YFV vaccine strains 17D-213, 17DD, 17D-204, 17D and YFV Asibi (wild type) strain. Conserved nucleotide bases are indicated by an asterix (" * "). Amino acid changes are shaded.

A. 7a.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50         60         70         80
CAPASIBI        SGRKAQGKTL GVNMVRRGVR SLSNKIKQKT KQIGNRPGPS RGVQGFIFFF LFNILTGKKI TAHLKRLWKM LDPRQGLAVL
CAP17D          SGRKAQGKTL GVNMVRRGVR SLSNKIKQKT KQIGNRPGPS RGVQGFIFFF LFNILTGKKI TAHLKRLWKM LDPRQGLAVL
CAP17D204       SGRKAQGKTL GVNMVRRGVR SLSNKIKQKT KQIGNRPGPS RGVQGFIFFF LFNILTGKKI TAHLKRLWKM LDPRQGLAVL
CAP17D213       SGRKAQGKTL GVNMVRRGVR SLSNKIKQKT KQIGNRPGPS RGVQGFIFFF LFNILTGKKI TAHLKRLWKM LDPRQGLAVL
CAP17DD         SGRKAQGKTL GVNMVRRGVR SLSNKIKQKT KQIGNRPGPS RGVQGFIFFF LFNILTGKKI TAHLKRLWKM LDPRQGLAVL
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....|
                         90        100        110        120
CAPASIBI        RKVKRVVASL MRGLSSRKRR SHDVLTVQFL ILGMLLMTGG
CAP17D          RKVKRVVASL MRGLSSRKRR SHDVLTVQFL ILGMLLMTGG
CAP17D204       RKVKRVVASL MRGLSSRKRR SHDVLTVQFL ILGMLLMTGG
CAP17D213       RKVKRVVASL MRGLSSRKRR SHDVLTVQFL ILGMLLMTGG
CAP17DD         RKVKRVVASL MRGLSSRKRR SHDVLTVQFL ILGMLLMTGG
Clustal Co      ********** ********** ********** **********
```

A. 7b.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50         60         70         80
17D213          KGTSYKICTD KMFFVKNPTD TGHGTVVMQV KVSKGAPCRI PVIVADDLTA AINKGILVTV NPIASTNDDE VLIEVNPPFG
17D             KGTSYKICTD KMFFVKNPTD TGHGTVVMQV KVSKGAPCRI PVIVADDLTA AINKGILVTV NPIASTNDDE VLIEVNPPFG
17D204          KGTSYKICTD KMFFVKNPTD TGHGTVVMQV KVSKGAPCRI PVIVADDLTA AINKGILVTV NPIASTNDDE VLIEVNPPFG
17DD            KGTSYKICTD KMFFVKNPTD TGHGTVVMQV KVPKGAPCRI PVIVADDLTA AINKGILVTV NPIASTNDDE VLIEVNPPFG
AsibiEDIII      KGTSYKMCTD KMSFVKNPTD TGHGTVVMQV KVPKGAPCKI PVIVADDLTA AINKGILVTV NPIASTNDDE VLIEVNPPFG
Clustal Co      ******:*** ** ******* ********** **.*****:* ********** ********** ********** **********
```

```
                ....|....| ....|....|
                        90        100
17D213          DSYIIVGRGD SRLTYQWHKE
17D             DSYIIVGRGD SRLTYQWHKE
17D204          DSYIIVGRGD SRLTYQWHKE
17DD            DSYIIVGRGD SRLTYQWHKE
AsibiEDIII      DSYIIVGTGD SRLTYQWHKE
Clustal Co      ******* ** **********
```

A. 7c.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       10         20         30         40         50         60         70         80
17DD            GAAEVLVVLS ELPDFLAKKG GEAMDTISVF LHSEEGSRAY RNALSMMPEA MTIVMLFILA GLLTSGMVIF FMSPKGISRM
17D204          GAAEVLVVLS ELPDFLAKKG GEAMDTISVF LHSEEGSRAY RNALSMMPEA MTIVMLFILA GLLTSGMVIF FMSPKGISRM
17D213          GAAEVLVVLS ELPDFLAKKG GEAMDTISVF LHSEEGSRAY RNALSMMPEA MTIVMLFILA GLLTSGMVIF FMSPKGISRM
17D             GAAEVLVVLS ELPDFLAKKG GEAMDTISVF LHSEEGSRAY RNALSMMPEA MTIVMLFILA GLLTSGMVIF FMSPKGISRM
NS4a            GAAEVLVVLS ELPDFLAKKG GEAMDTISVF LHSEEGSRAY RNALSMMPEA MTIVMLFILA GLLTSGMVIF FMSPKGISRM
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       90        100        110        120        130        140        150        160
17DD            SMAMGTMAGC GYLMFLGGVK PTHISYIMLI FFVLMVVVIP EPGQQRSIQD NQVAYLIIGI LTLVSAVAAN ELGMLEKTKE
17D204          SMAMGTMAGC GYLMFLGGVK PTHISYIMLI FFVLMVVVIP EPGQQRSIQD NQVAYLIIGI LTLVSAVAAN ELGMLEKTKE
17D213          SMAMGTMAGC GYLMFLGGVK PTHISYIMLI FFVLMVVVIP EPGQQRSIQD NQVAYLIIGI LTLVSAVAAN ELGMLEKTKE
17D             SMAMGTMAGC GYLMFLGGVK PTHISYVMLI FFVLMVVVIP EPGQQRSIQD NQVAYLIIGI LTLVSAVAAN ELGMLEKTKE
NS4a            SMAMGTMAGC GYLMFLGGVK PTHISYIMLI FFVLMVVVIP EPGQQRSIQD NQVAYLIIGI LTLVSVVAAN ELGMLEKTKE
Clustal Co      ********** ********** ******:*** ********** ********** ********** *****.**** **********

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      170        180        190        200        210        220        230        240
17DD            DLFGKKNLIP SSASPWSWPD LDLKPGAAWT VYVGIVTMLS PMLHHWIKVE YGNLSLSGIA QSASVLSFMD KGIPFMKMNI
17D204          DLFGKKNLIP SSASPWSWPD LDLKPGAAWT VYVGIVTMLS PMLHHWIKVE YGNLSLSGIA QSASVLSFMD KGIPFMKMNI
17D213          DLFGKKNLIP SSASPWSWPD LDLKPGAAWT VYVGIVTMLS PMLHHWIKVE YGNLSLSGIA QSASVLSFMD KGIPFMKMNI
17D             DLFGKKNLIP SSASPWSWPD LDLKPGAAWT VYVGIVTMLS PMLHHWIKVE YGNLSLSGIA QSASVLSFMD KGIPFMKMNI
NS4a            DLFGKKNLIP SSASPWSWPD LDLKPGAAWT VYVGIVTMLS PMLHHWIKVE YGNLSLSGIA QSASVLSFMD KGIPFMKMNI
Clustal Co      ********** ********** ********** ********** ********** ********** ********** **********
```

```
                ....|....| ....|....| ....|....| ....|....| ....|..
                        250        260        270        280
17DD            SVIMLLVSGW NSITVMPLLC GIGCAMLHWS LILPGIKAQQ SKLAQRR
17D204          SVIMLLVSGW NSITVMPLLC GIGCAMLHWS LILPGIKAQQ SKLAQRR
17D213          SVIMLLVSGW NSITVMPLLC GIGCAMLHWS LILPGIKAQQ SKLAQRR
17D             SVIMLLVSGW NSITVMPLLC GIGCAMLHWS LILPGIKAQQ SKLAQRR
NS4a            SVIILLVSGW NSITVMPLLC GIGCAMLHWS LILPGIKAQQ SKLAQRR
Clustal Co      ***:****** ********** ********** ********** *******
```

APPENDIX B

ELISA Raw data

B. 1. Raw data of IgG ELISA using YFV whole cell lysate antigen for detection of antibodies in human vaccinees.

	30/08	56/08	19/09	35/09	36/09	41/09	42/09	46/09	47/09	48/09	49/09	50/09	
Cell lysate	0.7480	0.3710	0.7770	0.6720	0.6870	0.5980	0.5350	0.7330	1.0080	0.4290	1.3570	1.3620	**1:500**
Cell lysate	0.7450	0.4080	0.7800	0.6770	0.7070	0.6440	0.4470	0.7910	1.0770	0.5020	1.4570	1.3630	
Mock lysate	0.1990	0.1460	0.2170	0.1530	0.2690	0.2210	0.1210	0.1680	0.2080	0.1680	0.2810	0.2440	**1:100**
Mock lysate	0.1550	0.1980	0.2410	0.1700	0.2740	0.1750	0.1150	0.1840	0.2470	0.1640	0.3820	0.3040	

	51/09	54/09	55/09	65/09	66/09	7/11	5/12	6/12	7/12	12/12	13/12	44/12	
Cell lysate	1.5990	0.3380	0.3690	0.7230	0.6540	0.3120	0.6170	0.6490	0.4760	0.7920	0.7800	1.5850	**1:500**
Cell lysate	1.5760	0.3330	0.3310	0.7160	0.6690	0.2930	0.5840	0.6720	0.5960	0.8930	0.8300	1.5940	
Mock lysate	0.2610	0.1810	0.1440	0.4190	0.2090	0.1430	0.1460	0.0970	0.1450	0.1610	0.1300	0.1500	**1:100**
Mock lysate	0.3040	0.1640	0.1600	0.4380	0.2440	0.1460	0.1550	0.1130	0.2050	0.2620	0.2370	0.2090	

B. 2. Raw data of IgG ELISA using YFV cell lysate and mock cell lysate antigen using negative flaviviral panel to determine cut-off.

	17/07	18/07	21/07	22/07	23/07	24/07	25/07	28/08	
YFV lysate	0.1700	0.1810	0.1360	0.1560	0.1900	0.2990	0.1280	0.1620	**1:500**
YFV lysate	0.1710	0.1990	0.1250	0.1500	0.1590	0.2860	0.1120	0.1860	
Mock lysate	0.1120	0.1210	0.1010	0.0730	0.1150	0.2830	0.1050	0.1550	**1:100**
Mock lysate	0.1290	0.1060	0.0860	0.0840	0.1100	0.3300	0.0970	0.1350	

B. 3. Raw data of IgG ELISA using recombinant YFV pQE-80L-EDIII antigen for detection of antibodies in human vaccinees.

	30/08	56/08	19/09	35/09	36/09	41/09	42/09	46/09	47/09	48/09	49/09	50/09	
pQE-80L-EDIII	0.3260	0.4380	0.6250	0.1720	0.2960	0.4880	0.1650	0.3340	0.3640	0.1930	0.5880	0.4700	**1:500**
pQE-80L Mock	0.2170	0.1850	0.2250	0.0920	0.2240	0.2580	0.1270	0.1710	0.2310	0.1350	0.3030	0.1660	**1:500**
	51/09	**54/09**	**55/09**	**65/09**	**66/09**	**7/11**	**5/12**	**6/12**	**7/12**	**12/12**	**13/12**	**44/12**	
pQE-80L-EDIII	0.5730	0.2920	0.2350	0.3770	0.3820	0.2040	0.3320	0.1300	0.2170	0.2630	0.2560	0.3620	**1:500**
pQE-80L Mock	0.2430	0.1540	0.1540	0.2430	0.1890	0.1460	0.1690	0.1050	0.1460	0.1590	0.1660	0.1730	**1:500**

B. 4. Raw data of IgG ELISA using recombinant YFV pQE-80L-EDIII and pQE-80L mock antigen using negative flaviviral panel to determine cut-off.

	17/07	18/07	21/07	22/07	23/07	25/07	
pQE-80L-EDIII	0.2080	0.1650	0.1980	0.1460	0.2290	0.1910	**1:500**
pQE-80L-EDIII	0.2610	0.2020	0.2000	0.1380	0.2350	0.1620	
pQE-80L Mock	0.3060	0.3130	0.1990	0.1810	0.2820	0.1640	**1:500**
pQE-80L Mock	0.3330	0.3330	0.1960	0.1630	0.2460	0.1610	

B. 5. Raw data of IgG ELISA using recombinant YFV pColdTF-EDIII and pCold TF mock antigen using a high positive and negative sample.

	1:320	1:320	1:640	1:640	1:1280	1:1280
pColdTF-EDIII (C++)	0.1840	0.1470	0.1240	0.1280	0.1450	0.2370

pColdTF Mock (C++)	0.2950	0.3070	0.2400	0.2000	0.2430	0.2420
pColdTF Mock (C-)	0.1130	0.1040	0.1050	0.0930	0.0960	0.1120

B. 6. Raw data of IgG ELISA for stability of the recombinant YFV pQE-80L-EDIII antigen over four months.

		1 week	2 week	5 week	7 week	19 week
pQE-80L-EDIII	**C++**	1.191	0.8750	0.4110	0.4250	0.2730
	C+	0.525	0.3710	0.2760	0.2320	0.1690
	C-	0.249	0.1800	0.1940	0.1170	0.1340

B. 7. Raw data of IgG ELISA using YFV whole cell lysate antigen for detection of antibodies in mouse sera from mice immunized with recombinant YFV pQE-80L-EDIII

	Mouse1	Mouse2	Mouse3	Mouse4	Mouse5	
YFV lysate	0.3590	0.3550	0.6170	0.3540	0.5370	**1:500**
YFV lysate	0.3880	0.3080	0.5020	0.2930	0.5240	
Mock lysate	0.1630	0.1850	0.1070	0.0670	0.3630	**1:100**
Mock lysate	0.2350	0.1820	0.1080	0.0690	0.3760	

B. 8. Raw data of IgG ELISA using YFV whole cell lysate antigen for detection of antibodies in mouse sera from mice immunized with 17D

	Mouse1	Mouse2	Mouse3	Mouse4	Mouse5	Mouse6	
YFV lysate	0.2360	0.7600	0.9320	0.2950	0.3850	0.2550	**1:500**
YFV lysate	0.2190	0.7930	0.9580	0.2810	0.4190	0.2670	
Mock lysate	0.1240	0.0930	0.1180	0.0780	0.0680	0.0680	**1:100**
Mock lysate	0.1030	0.0980	0.1470	0.1650	0.1110	0.0910	

B. 9. Raw data of IgG ELISA using YFV cell lysate and mock cell lysate antigen using negative sera to determine cut-off.

	Sera 1	Sera 2	
YFV lysate	0.3220	0.2990	**1:500**
YFV lysate	0.3650	0.3440	
Mock lysate	0.2940	0.3130	**1:100**
Mock lysate	0.4410	0.3610	

B. 10. Raw data of IgG ELISA using YFV cell lysate and recombinant YFV pQE-80L-EDIII antigens for detection of antibodies from convalescent WNV horse sera.

	Horse 1	Horse 2	Horse 3	Horse -	
YFV lysate	0.4380	0.6960	0.2510	0.4180	**1:500**
YFV lysate	0.4270	0.6860	0.2340	0.4210	
Mock lysate	0.1920	0.0250	0.1390	0.2240	**1:100**
Mock lysate	0.2030	0.3730	0.1440	0.2210	

	Horse 1	Horse 2	Horse 3	Horse -	
pQE-80L-EDIII	0.3200	0.5380	0.2540	0.3670	**1:500**
pQE-80L-EDIII	0.3000	0.0310	0.2340	0.3180	
pQE-80L Mock	0.1570	0.4080	0.1460	0.2450	**1:500**
pQE-80L Mock	0.2250	0.4700	0.1900	0.2820	

B. 11. Raw data of IgG ELISA for detection of antibodies from human vaccinees against NS4a (1-5), C (6-9) and EDIII (10-28) peptides.

1	2	3	4	5	6	7	8	9	10	11	12
NS4A_1	NS4A_2	NS4A_3	NS4A_4	NS4A_5	C6	C7	C8	C9	ED10	ED11	ED12
ED13	ED14	ED15	ED16	ED17	ED18	ED19	ED20	ED21	ED22	ED23	ED24
ED25	ED26	ED27	ED28								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 30/08	0.7560	0.5300	0.5500	0.4910	0.6890	0.6720	0.7960	0.6470	0.8810	0.9870	0.7570	0.8090
	0.9250	0.8540	0.5990	0.8730	0.6050	1.8910	1.5900	1.9550	1.6990	1.2910	1.8680	0.7520
	0.3990	0.3680	0.4270	0.3260								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 35/09	0.7510	0.7340	0.6240	0.8000	0.8040	0.7110	0.5830	0.5880	0.7600	0.6060	0.6160	0.6740
	0.7430	0.7050	0.6800	0.7330	1.3140	0.7610	0.6590	0.6270	0.6400	0.6060	0.6310	0.6460
	0.7590	0.7920	0.6250	0.5780								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 19/09	0.1990	0.2520	1.1550	0.2920	0.2270	1.5210	0.1920	0.1920	0.4370	0.4100	1.3690	0.4600
	0.2470	0.4120	1.3470	0.4990	0.2960	0.2490	0.2340	0.1940	0.3910	1.5950	1.3060	0.3230
	0.4710	0.3200	1.0530	1.3660								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 36/09	0.5390	0.5530	0.4060	0.4960	0.6090	0.4610	0.4550	0.4080	0.3850	0.3700	0.4560	0.5550
	0.4770	0.3730	0.3220	0.3040	0.3200	0.2910	0.3740	0.2840	0.2670	0.3280	0.3080	0.2920
	0.3590	0.3210	0.2730	0.3350								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 41/09	0.4460	0.4060	0.4550	0.4570	0.3850	0.3800	0.3790	0.3630	0.2300	0.2350	0.2340	0.2770
	0.4590	0.4740	0.4440	0.4360	0.6010	0.4150	0.3710	0.3680	0.2300	1.1200	0.2240	0.3040
	0.2510	0.2290	0.2630	0.2420								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 42/09	0.3330	0.3700	0.3740	0.2980	0.2930	0.2960	0.2470	0.2540	0.1490	0.1500	0.2340	0.1890
	0.1660	0.4110	0.3860	0.3330	0.3950	0.3250	0.2790	0.2670	0.1530	0.7040	0.1690	0.1860
	0.1990	0.1930	0.2370	0.2010								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 46/09	1.0300	0.9850	0.9680	0.9090	0.7480	0.6970	0.6580	0.6920	0.2640	0.2840	0.2260	0.2020
	1.0610	0.9280	0.9500	0.9070	1.1460	0.6180	0.6000	0.5820	0.4160	1.2490	0.3770	0.6690
	0.3860	0.3750	0.3620	0.3960								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 47/09	0.5730	0.4720	0.6250	0.4640	0.5030	0.4700	0.5060	0.5470	0.4460	0.4580	0.4680	0.4830
	0.4920	0.4820	0.4450	0.4240	0.9370	0.4780	0.4840	0.5450	0.4560	0.9860	0.4500	0.4840
	0.4160	0.3790	0.4150	0.4010								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 48/09	0.1610	0.1430	0.1340	0.1820	0.1700	0.2280	0.1770	0.2070	0.1370	0.1360	0.1490	0.1680
	0.3150	0.1740	0.2360	0.2470	0.3560	0.2680	0.4070	0.2570	0.1630	0.6890	0.1500	0.1570
	0.1680	0.1650	0.1400	0.1340								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 49/09	0.2520	0.2520	0.2740	0.2650	0.2690	0.2430	0.2610	0.2470	0.2480	0.2720	0.2740	0.3140
	0.2420	0.2670	0.2540	0.2600	0.3000	0.2650	0.2700	0.2650	0.2350	0.9370	0.2480	0.2590
	0.3000	0.2350	0.2590	0.3880								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 50/09	0.0820	0.0740	0.5670	0.1440	0.4590	0.0870	0.0740	0.0800	0.1460	0.1460	0.6060	0.1910
	0.1350	0.1390	0.7110	0.1080	0.4050	0.0850	0.0900	0.0890	0.1160	0.1490	0.4730	0.1860
	0.1330	0.1120	0.4000	0.6920								

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
VBD 51/09	0.1840	0.3020	1.1220	0.2620	0.1550	1.5360	0.1310	0.1490	0.2980	0.2960	1.0630	0.7200
	0.2190	0.3750	1.2250	0.3390	0.2050	0.1900	0.1390	0.1340	0.2930	0.2930	0.9630	0.3430
	0.3070	0.2590	0.9020	1.3360								

B. 11. Raw data of IgG ELISA against NS4a (1-5), C (6-9) and EDIII (10-28) peptides performed on a negative serum sample.

Peptide 1:500	1	2	3	4	5	6	7	8	9	10	11	12
C-	0.1680	0.1690	0.1810	0.1950	0.2200	0.2250	0.2040	0.2120	0.2390	0.2740	0.1920	0.2100
	0.1740	0.2100	0.1980	0.1990	0.2130	0.1990	0.2310	0.2190	0.1950	0.3130	0.1970	0.2210
	0.2180	0.2040	0.1950	0.1990								

APPENDIX C

Abstract of poster presented at the South African Society for Biochemistry and Molecular Biology (SASBMB), 18-20th January 2010, Bloemfontein

TITLE: RECOMBINANT YELLOW FEVER VIRAL ENVELOPE PROTEIN FOR THE DETECTION OF ANTIBODIES

Shannon Smouse[1], Felicity Burt[1,2]

[1]Department of Medical Microbiology and Virology, [2]NHLS Universitas, Faculty of Health Sciences, University of the Free State, Bloemfontein, 9300.

Email: smouses.MD@ufs.ac.za

Introduction and aim: Yellow fever virus is a mosquito-borne virus that causes viral hemorrhagic fever in both Africa and South America. The virus is the prototype of the family *Flaviviridae* and belongs to the genus *Flavivirus* which comprises of approximately 70 viruses. It is transmitted to vertebrates by the bite of an infected female mosquito, usually the *Aedes* species. It is a re-emerging pathogen with case-fatality rates that can exceed 50% in individuals. The diagnosis of infection and testing of the immune status of vaccinees require reagents that are prepared in biosafety level 3 and 4 facilities. The viral envelope (E) protein plays an important role in eliciting antibodies and hence serves as an ideal diagnostic and research tool for the detection of antibodies. The aim of this research was to prepare a safe recombinant yellow fever viral envelope protein for use as a diagnostic tool for the detection of antibodies against yellow fever in vaccinated and infected individuals.

Methodology: The domain III region of the envelope (EDIII) gene was amplified with primers identified using sequence data retrieved from Genbank. The EDIII gene was cloned into a plasmid vector and subsequently cloned into an *E.coli* bacterial expression system. The expressed protein was characterized using a western blot and purified using the His-tag and Ni^{+} columns. An ELISA for detection of antibodies against yellow fever virus was developed using the recombinant antigen.

Results: The EDIII gene was amplified as confirmed on an agarose gel. A 13 kDa protein was expressed from the construct and purified. A direct ELISA was developed using the recombinant antigen for detection of IgG antibodies against yellow fever virus.

Conclusion: Recombinant antigens are useful diagnostic tools because they are safe to use, simple, cost effective and extremely stable. Antibody reactivity against the recombinant EDIII (r-EDIII) protein indicated that the EDIII region is a suitable antigen for use as a diagnostic tool.

Abstract of poster presented at the 13th International Congress on Infectious Diseases (ICID) 9 – 12th March 2010, Miami, USA

TITLE: PREPARATION OF ANTIGENICALLY ACTIVE RECOMBINANT YELLOW FEVER VIRAL ENVELOPE DOMAIN III PROTEIN

S Smouse and FJ Burt

Background: Yellow fever virus belongs to the genus *Flavivirus*, of the family *Flaviviridae*. It is amosquito-borne virus endemic in tropical regions of Africa and South America. Although an effective vaccine is available, the virus remains a major public health threat, particularly in Africa where vaccination is limited by poverty, civil wars and the inaccessibility of rural areas prone to outbreaks. It is a re-emerging pathogen with case-fatality rates that can exceed 50%. The diagnosis of infection and testing of the immune status of vaccinees require reagents that are prepared in biosafety level 3 and 4 facilities. The viral envelope protein plays an important role in eliciting antibodies and hence serves as an ideal diagnostic and research tool for the detection of antibodies. The aim of the study was to compare various bacterial expression systems for the preparation a recombinant yellow fever viral envelope protein for the detection of antibodies against yellow fever in vaccinated individuals.

Methods: The domain III region of the envelope (EDIII) gene was amplified with primers identified using sequence data retrieved from GenBank. The EDIII gene was cloned into three different *E.coli* bacterial expression systems for comparison of protein yield, solubility and suitablity in an ELISA. Proteins were all expressed with an N+ terminal His tag for purification. A direct ELISA was developed in which the plates were coated with antigen and reacted with serum samples from vaccinees.

Results: Each antigen was evaluated using serum samples collected from vaccinees and serological cross reactivity of the antigen against heterologous flaviviral antibody was determined using convalescent serum samples from patients with known flaviviral infections, such as West Nile. Protein yields varied significantly between the different expression systems with higher yields and increased solubility obtained from expression at lower temperatures. Higher reactivity was associated with homologous antibody.

Conclusion: Preliminary results suggest that bacterially expressed recombinant EDIII protein is antigenically and potentially useful for detecting antibodies against yellow fever virus.

Abstract of poster presented at the Faculty of Health Sciences, Research Forum 26-27th August 2010, University of the Free State, Bloemfontein

TITLE: INDUCTION OF HUMORAL IMMUNITY FROM LINEARIZED EPITOPES ON THE DOMAIN III REGION OF YELLOW FEVER VIRUS ENVELOPE PROTEIN

Shannon L Smouse, Felicity Burt

Introduction and aim: Yellow fever virus is a mosquito-borne virus that causes viral hemorrhagic fever in Africa and South America. The viral envelope protein plays an important role in eliciting neutralising antibodies and is a useful diagnostic and research tool for the detection of antibodies. We have previously expressed the domain III region of the envelope (EDIII) protein as an antigen for detection of antibodies against yellow fever virus in an ELISA. The aim of this research was to determine if linearized epitopes on the domain III protein are immunogenic in an animal model which would be important for vaccine development.

Materials and Methods: The EDIII gene was expressed using an E.coli bacterial expression system. To determine the immunogenicity of linearized epitopes on the recombinant EDIII protein, groups of mice were immunized with the recombinant protein in combination with an adjuvant. An ELISA was developed using a cell lysate antigen prepared from Vero cells infected with 17D vaccine strain. Serum samples collected from the mice after booster inoculations were tested for antibody against yellow fever virus using the ELISA and neutralisation assays.

Results: The recombinant EDIII protein was shown to be immunogenic in mice and was able to induce a detectable humoral antibody response in serum samples collected from mice that received two immunizations plus adjuvant.

Conclusion: Currently there is a highly efficacious vaccine against yellow fever virus, however it is a live attenuated vaccine and, although rare, there have been reports of serious adverse events including fatalities. Hence there is scope for development of a subunit vaccine. The identification of epitopes that induce a protective immune response is essential for development of new generation vaccines.

Abstract of poster presented at the Virology Africa 2011 Congress, UCT Graduate School of Business, V&A Waterfront, Cape Town, South Africa

TITLE: PREPARATION OF RECOMBINANT YELLOW FEVER VIRAL ENVELOPE DOMAIN III PROTEIN FOR USE AS A DIAGNOSTIC REAGENT AND INDUCTION OF NEUTRALISING ANTIBODY RESPONSE

SHANNON L. SMOUSE[1], FELICITY J. BURT[1, 2]

[1]Department of Medical Microbiology and Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein, [2]National Health Laboratory Services, Universitas, Bloemfontein

Yellow fever virus belongs to the genus *Flavivirus*, of the family *Flaviviridae*. It is a mosquito-borne virus endemic in tropical regions of Africa and South America. Although an effective vaccine is available, the virus remains a major public health threat, particularly in Africa where vaccination is limited by poverty, civil wars and the inaccessibility of rural areas prone to outbreaks. It is a re-emerging pathogen with case-fatality rates that can exceed 50%. The diagnosis of infection and testing of the immune status of vaccinees require reagents that are prepared in biosafety level 3 and 4 facilities. The viral envelope protein plays an important role in eliciting antibodies and hence serves as an ideal diagnostic and research tool for the detection of antibodies. The aim of the study was to prepare a recombinant yellow fever viral envelope protein for the detection of antibodies against yellow fever and to determine if epitopes on the domain III protein are immunogenic and can induce a neutralising antibody response in an animal model which may be important for vaccine development. The domain III region of the envelope (EDIII) was amplified with primers identified using sequence data retrieved from GenBank. The EDIII protein was expressed using a bacterial expression system. A direct ELISA was developed in which the plates were coated with antigen and reacted with serum samples from vaccinees. An ELISA was developed using a cell lysate antigen prepared from Vero cells infected with 17D vaccine strain. Each antigen was evaluated using serum samples collected from vaccinees and serological cross reactivity of the antigen against heterologous flaviviral antibody was determined using convalescent serum samples from patients with known flaviviral infections, such as West Nile. Higher reactivity was associated with homologous antibody. To determine the immunogenicity of epitopes on the recombinant EDIII protein, groups of mice were immunized with the recombinant protein in combination with an adjuvant. Serum samples were collected from the mice after booster inoculations and were tested for antibody against yellow fever virus using ELISA and neutralisation assays. The recombinant EDIII protein was shown to be immunogenic in mice and was able to induce a detectable humoral antibody response. Results suggest that bacterially expressed recombinant EDIII protein is antigenically active and potentially useful for detecting antibodies against yellow fever virus. Currently there is a highly efficacious vaccine against yellow fever virus, however it is a live attenuated vaccine and, although rare, there have been reports of serious adverse events inlcuding fatalities. Hence there is scope for development of a subunit vaccine. The identification of epitopes that induce a neutralising immune response may have application in development of new generation vaccines.

Abstract of poster presented at the Faculty of Health Sciences, Research Forum 23-24th August 2012, University of the Free State, Bloemfontein

TITLE: IDENTIFICATION OF LINEAR B-CELL EPITOPES IN THE CAPSID, NS4A AND DOMAIN III REGION IN THE E GLYCOPROTEIN OF YELLOW FEVER VIRUS

Shannon Smouse, Felicity Burt

Introduction and aim: Yellow fever virus (YFV) is a mosquito-borne virus that causes viral hemorrhagic fever in tropical parts of both Africa and South America. The virus has re-emerged and has become a public health concern across the world, despite the availability of a highly efficacious vaccine. This vaccine cannot be administered to immune-compromised individuals, thus the identification and mapping of viral epitopes is important for development of subunit vaccines and improved diagnostics. Our aim was to identify immuno-dominant antigenic regions and screen for linear B-cell epitopes on the capsid, NS4a and domain III (EDIII) region of the E glycoprotein of YFV.

Materials and Methods: Putative hydrophobic and hydrophilic regions along the length of each protein were predicted using BepiPred and ABCpred. Surface accessibility and antigenicity were done using the Immune Epitope Database (IEDB) version 2.0 software (www.ImmuneEpitope.org). Bioinformatics was used to identify 28 overlapping peptides (8 mer length offset by 3) representing predicted amino acids of the capsid protein, NS4a protein and EDIII protein. Using an in-house ELISA, the peptides were screened for reactivity against YFV antibody using immune sera from 12 patients with a history of vaccination against YFV. A serum sample from a patient with no history of vaccination was used as a negative control to determine a cut-off value for the ELISA.

Results: Overlapping peptides 1-5 covering the region (IPSSASPWSWPDLDLKPGAA) of NS4a reacted with 6/12 patient sera. Overlapping peptides 6-9 covering the region (KTKQIGNRPGPSRGVQG) of capsid reacted with 4/12 patient sera. Overlapping peptides 10-28 of EDIII reacted with 4/12 patient sera. Peptides 14, 15, 22, 23 and 28 of EDIII showed consistent reactivity against all the patient sera. Overlapping peptides 22 and 23 showed higher reactivity compared to the other peptides.

Conclusion: These peptides may be important tools for novel diagnostic assays. Further studies to determine their role in protective immunity would indicate their usefulness in vaccine development.

OPSOMMING

Die Geelkoors virus is ‘n virus wat deur geleedpotiges gedra word en wat virale haemoragiese koors veroorsaak in mense in die tropiese gebiede van beide Afrika en Suid-Amerika. Die virus is lid van die familie Flaviviridae, van die genus Flavivirus wat ongeveer 70 virusse insluit. Dit word na gewerwelde diere oorgedra deur die byt van ‘n geïnfekteerde vroulike muskiet van hoofsaaklik die *Aedes* spesie. Dit is ‘n patogeen wat weer opnuut sy verskyning gemaak het en wat ‘n mortaliteit van meer as 50% in menslike gevalle het. Geelkoors kan ‘n akute siekte in mense veroorsaak, met gepaardgaande koors wat kan vererger tot ernstige siekte met lewer en nierversaking. Die diagnose van infeksie en toets van die immuunstatus van ingeënte mense vereis reagense wat in bioveiligheidsvlak (BSL) drie en vier fasiliteite voorberei moet word. Die ontwikkeling van ‘n rekombinante antigeen, wat nie BSL drie fasiliteite vir voorbereiding vereis nie, sal dus ‘n belangrike rol speel in ‘n diagnostiese laboratorium vir die opsporing van antiliggaampies in geïnfekteerde individue en geïmmuniseerde mense. Ondanks die beskikbaarheid van ‘n lewende-geattenueerde effektiewe entstof, word dit nie aanbeveel vir immuunonderdrukte mense nie en sal die ontwikkeling van ‘n nuwe generasie entstof belangrike implikasies vir publieke gesondheid inhou. Identifisering en kartering van antigeniese gedeeltes en virale epitope is belangrik vir die ontwikkeling van subeenheid entstowwe en verbeterde diagnose. Subeenheid entstowwe wat fokus op antigene wat ‘n beskermende immuunreaksie induseer bied ‘n veilige benadering tot die ontwikkeling van entstowwe teen siektes wat ernstige en gereeld dodelike haemoragiese koors veroorsaak. Die doel van hierdie studie was die identifisering van immuundominante virale proteïene wat opspoorbare antigeniese reaksies veroorsaak wat gebruik kon word vir die ontwikkeling van diagnostiese toetse en om liniêre B sel epitope op geselekteerde virale proteïene te identifiseer.

Die volledige oop leesraam van die gene wat kodeer vir die domein III gedeelte van die omhulsel proteïen, kapsel en NS4a proteïene van geelkoors, is geamplifiseer vanaf die 17D stam van geelkoors deur middel van RT-PKR en die gebruik van priemstukke wat spesifiek ontwerp is vanaf basispaarfragmentdata wat van GenBank afgetrek is. Oligonukleotied priemstukke is gemodifiseer met *Bam*HI and *Hind*HII beperkingsensiem posisies wat “downstream” kloning fasiliteer. Elke amplikon is in ‘n pGEM®-T Easy cloning vektor gekloneer deur gebruik van T/A klonering. Elke geen is vanuit die rekombinante plasmied herwin deur gebruik van die *Bam*HI en *Hind*III beperkingsensiem posisies en ingebou in ‘n bakteriese

uitdrukkingssisteem vektor, pQE-80L. In ‘n vorige studie is die geelkoors EDIII geen in pQE-80L gekloneer en uitgedruk in JM109 *Escherichia coli* selle, maar ‘n uitermate lae opbrengs is verkry. In hierdie studie is die vlakke van uitdrukking verhoog deur die gebruik van verskillende sellyne en optimisasie van die inkubasie kondisies. ‘n Onoplosbare 13 kDa proteïen is uitgedruk met behulp van die konstruk en bevestig deur Westelike blot analise. Die proteïen is uitgedruk met ‘n 6 x Histidine merker wat gebruik is om suiwering te fasiliteer deur gebruik van ‘n Ni^{2+} kolom onder denaturerings kondisies. Pogings om die geelkoors C en NS4a proteïene uit te druk was nie suksesvol nie en is gestaak. In ‘n poging om die oplosbaarheid te verbeter is die geelkoors EDIII geen uitgesny uit die pGEM®-T Easy vektor en daarna in ‘n pCold TF bakteriese vektor gekloneer. ‘n ~65 kDa oplosbare proteïen is met behulp van die konstruk uitgedruk en gesuiwer onder natuurlike toestande.

Die funksionele aktiwiteit van die rekombinante antigene in ELISA is vergelyk met heelsel lisaat antigeen wat voorberei is vanaf selle wat met geelkoors geïnfekteer is. Die biologiese aktiwiteit van die rekombinante geelkoors pQE-80L-EDIII antigeen is bevestig in immuuntoetse deur gebruik van serum monsters van mense wat met geelkoors ingeënt is. Positiewe sera wou nie in ELISA reageer wanneer pCold TF uitdrukkings antigeen gebruik is nie en hierdie antigeen is uitgelaat uit verdere toetse. ‘n Totaal van 20/24 serum monsters van geënte mense tydens verskillende stadiums na inenting, het gereageer in ‘n ELISA met die rekombinante geelkoors pQE-80L-EDIII proteïen en 24/24 het in ELISA gereageer met heelsel lisaat antigeen. Dit kon aangetoon word dat die EDIII gedeelte van die omhulsel proteïen in staat was om tussen West Nile Virus infeksie en geelkoors infeksie te onderskei in ‘n beperkte aantal sera van herstelde perde. Die rekombinante EDIII proteïen is gebruik om muise in te ent. Serum monsters wat van die muise versamel is, het reageer teen heelsel lisaat antigeen in ELISA en dit is aangetoon dat die muise neutraliserende antiliggame het, in ‘n *in vitro* neutralisasie toets. Die EDIII gedeelte van die E proteïen het dus waarskynlik ‘n belangrike beskermende immuunrespons uitgelok. Laastens is bioinformatika gebruik om moontlike epitoop gedeeltes te voorspel en deur peptiedbiblioteke te gebruik van al die voorspelde gedeeltes, kon een potensiële epitoop gedeelte in die EDIII proteïen geïdentifiseer word. Voorspelde epitopiese en antigeniese gedeeltes oor die lengte van die kapsied, NS4a en EDIII proteïene van elke stam is voorspel deur gebruik van die BCPREDS en ABCpred sagteware.

Ter afsluiting. Die EDIII proteïen, ‘n immuundominante antigeen van geelkoors, wat in hierdie studie voorberei is, het effense potensiaal vir die onderskeiding van flavivirus antiliggaampies

alhoewel dit nie sensitief genoeg is vir roetine diagnose nie. ‘n Potensiële epitoop, TGHGTVVMQ, van aminosuur 21 tot 29 op die EDIII proteïen is geïdentifiseer en deur gebruik van bioinformatika is aangetoon dat dit reaktiwiteit teen immuun sera het. Die belang van hierdie epitoop moet egter verder bestudeer word. Uiteindelik toon die EDIII gedeelte van die geelkoors proteïen potensiaal as ‘n teiken gedeelte vir entstof ontwikkeling soos aangetoon vir ander flavivirusse, maar wat nog nie voorheen vir geelkoors gepubliseer is nie.a