Engineering yeast strains for the expression of South African G9P[6] rotavirus VP2 and VP6 structural proteins

By

Mohau Steven Makatsa

Dissertation submitted in fulfilment of the requirements for the degree

Magister Scientiae

in the

Department of Microbial, Biochemical and Food Biotechnology

Faculty of Natural and Agricultural Sciences

University of the Free State

Bloemfontein

South Africa

Supervisor: Dr. Hester G. O'Neill

Co-supervisor: Prof. Jacobus Albertyn

February 2015

I can do all things through Christ who strengthens me.

<div align="right">

- Philippians 4:13

</div>

Problems are not stop signs, they are guidelines.

<div align="right">

- Robert H. Schuller

</div>

Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.

<div align="right">

- Thomas A. Edison

</div>

The secret of getting ahead is getting started.

<div align="right">

- Mark Twain

</div>

To my daughter Reitumetse Makatsa, you are the greatest blessing in my life

Declaration

"I Mohau Steven Makatsa declare that the dissertation hereby submitted by me for the Magister Scientiae degree at University of the Free State is my own independent work and has not previously been submitted by me at another university/faculty. I further more cede copyright of the dissertation in favour of University of the Free State"

Signature: _____

Mohau Steven Makatsa

Department of Microbial, Biochemical and Food Biotechnology,

University of the Free State,

South Africa.

Acknowledgements

Contents

CHAPTER 3

Construction of dual expression vectors containing yeast codon-optimized sequences encoding rotavirus proteins VP2 and VP6.. 23

CHAPTER 4

Engineering recombinant yeast strains and expression of rotavirus VP6 protein in yeast. 46

List of Figures

List of Tables

Abbreviations

ASC: Antibody secreting cell

attHRV: Attenuated Wa human rotavirus

BVMO: Baeyer–Villiger monooxygynase

bp: base pairs

Da: dalton

DLPs: Double-layered particles

DNA: deoxyribonucleic acid

ds: Double-stranded

dsRNA: Double-stranded Ribose Nucleic Acid

EDTA: ethylene-diamine-tetra-acetic acid

EIA: Enzyme immunoassay

ER: Endoplasmic reticulum

FBS: fetal bovine serum

hph: Hygromycin B resistance gene

IFN: Interferon

IgA: Immunoglobulin A

IgG: Immunoglobulin G

IIR: Innate immune responses

IN: Intranasal

MDA5: Melanoma differentiation-associated gene 5

NEAA: Non-essential amino acids

NSPs: Non-structural proteins

ORS: Oral dehydration Salts

ORT: Oral dehydration therapy

PAGE: Polyacrylamide gel electrophoresis

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

RER: Rough endoplasmic reticulum

RLP: Rotavirus-like particles

RNA: ribonucleic acid

RV: Rotavirus

SA11: simian agent 11

SDRs: short-chain dehydrogenase/reductase genes

ss: Single-stranded

VHH: Llama-Derived Antibody Fragments

VLPs: Virus-like proteins

TAE: Tris Acetate EDTA

TBST: Tris-buffered saline with Tween-20

TGS: Tris Glycine SDS

TLPs:Triple-layered particles

tlRLPs: Triple-layered rotavirus virus-like particles

T_M: Melting temperature

U: Unit

USA: United States of America

Conference presentations

Makatsa, M.S., Albertyn, J., O'Neill, H.G. Expression of rotavirus VP2 and VP6 structural proteins in yeast. SASM 2013, Forever Resorts Warmbaths, Bela-Bela, 24-27 November 2013. (Poster Presentation)

CHAPTER 1

Introduction to the study

1.1   Background

Rotavirus is the leading cause of severe gastroenteritis in children less than five years of age worldwide. In 2008, rotavirus was reported to be responsible for 453 000 deaths out of 1.3 million deaths caused by diarrhoea in children younger than five years. More than 50% of these deaths occurred in Asia and Sub-Saharan Africa (Tate *et al.*, 2012).

Rotaviruses were first visualized in 1958 in a vervet monkey, *Cercopithecus aethiops pygerythrus,* and were grouped as simian agent (SA) 11 based on cytopathic effects observed in vervet kidney tissue cultures stained with haematoxylin and eosin (Malherbe & Harwin, 1963). In 1973, Bishop and co-workers discovered that SA11 has the same structure as the infectious agent from the duodenal epithelium of a child suffering from diarrhoea. This agent was then named rotavirus based on the virion structure (rota, means "wheel" in Latin) (Bishop *et al.*, 1973). This virus is a member of the *Reoviridae* family which is a group of double stranded RNA viruses. It contains a segmented genome and consists of six structural (VP) and six non-structural (NSP) proteins (Estes, 2013).

Rotavirus infection is zoonotic, causing disease in both humans and animals. The main mode of rotavirus transmission is by the faecal-oral route. Transmission can occur as a result of person-to-person contact, respiratory secretions or contaminated environment (Parashar *et al.*, 1998). In temperate climates rotaviruses are more prominent in winter but in tropical countries rotavirus infections are observed throughout the year (Patel *et al.*, 2012). Rotaviruses are shed in large numbers during episodes of diarrhoea and usually are detectable by antigen enzyme immunoassay (EIA) up to one week after infection or more than 30 days in immunocompromised patients (Parashar *et al.,*. 1998).

There are currently two globally licensed rotavirus vaccines named RotaTeq® (MERCK) and Rotarix^TM (GlaxoSmithKline), both of which have been recommended for global use by the World Health Organisation (WHO). The development of RotaTeq®, a pentavalent vaccine, was

based on a bovine WC3 strain with genotype G6P7[5]. It comprises of five reassortants in which the external capsid protein, VP7, of WC3 is replaced by human VP7 (G1-G4) and the spike protein, VP4, replaced by human strain P[8]. The monovalent vaccine (Rotarix™) was developed by attenuation of the human rotavirus 89-12, a G1P1A[8] strain (Ruiz-Palacios *et al.*, 2006). This vaccine provides heterologous protection against at least G3P[8], G4P[8] and G9P[8] (Vesikari *et al.*, 2007). Both these vaccines have been found to be safe and effective against severe diarrhoea caused by rotavirus (Bhandari *et al.*, 2006). A very low risk associated with intussusception has been identified (Patel *et al.*, 2011). These vaccines are however expensive so roll-out in low-income countries, where the need is the highest, is slow. Introduction of the vaccines in Sub-Saharan Africa occurs mostly with support from the Global Alliance for Vaccines and Immunisation (GAVi).

1.2   Problem identification

Rotavirus remains a worldwide problem as a cause of severe gastroenteritis in children under five years (Tate *et al.*, 2012). Although there are currently two worldwide licensed vaccines (RotaTeq® and Rotarix™) and two local vaccines (China (LLR) and India (116E)) against rotavirus that have been proven to be effective and safe, a need exists for an efficacious and low-cost vaccine for use in Africa (Bhandari *et al.*, 2014; Fu *et al.*, 2012; Goveia *et al.*, 2011; Heaton *et al.*, 2005).

The use of live attenuated rotavirus vaccines creates a risk of intussusception and reassortment (Murphy *et al.*, 2001; Zanardi *et al.*, 2001). If the vaccine strains co-infect a host cell that is infected with a distinct rotavirus strain, they can exchange genome segments (reassort) and form novel strains that could be more virulent. Shedding of RotaTeq® vaccine strains in infants experiencing gastroenteritis after a recent vaccination with RotaTeq® vaccine has been reported (Ruiz-Palacios *et al.*, 2006, Vesikari *et al.*, 2006, Donato *et al.*, 2012). In Sub-Saharan Africa there is a high level of mixed rotavirus infections (Mwenda *et al.*, 2010), which creates a possibility for more reassortment and more rotavirus strain diversity.

Virus-like particles (VLPs) are formed when viral structural proteins are expressed in a heterologous system (Lawton *et al.*, 1997). Such particles do not contain the genetic material of

the virus. Production of VLPs provide an alternative approach to the development of rotavirus vaccines as they mimic the overall structure of virus particles, while having the ability to bind and enter cells using appropriate receptors and provoking immune response (Crawford *et al.*, 1994). Production of rotavirus-like particles (RLPs) in insect cells is regarded as the gold standard for RLPs production, but production of VLPs in insect cells is an expensive and cumbersome process (Palomares & Ramírez, 2009). Heterologous protein production in yeast provides an easy and cost-effective alternative for production of RLPs (Rodríguez-Limas & Tyo, 2011).

1.3   Rationale

Rotavirus infects approximately 95% of children by the age of three to five years, worldwide. Rotavirus is present in both developing and well-developed countries, implying that improving sanitation is not the answer to prevent rotavirus infection (Parashar *et al.*, 1998).

There is currently no treatment for rotavirus; the major strategy for rotavirus treatment is oral rehydration therapy (ORT). Infected patients can also be given oral ingestions of immunoglobulins containing antibodies against rotavirus serotypes (Sarker *et al.*, 1998). The use of live attenuated rotavirus vaccines has been effective in controlling rotavirus disease. However, live attenuated vaccines are associated with adverse effects such as intussusception and there is also evidence of reassortment (Iturriza-Gómara, 2001; Matthijnssens *et al.*, 2008; Murphy *et al.*, 2001; Patel *et al.*, 2011; Payne *et al.*, 2010; Zanardi *et al.*, 2001). Consequently, the use of virus-like particles (VLPs) as alternative vaccines has been studied over years (Federico, 2010). VLPs resemble the virion without viral genetic material. Production of VLPs requires expression in a heterologous expression system (Lawton *et al.*, 1997).

Yeasts provide an attractive alternative for production of RLPs and it is advantageous because of the low cost of production, ability to express heterologous proteins and ease of scale up (Federico, 2010). Triple-layered rotavirus virus-like particles (tlRLPs) have been successfully produced in yeast (*Saccharomyces cerevisiae*) (Rodríguez-Limas *et al.*, 2011). In this study, are using the similar approach but we are expressing rotavirus particles in a range of different yeast strains to identify the yeast strains that best express these proteins and also safe to use for

human vaccine development as using *S. cerevisiae* to produce human vaccines has setbacks which include that glycoproteins are often over-glycosylated, and terminal mannose residues in N-linked glycans are added by an α -1,3 bond which is suspected to be allergenic (Jigami & Odani,1999).

1.4   Preliminary data

The wide-range yeast expression system containing 18S rDNA from genomic *Kluyveromyces marxianus*, *Yarrowia lipolytica* TEF promoter, *K. marxianus* inulinase region and the hygromycin B resistance gene (hph) has been developed by researchers at UFS (Albertyn *et al.*, 2011). The rotavirus strain RVA/HUMAN-WT/ZAF/GR10924/1999/G9P[6] was molecularly characterized from a stool sample of a neonate experiencing severe diarrhoea at the Dr. George Mukhari Hospital, University of Limpopo, Medunsa Campus, South Africa (Jere *et al.*, 2011). The whole genome consensus sequence of this strain was obtained using GS20/FLX technology (Potgieter *et al.*, 2009). Double-layered RLPs (VP2/6), triple-layered (VP2/6/7 or VP2/6/7/4) and chimeric tlRLPs were produced in insect cells using codon-optimized sequences of genome segments 2 (VP2), 4 (VP4), 6 (VP6) and 9 (VP7) for insect cell expression (Jere *et al.*, 2014).

Wild-type coding regions of VP2 and VP6 from strain RVA/Human-wt/ZAF/GR10924/1999/G9P[6] were obtained from GenScript and cloned into the yeast expression vectors, pKM173 and pKM177, respectively. A dual expression vector was generated by cloning the VP6 coding region containing expression cassette into a pKM173 vector containing the VP2 coding region, yielding a yeast expression vector capable of expressing both VP2 and VP6. Eight yeast strains; named *Kluyveromyces marxianus, Kluyveromyces lactis, Debaryomyces hansenii, Yarrowia lipolytica, Hansenula polymorpha, Pichia pastoris Candida deformans* and *Arxula adeninivorans* collected from the UNESCO-MIRCEN yeast culture collection at the UFS were selected for evaluation. *Saccharomyces cerevisiae* was included as a positive control. Protein expression was evaluated using western blot analysis, but no protein expression was obtained in any of the yeast strains tested. Therefore, optimized open reading frames (ORF) were preferential for further cloning and expression in yeast.

1.5 Aim and Objectives

The aim of this study was to assess different yeast strains for the expression of rotavirus VP2 and VP6 in the production of double-layered rotavirus-like particles, using the open reading frames (ORFs) encoding rotavirus VP2 and VP6 structural proteins from the RVA/HUMAN-WT/ZAF/GR10924/1999/G9P[6] strain.

Specific objectives of the study were to:

I. Construct yeast expression vectors containing the yeast optimized VP2 ORFs (pKM173 vector) and VP6 ORFs (pKM177 vector).
II. Engineer yeast strains capable of expressing rotavirus VP6 and VP2/VP6.
III. Generate an appropriate rotavirus VP6 control to assist with the evaluation of VP6 expression by the recombinant yeast strains.
IV. Examine expression of rotavirus VP6 protein by yeast strains using western blot analysis.

1.6 Dissertation Structure

This dissertation is divided into seven chapters; chapters 3 and 4 comprise a brief introduction, materials and methods, results and discussion. The dissertation structure is as follows:

CHAPTER 1 Introduction to the study

CHAPTER 2 Literature Review

CHAPTER 3 Construction of dual expression vectors containing yeast codon-optimized sequences encoding rotavirus proteins VP2 and VP6

CHAPTER 4 Engineering recombinant yeast strains and expression of rotavirus VP6 protein in yeast

CHAPTER 5 Concluding remarks and future prospects

CHAPTER 6 Summary

CHAPTER 7 Opsomming

CHAPTER 2

Literature Review

2.1 Introduction

Rotavirus-induced diarrhoea is responsible for 453 000 annual deaths in children under five years of age worldwide (Black *et al.*, 2010; Parashar *et al.*, 2003). Almost all the deaths due to rotavirus diarrhoea occur in children living in developing countries (Tanaka *et al.*, 2007).

The introduction of rotavirus vaccines has been identified as the best strategy to reduce rotavirus disease burden (WHO, 2000). Primary rotavirus infections can lead to disease that ranges from mild gastroenteritis to severe or fatal diarrhoea with dehydration. This primary infection can lead to protection from subsequent infections as a result of induced immunity against rotavirus disease; immunity increases with subsequent infections (Velázquez *et al.*, 1996).

2.2 Molecular Biology

Rotavirus is a member of the *Reoviridae* family, which is a group of icosahedral double-stranded RNA (dsRNA) viruses. The genome comprises of 11 genome segments encoding six structural (VP1, VP2, VP3, VP4, VP6 and VP7) and six non-structural proteins (NSP1-NSP6) (Figure 2.1a). The rotavirus structure is divided into three complex layers, namely the inner layer (VP1, VP2 and VP3), the middle capsid (VP6) (Figure 2.1c), and the outer capsid (VP7 and projections of VP4) (Figure 2.1b). Each genome segment codes for a single viral protein, except segment 11 which codes for NSP5 and NSP6 (Estes, 2013).
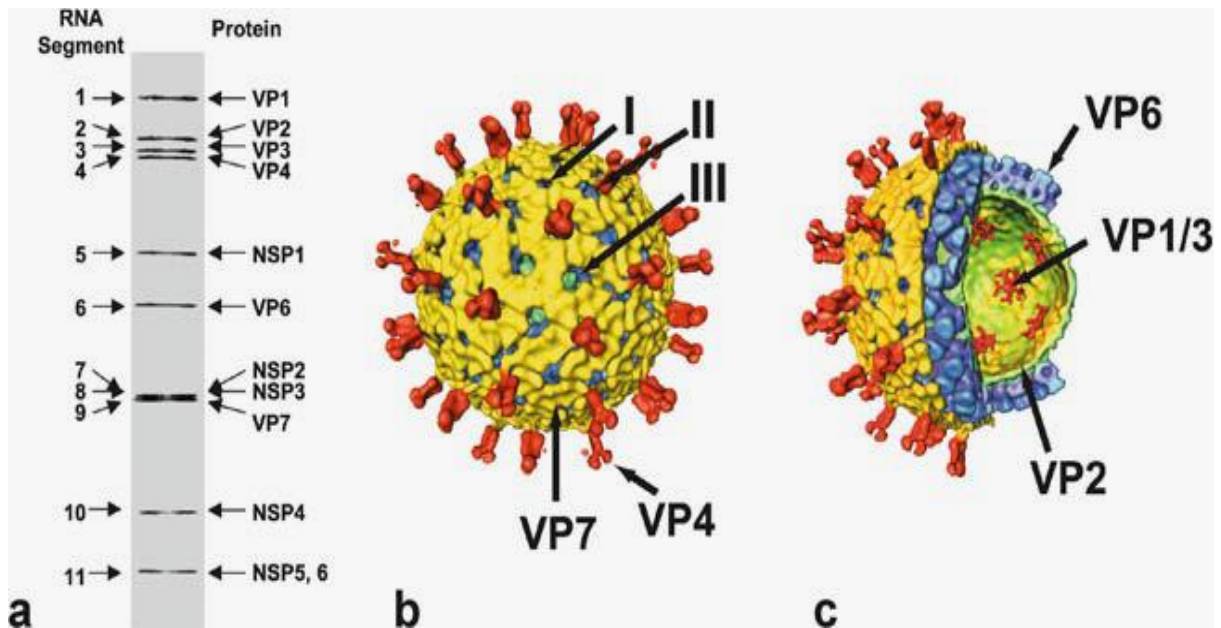
Figure 2.1: Aspects of rotavirus structural composition (a) RNA genome segments on the left numbered according to polyacrylamide gel migration and encoded proteins on the right. (b) Surface presentation of proteins VP7 (yellow) and VP4 (red) and three aqueous channels designated I, II and III. (c) Cut TLP structure showing middle capsid VP6 (blue), inner layer VP2 (green) and VP1/3 complex (red) (Taken from Pesavento *et al.*, 2006).

Table 2.1 summarises the known protein functions of rotavirus proteins. VP6 is the most abundant viral protein. It is the major determinant of group specificity and target of common diagnostic assays (Parashar *et al.*, 1998).

Table 2.1: Summary of characteristics and functions of rotavirus proteins

Protein	Function	References
VP1	RNA-dependent RNA polymerase; ssRNA binding; located at the five–fold axis inside the inner capsid; forms a transcription complex with VP3.	Zeng *et al.*, 1996
VP2	Inner capsid structural (core) protein; non-sequence-specific RNA-binding activity; required for replicase activity of VP1.	Bican *et al.*, 1982; Boyle & Holmes, 1986
VP3	Guanylyltransferase and methyltransferase; part of the virion transcription complex with VP1.	Anthony *et al.*, 1991; Liu *et al.*, 1992; Pizarro *et al.*, 1991
VP4	Outer capsid spike protein; P-type-specific neutralization antigen; virulence determinant; haemagglutinin; cell-attachment protein; cleavage by trypsin into VP5* and VP8* enhances infectivity.	Anthony *et al.*, 1991; Ericson *et al.*, 1983; Fiore *et al.*, 1991; Greenberg & Flores, 1983; Kalica *et al.*, 1983
VP6	Major virion protein; middle capsid structural protein; subgroup antigen.	Greenberg & McAuliffe, 1983; Prasad *et al.*, 1988
VP7	Outer capsid structural glycoprotein; G-type neutralization antigen.	Ericson *et al.*, 1983; Greenberg & Flores, 1983
NSP1	Associates with the cytoskeleton; antagonist of the host innate immune response system	Graff *et al.*, 2002
NSP2	NTPase and helicase activity; non-specific ssRNA binding; involved in viroplasm formation; essential for dsRNA synthesis.	Taraporewala *et al.*, 1999

NSP3	Homodimer, involved in translational regulation.	Mattion *et al.*, 1992, Padilla-Noriega *et al.*, 2002
NSP4	Viral enterotoxin; receptor for budding of double-layered particles through the ER membrane during virus replication.	Au *et al.*, 1989, Ball *et al.*, 1996, Ericson *et al.*, 1983
NSP5	Interacts with NSP2 and NSP6; forms homo-multimers; O-linked glycosylation.	Afrikanova & Fabbretti, 1998
NSP6	Product of the second out-of-frame open-reading frame of genome segment II; interacts with NSP5; localizes to the viroplasm.	González *et al.*, 1998

ER= endoplasmic reticulum; ss= single-stranded, * resulting from VP4 trypsin cleavage

2.3   Classification of rotavirus

Rotaviruses are classified into seven groups (A-G) with possibly an additional group (H) using the middle capsid VP6 protein. Group A is further divided into G types (based on the glycoprotein VP7) and P types (based on the protease-sensitive VP4) (Matthijnssens *et al.*, 2012). Currently, at least 27 G types and 37 P types have been identified (Matthijnssens *et al.*, 2012; Trojnar *et al.*, 2013). Group A rotaviruses are the most common cause of severe gastroenteritis in humans. In this group; types G1, G3, G4 and G9 with P[8] VP4 specificity and G2P [4] are most common worldwide (Santos *et al.,* 2005).

In 2008, Matthijnssens and co-workers proposed a full genome-based classification of rotaviruses to elucidate the epidemiology of rotavirus. The suggested nomenclature is: Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, representing the genotypes of, respectively, the VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5 genes, with x indicating the numbers of the corresponding genotypes (Matthijnssens *et al.*, 2008). The full genome classification system is

recommended in order to clearly assign novel strains into known genotypes or determine a new genotype; complete ORFs nucleotide sequences should be obtained for this purpose. Guidelines for classification of RV strains are: (a) wild type, (b) tissue culture-adapted or *in vivo* passaged, (c) generated in a laboratory for which host species can be assigned unambiguously, (d) generated in a laboratory for which host species cannot be assigned unambiguously and (e) vaccine strains have been proposed (Matthijnssens *et al.*, 2012).

2.4   Replication and Pathogenesis

Rotavirus enters the body through contact with water, food or any object contaminated with rotavirus. VP4 and VP7 facilitate virion entry into host cells (Figure. 2.2) (Fleming *et al.*, 2014). The infections triple-layered rotavirus particle uses VP4 spikes to interact with cellular receptors and then undergo a conformational change exposing the lipophilic domains of VP5* which are normally hidden below VP8* (Kim *et al.*, 2010; Settembre *et al.*, 2011). Rotavirus particles use their own transcription complexes comprising of the viral RNA-dependent RNA polymerase (RdRp) (VP1) and the viral capping enzyme (VP3). DLPs in the cytoplasm give rise to capped mRNA from all 11 genomic segments which are translated into proteins and replicated into new genomic RNA (Silvestri *et al.*, 2004). The non-structural proteins NSP2 and NSP5 localize in viroplasms together with other viral proteins, including the polymerase VP1, VP3 and VP2. Interaction between NSP2 and NSP5 activates NSP5 hyper-phosphorylation and the formation of viroplasm-like structures (VLSs) (Eichwald *et al.*, 2004; Fabbretti *et al.*, 1999). After the newly assembled DLPs leave the viroplasms, they bud through the endoplasmic reticulum (ER) for maturation. In this process, a transmembrane glycoprotein (NSP4), which is mainly located in the ER acts as intracellular receptor for maturation of DLPs into mature triple-layered particles (TLPs) by interacting with VP6. Mature TPLs are then released by cell lysis (Taylor *et al.*, 1996).

Rotavirus infection results in loss of digestive enzymes which results in reduction of intestinal absorption, glucose and water which ultimately leads to diarrhoea (Uhnoo *et al.*, 1986). Infection with rotavirus can be symptomatic or asymptomatic. Symptomatic infections are prominent in children less than five years of age. Rotavirus infection has an incubation period of less than 48 hours with a sudden onset of vomiting, a high frequency of dehydration and

diarrhoea lasting five to six days. Rotavirus diagnosis is usually carried out using enzyme-linked immune assays detecting VP6 antibodies (Desselberger *et al.*, 2009). Re-infection of rotavirus can occur at any time and subsequent infection results in less severe disease (Velázquez *et al.*, 1996). This observation is what indicated that more than one dose of rotavirus vaccines would be needed for higher efficacy.
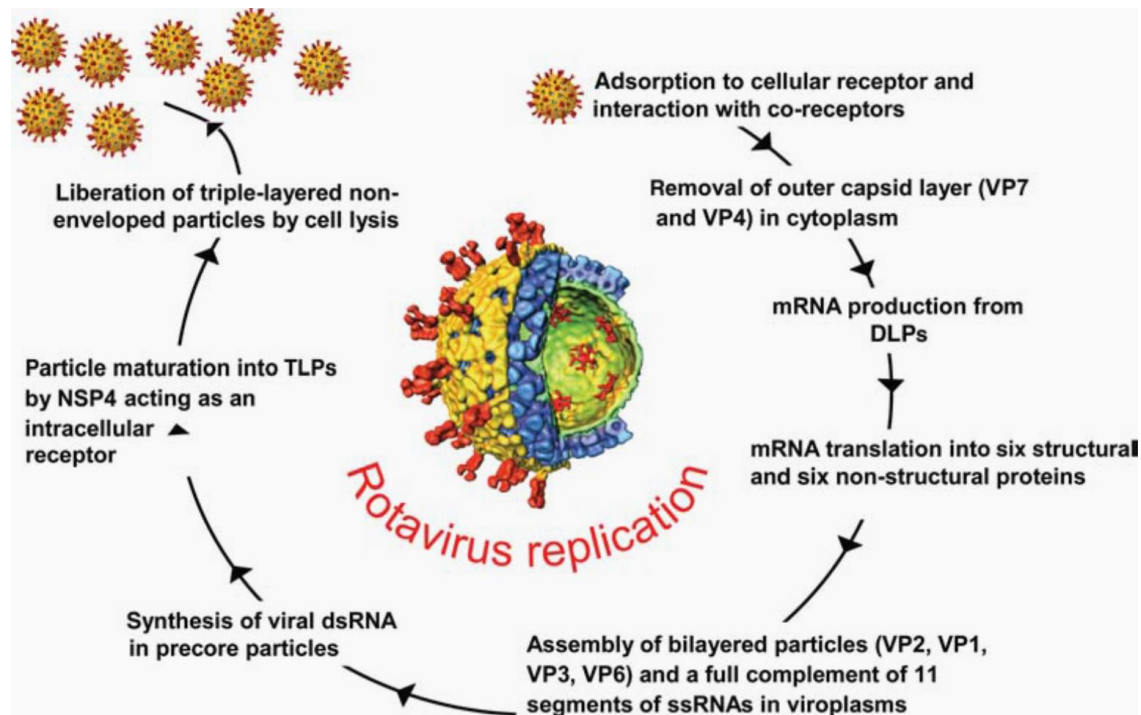
Figure 2.2: An overview of rotavirus replication cycle (Taken from Desselberger *et al.*, 2009).

2.5 Epidemiology and Prevalence

Rotaviruses are shed in high concentrations in stool of infected persons. The mode of transmission of rotavirus is the faecal-oral route; the virus spreads from person-to-person, by fomites and probably by contaminated food or water (Parashar *et al.*, 1998).

Rotavirus is prominent during the winter in temperate climates but it is not seasonal in tropical countries as it is detected all year around (Stoll *et al.*, 1982). In South Africa, rotavirus infections occur throughout the year and are more prominent during winter months (Steele *et al.*, 2003).

Rotavirus is a worldwide problem. The prevalence of rotavirus is diverse depending on a specific country. In 2004 rotavirus was estimated to be responsible for 527 000 deaths (95%

confidence interval, 475 000–580 000 deaths) in children less than five years of age (Parashar *et al.,* 2009). This value is different to the 2008 estimate reported by Tate and colleagues where rotavirus caused 453 000 deaths (Tate *et al.*, 2012). The difference is also observed in estimated deaths due diarrhoea which decreased from 1.8 million in 2003 to 1.3 million in 2008 (Black *et al.*, 2010; Tate *et al.*, 2012). Differences in estimates could be because of methods used to obtain the estimates and/or better management of diarrhoea.

Although improvement in sanitation and hygiene has high impact on diarrhoea caused by bacteria and parasites, it does not have much influence on rotavirus diarrhoea. Although both studies used the data before the introduction of vaccines, they did not use the same data, which could be the reason for different numbers. Better nutrition and early diagnosis could also explain the decline in rotavirus diarrhoea from 527 000 in 2004 to 453 000 in 2008 (Parashar *et al.*, 2009; Tate *et al.*, 2012).

Rotavirus surveillance is carried out by PCR-genotyping. In 1973 to 2003, the rotavirus strain, P[8]G1, of a globally common rotavirus P-G combination (Fig. 2.3), was responsible for 52% of the rotavirus diarrhoea among children worldwide, but only 17% of the rotavirus diarrhoea among children in Africa (Santos & Hoshino, 2005). The distribution of various P-G combinations varies drastically from one continent to another. Strains of other P-G combinations that are considered unusual have also been described all over the world and the percentages of such unusual strains were much higher in Africa (27%), Asia (14%) and South America (11%) than in North America, Europe and Australia (5%, 1.4% and 0.1%, respectively). Rotavirus strains carrying P[4]G1, P[8]G2, P[4]G3, P[9]G4, P[4]G4, P[4]G9, P[10]G9, P[6]G12 or P[9]G12 specificity may represent naturally occurring reassortants among various human rotavirus genotypes (Gouvea & Brantly, 1995; Iturriza-Gómara, 2001). In a study comprising of 2 555 of rotavirus enzyme immunoassay (EIA) positives in Africa, G1 was the most predominant (28.8%), followed by G9 (17.3%), G2 (16.8%), G8 (8.2%), G12 (6.2%) and G3 (5.9%). Similarly, the P[8] strain was the most prevalent (40.6%), followed by P[6] (30.9%) and P[4] (13.9%). The highest G/P combinations detected were G1P[8] (18.4%), G9P[8] (11.7%), G2P[4] (8.6%), G2P[6] (6.2%), G1P[6] (4.9%), G3P[6] (4.3%), G8P[6] (3.8%) and G12P[8] (3.1%) (Seheri *et al.,* 2014).

In Sub-Saharan Africa (countries: Ghana, Kenya, Uganda, Zambia, Cameroon, Ethiopia, Tanzania and Zimbabwe) it was shown that rotavirus has a 12% rate of mixed infections where a total of 2 200 rotavirus-positive specimen were tested in a period of two years (2006-2008) (Mwenda *et al.*, 2010). Predominant types included G1P[8] (21%), G2P[4] (7%), and P [8] (29%); however, unusual types were also detected, including G8P[6] (5%), G8P[8] (1%), G12P[6] (1%), and G12P[6] (1%) (Mwenda *et al.*, 2010).

2.6   Immune response

Rotavirus infection elicits an innate immune responses (IIR) immediately after primary infection (Angel *et al.*, 2012). Mechanisms of this IIR is poorly understood, however, it has been shown that NSP1 interacts with certain cellular proteins which include: interferon (IFN) regulatory factors (IRF) 3, melanoma differentiation-associated gene 5 (MDA5)/mitochondrial anti-viral signalling protein (MAVS) (Nandi *et al.*, 2014; Sen *et al.*, 2011), the tumour suppressor protein p53 and the TNF receptor associated factor 2 (TRAF2), leading to their proteasomal degradation and thus preventing or down-regulating the early triggering of an IFN response (Bagchi *et al.*, 2013; Bhowmick *et al.*, 2013).

Rotavirus infection also elicits acquired immune responses from B cells producing antibodies directed against virus-specific proteins, and from T cells recognizing T cell-specific rotavirus epitopes (Franco & Greenberg, 1997; Jiang *et al.*, 2002).

Infection with rotavirus usually confers a level of protection against subsequent infections; the level of protection is higher against moderate to severe disease. Complete protection may result after at least two to three subsequent infections. This finding suggested that at least two doses of vaccine is required to induce sufficient protection (Fischer *et al.*, 2002; Gladstone & Ramani, 2011; Velázquez *et al.*, 1996). Protection can also occur as a result of rotavirus-specific antibodies passed transplancentally which result in asymptomatic or mild disease in most neonates (Bishop *et al.*, 1983). However, these antibodies interfere with immune responses to rotavirus vaccination (Appaiahgari *et al.*, 2014; Johansson *et al.*, 2008). Serum IgA or IgG can be used to indicate rotavirus immunity after infection and vaccination. Although the correlates of protection is not known, rotavirus-specific IgA and IgG neutralizing antibodies elicited by VP4

and VP7 are associated with protection (Desselberger & Huppertz, 2011; Franco *et al.*, 1996; Velázquez *et al.*, 2000).

VP6 is the major rotavirus protein that elicits a human antibody response (Svensson *et al.*, 1987). Human antibodies against rotavirus VP6 protein have also been reported to inhibit viral transcription by interacting with intracellular DLPs (Aiyegbo *et al.*, 2013). It has been suggested that mucosal RV VP6-specific IgA plays a significant role for the inhibition of RV replication *in vitro* and *in vivo* using IgA antibody depletion and blocking experiments using recombinant VP6 which confirmed that neutralization was mediated by anti-VP6 IgA antibodies (Lappalainen *et al.*, 2014). Oral administration of G6P[1] VP6-specific llama-derived antibody fragments (VHH nano Abs) against Wa G1P[8] human RVA-induced diarrhoea was shown to be reactive and effective in neonatal gnotobiotic pigs (Vega *et al.*, 2013).

2.7   Treatment

There is currently no therapy for rotavirus infection. Oral rehydration therapy (ORT) is usually used to treat dehydration caused by rotavirus-induced diarrhoea. ORT is based on the use of oral rehydration salts (ORS) for replacement of fluids and electrolytes lost due to diarrhoea (Santosham *et al.*, 1997). Factors such as breastfeeding, female education, measles immunization and socioeconomic status may also have an impact on rotavirus control (Victora *et al.*, 2000). Breastfeeding has been indicated to reduce the risk of rotavirus infection (Plenge-Bönig *et al.*, 2010); the antiviral activity of breast milk may be because breast milk contains bioactive glycans that inhibit pathogens, this includes lactadherin which inhibits rotavirus (Newburg, 2009); or/and presence of cytokines (Chirico *et al.*, 2008) and maternal antibodies, especially immunoglobulin A (IgA) (Moon *et al.*, 2010). The presence of these bioactive glycans in breast milk results in lower immunogenicity and efficacy of the current live vaccines (Groome *et al.*, 2014). Current research suggests that probiotic treatment reduces the duration of diarrhoea caused by rotavirus infection (Grandy *et al.*, 2010); *Bifidobacterium lactis* HN019 confers protection against weanling diarrhoea (Shu *et al.*, 2001).

2.8   Vaccines

Several approaches have been used to develop rotavirus vaccines. These include live attenuated vaccine developments namely: Rotashield (Murphy *et al.*, 2001; Zanardi *et al.*, 2001), Lanzhou lamb rotavirus (LLR) (Fu *et al.*, 2007), RotaTeq® (Heaton *et al.*, 2005), RV3 vaccine (RV3 strain P[6]G3) (Barnes *et al.*, 2002), 116E vaccine (116E strain P[11]G9)) (Bhandari *et al.*, 2014) and I321 vaccine (I321 strain P[11]G10) (Glass *et al.*, 2005) and Rotarix™ (Ruiz-Palacios *et al.*, 2006), and non-live vaccine development which mainly focuses on the use of rotavirus particles as vaccines (Jiang *et al.*, 2008).

2.8.1   Live attenuated rotavirus vaccines

Studies to develop rotavirus vaccines began in the mid-1970s. Three monovalent vaccines, bovine NCDV strain P[1]G6, WC3 bovine WC3 strain P[5]G6 and rhesus rotavirus vaccine (RRV) rhesus MMU 18006 strain P[3]G3 were developed (Christy *et al.*, 1988; Clark *et al.*, 1988; Vesikari *et al.*, 1984) but were discontinued because of inconsistent efficacy results (Hanlon *et al.*, 1987; Lanata *et al.*, 1989). In 2000, a monovalent live oral Lanzhou lamb rotavirus (LLR) strain P[12]G10 rotavirus vaccine was developed, licensed and it is currently used in China. The efficacy of the LLR vaccine has not been evaluated in a randomized controlled trial. One case-controlled study showed that the vaccine confers 73.3% protection against rotavirus diarrhoea requiring hospitalization (Fu *et al.*, 2007). A recent study indicates that one dose of the LLR vaccine confers partial human protection against rotavirus disease (Fu *et al.*, 2012).

A monovalent live oral human rotavirus vaccine was developed (GlaxoSmithKline) by tissue culture passage of a wild-type human rotavirus isolate 89-12 strain P[8]G1 (Bernstein *et al.*, 1999). This vaccine (Rotarix™) was evaluated for association with intussusception in a large-scale, double blind, placebo-controlled trial of more than 63 000 participants from 11 Latin American countries (Argentina, Brazil, Chile, Colombia, the Dominican Republic, Honduras, Mexico, Nicaragua, Panama, Peru and Venezuela) and Finland. There was a lack of association in vaccination and intussusception. In this study 20 000 infants were monitored for vaccine efficacy. The results indicated a protection rate of 85% against severe rotavirus gastroenteritis and 100% protection against the most severe rotavirus gastroenteritis episodes (Ruiz-Palacios

et al., 2006). The results also demonstrated that Rotarix[TM] is efficacious in preventing rotavirus gastroenteritis of any severity caused by the predominant G1 serotype (92% efficacy) and serotypes G3, G4, or G9 (88% efficacy)(Ruiz-Palacios *et al.*, 2006).

In another study from six European countries (Czech Republic, Finland, France, Germany, Italy and Spain) with 3994 participants, Rotarix[TM] had an efficacy of 79% against rotavirus gastroenteritis of any severity, 90% against severe rotavirus disease and 96% against hospitalization due to rotavirus. For severe rotavirus gastroenteritis, the vaccine had efficacies of 96% against G1P[8] and 88% against non-G1P[8] RV strains (Vesikari *et al.*, 2007). In a study conducted in South Africa and Malawi, the efficacy of the vaccine was 76.9% and 49.4% respectively but the number of cases of severe rotavirus diarrhoea prevented were higher in Malawi (6.7 cases per 100 vaccinated infants per year) than South Africa (4.2 cases per 100 vaccinated infants per year) (Madhi *et al.*, 2010). Rotarix[TM] was first licensed in 2004 in Mexico and the Dominican Republic and was later (2007) approved in 90 countries worldwide including South Africa. The vaccine is administered as two oral doses at two and four months of age (Dennehy, 2008).

The ability of rotavirus to reassort during mixed rotavirus infections *in vitro* permitted the development of reassortant vaccines (Kapikian *et al.*, 1986). Rotashield (Wyeth Ayerst (USA)) was the first multivalent vaccine to be developed and was licensed in 1998. Rotashield is a quadrivalent live oral human-rhesus vaccine (RRV) that incorporates rhesus rotavirus strain MMU 18006 (serotype G3) with human serotypes G1, G2 and G3. Rotashield was withdrawn from the market by the manufacturer in 1999 because of reports of cases of intussusceptions among recipients of the vaccine (Murphy *et al.*, 2001; Zanardi *et al.*, 2001). Intussusception with Rotashield was associated with the age of vaccine recipients at the time of vaccination. Babies receiving the first dose of the vaccine after 90 days of age, developed intussusception. This vaccine was found to be 60.7% efficacious against rotavirus disease of any severity cause by any rotavirus genotype in a study conducted in Ghana, West Africa. The vaccine was administered in a two dose schedule where the first dose was given to neonatal babies and the second dose given before the age of 60 days (Armah *et al.*, 2013).

In 2006, a pentavalent live oral human-bovine (WC3) reassortant vaccine (RotaTeq®) was licensed in the United States of America (USA). RotaTeq® (Merck (USA)) consists of five reassortants: four human rotavirus capsid proteins (G1, G2, G3, or G4) with bovine protein (P7[5]) and human rotavirus protein P1A[8] with bovine rotavirus protein G6 (Heaton *et al.*, 2005).

RotaTeq® is recommended as three oral doses at two, four, and six months of age; the vaccine has been licensed in over 95 countries including Africa, Australia, Canada, the European Union, Asia, and Latin America (Goveia *et al.*, 2011).

The efficacy of RotaTeq® against all G1-G4 and G9 rotavirus serotypes has been observed. In a study with 34 035 infants in the vaccine group and 34 003 in the placebo group, RotaTeq® had an efficacy of 74% (67%-79%) against rotavirus diarrhoea of any severity and an efficacy of 98% (90%-100%) against severe rotavirus diarrhoea (Table 2.2). The risk of intussusception was similar in vaccine and placebo recipients and it was estimated to be 1.6; 95 percent confidence interval, 0.4 to 6.4 (Vesikari *et al.*, 2006). In a large study consisting of 5 673 persons, RotaTeq® had an efficacy of 63% (44%-75%) against rotavirus diarrhoea of any severity and an efficacy of 88% (49%-99%) against severe rotavirus diarrhoea (Armah *et al.*, 2010; Ruiz-Palacios *et al.*, 2006). The efficacy of RotaTeq® is lower in Asian and African children through both the entire efficacy follow-up period of nearly two years (Asia: 48.3%; Africa: 39.3%) and the first year of life (Asia: 51.0%; Africa: 64.2%) (Armah *et al.*, 2010; Zaman *et al.*, 2010).

A candidate tetravalent live attenuated human-bovine (UK) reassortant rotavirus vaccine with serotypes UK_Wa (P7[5],G1), UK_DS1 (P7[5],G2), UK_P(P7[5],G3), and UK_ST3 (P7[5],G4) has been developed (Kapikian *et al.*, 1986; Midthun *et al.*, 1985). This vaccine is immunogenic in infants less than six months of age and was reported to be safe and well tolerated in the small number of adults and paediatric subjects (Clements-mann *et al.*, 1999).

Table 2.2: Efficacy of current worldwide licensed live vaccines in developing countries (adapted from Babji & Kang, 2012).

	Vaccine	No. of children enrolled	Percent efficacy (CI*)	Reference
Asia				
Taiwan	Rotarix[TM]	1 141	96.1 (85.1–99.5)	Phua *et al.*, 2009
Singapore		6 542	85.1–99.5	
Hong Kong		3 025	96.1 (85.1–99.5)	
Bangladesh	Rotateq®	1 136	42.7 (10.4–63.9)	Zaman *et al.*, 2010
Vietnam		900	63.9 (7.6–90.9)	
Africa				
South Africa	Rotarix[TM]	1 944	76.9 (56.0–88.4)	Madhi *et al.*, 2010
Malawi		1 030	49.4 (19.2–68.3)	
Ghana	Rotateq®	2 162	55.5 (28.0–73.1)	Armah *et al.*, 2010
Kenya		1 221	63.9 (5.9–89.8)	

*CI= confidence interval

Neonatal human strains have also been explored as rotavirus vaccine candidates because they appear to be naturally attenuated (Danchin *et al.*, 2013; Dennehy, 2008). At least two candidate monovalent live oral human neonatal vaccines are currently in evaluation. These include RV3 vaccine (RV3 strain P[6]G3) (Barnes *et al.*, 2002) and I321 vaccine (I321 strain P[11]G10) (Glass *et al.*, 2005). 116E vaccine (116E strain P[11]G9)) (Bhandari *et al.*, 2014) has been rolled out in India. RV3 has been proven to be safe and well tolerated in infants of up to three months of age but a small phase II study indicated that three doses of 10^5 PFU of the vaccine resulted in low immunogenicity (Barnes *et al.*, 2002). The two latter vaccines were developed in India. Two phase I studies in adult recipients involving both vaccines indicated that both vaccines are safe and well tolerated but they replicated poorly in recipients (Glass *et al.*, 2005). The 116E vaccine

has also been found to be safe and well tolerated in a study including 4 532 infants. It was launched in 2014 in India and it is cheaper than the two vaccines that are currently licensed worldwide. However, this vaccine has low efficacy of 53.6% against severe rotavirus diarrhoea. In a study of 4 532 infants that were assigned to receive the 116E vaccine and 2 267 infants that received placebo six cases of intussusception were reported for the vaccine group and two cases of intussusception were reported for the placebo group (Bhandari *et al.*, 2014).

2.8.2   Non-live rotavirus vaccines

Inactivated and subunit vaccines have been studied extensively as candidate rotavirus vaccines. These two approaches provide advantages over the currently licensed vaccines as inactivated and subunit vaccines are safer to use as they are not associated with any adverse effects that are associated with live attenuated vaccines such as intussusception, reassortment etc., and there are well-established techniques for development of inactivated and subunit vaccines (Jiang *et al.*, 2008).

Virus inactivation can be achieved by treatment with chemical substances or by heat inactivation. Rotavirus can be efficiently inactivated by treating with 5 mM EDTA or by heating at 50°C (Estes & Graham, 1979). Inactivated human rotavirus has been reported to protect against rotavirus shedding and induced rotavirus-specific IgG antibody and neutralizing activity in gnotobiotic pigs (Wang *et al.*, 2010).

Virus-like proteins (VLPs) are acquired when structural proteins are produced in a heterologous expression system without viral genetic material. VLPs are able to induce immune responses and therefore have successfully been used as vaccines, including HBcAg vaccine for hepatitis B virus (HBV) (Ludwig *et al.*, 2007) and two human papillomavirus (HPV) *Gardasil®* (Maver *et al.*, 2010) and *Cervarix®* (Romanowski, 2011) vaccines based on major capsid protein L1.

Different systems are available for production of VLPs. These include expression in bacteria, insect cells, and yeast (Gerngross, 2004). Baculovirus-based expression of rotavirus proteins in insect cells is regarded as the main method of production of rotavirus-like particles. Baculoviruses are the dominant virus affecting insects; these viruses are non-pathogenic to plants and mammals which make them great vectors in insect cell expression (Contreras-Gómez

et al., 2013). The baculovirus-insect cell expression system is based on insertion of a foreign gene into the viral genome using appropriate vectors and infecting insect cell cultures with the resulting recombinant virus (Contreras-Gómez *et al.*, 2013). Although insect cell expression can produce high levels of expression and results in correctly folded proteins, expression of recombinant proteins in insect cells is labour-intensive, time-consuming and expensive (Federico, 2010).

Bertolotti-Ciarlet and colleagues reported that 2/6-VLPs produced by a single (rotavirus simian SA11 genes 2 and 6) or dual (bovine RF strain VP2 and simian SA11 strain VP6) baculovirus expression vector was immunogenic and induced protection from challenge with wild-type murine rotavirus in adult mice (Bertolotti-Ciarlet *et al.*, 2003). In the neonatal gnotobiotic pig model, three doses of 2/6-VLP vaccine administered intranasal (IN) following initial prime with attenuated Wa human rotavirus (attHRV) gives a protection rate equal to three doses of live attHRV and higher antibody secreting cell (ASC) responses (Azevedo *et al.*, 2010). This suggests that VLP vaccines can also be used effectively as boosters to reduce the risk of adverse effects caused by live attenuated vaccines.

Chimaeric RLPs were produced from the consensus sequences of African rotaviruses (G2, G8, G9 or G12 strains associated with either P[4], P[6] or P[8] genotypes). Sequences encoding rotavirus proteins VP2, VP4, VP6 and VP7 were codon-optimized for expression in insect cells (Jere *et al.*, 2014).

Triple-layered rotavirus virus-like particles (tlRLPs) have been successfully produced in yeast (*Saccharomyces cerevisiae*) (Rodríguez-Limas *et al.*; 2011). The approach followed for expression of the three proteins was based on the cloning of the coding regions of the three proteins into three individual plasmids for expression in yeast. This approach was, however, shown to result in inadequate co-expression; cloning all three ORFs into one plasmid resulted in sufficient co-expression (Rodríguez-Limas *et al.*; 2011). *Saccharomyces cerevisiae* extracts containing rotavirus-like particles (RLP) were used as a vaccine candidate in an adult mouse model. Two doses of 1 mg of yeast extract containing rotavirus proteins resulted in an immunological response capable of reducing rotavirus replication after infection (Rodríguez-

limas *et al.*, 2014). These results support the use of RLPs as an alternative rotavirus vaccine and also indicate that yeast expression systems can be successfully used for production of RLPs.

2.9   Heterologous gene expression

When selecting an expression system, several factors have to be taken into consideration. These factors are as follows: the cost of production, the ability to control final product (post-translational modifications), time and labour required to produce a protein and the regulatory path to approve a produced protein in a given expression system (Gerngross, 2004).

S. cerevisiae was successfully used in the production of insulin and human papilomavirus (HPV) and hepatitis B vaccines. It has disadvantages of glycoproteins being over-glycosylated and terminal mannose residues in N-linked glycans are added by an α-1,3 bond which can be allergenic (Jigami & Odani, 1999). *P. pastoris* and *H. polymorpha* can be used instead to overcome allergenic properties as they comprise of terminal α-1,2 bonds which are non-allergenic (Bretthauer & Castellino, 1999). This indicates that all yeast systems have their own advantages and disadvantages and favour expression of certain proteins over others; therefore a range of yeast systems have to be considered for expression. In order to select a yeast system that is better suited for expression of a specific protein, multiple yeasts have to be assessed in parallel and a vector that can be targeted to facilitate expression in different systems would minimise workload. Such wide-range expression vectors are becoming available. A vector needs certain properties to be classified as a wide-range expression vector. Firstly, the element targeted by the vector for integration in the host cell has to be suitable for all the selected organisms. Secondly, the promoter that drives expression has to be functional in all organisms and thirdly, the vector needs a selection marker that can complement the auxotrophy in all organisms (Udem & Warner, 1972).

In this study the inner capsid protein (VP2) and middle capsid protein (VP6), making up the double-layered rotavirus particle were expressed using a wide-range yeast expression vector developed at the UFS. This vector is a shuttle vector as it contains both bacterial and yeast elements to facilitate easy plasmid propagation and selection in *Escherichia coli* by the use of a

bacterial origin of replication and the antibiotic resistance gene kanamycin (Albertyn *et al.*, 2011).

The proteins were produced based on the yeast codon-optimized VP2 and VP6 coding sequences of rotavirus strain RVA/HUMAN-WT/ZAF/GR10924/1999/G9P[6] obtained from a child experiencing severe diarrhoea in the Dr. George Mukhari Hospital, University of Limpopo, Medunsa Campus (Jere *et al.*, 2011).

CHAPTER 3

Construction of dual expression vectors containing yeast codon-optimized sequences encoding rotavirus proteins VP2 and VP6

3.1 Introduction

Recombinant gene expression is one of the most essential techniques used in molecular and biomedical research. The use of recombinant proteins as subunit vaccines has been increasing. These vaccines can be developed by production of recombinant proteins that have similar properties to native proteins (Sørensen, 2010). Different systems are available for the production of recombinant genes; these include expression in bacteria, insect, mammalian, plants and yeast cells (Gerngross, 2004). Baculovirus based expression in insect cells is regarded as the main method of production of rotavirus-like particles but has the major disadvantage of being expensive especially to scale-up production (Azevedo *et al.*, 2010).

Yeasts are preferred hosts for expression of recombinant proteins, because they provide ease of genetic manipulation, requires inexpensive medium to grow and its ease to scale up production (Romanos *et al.*, 1992). Yeasts also provide the ability of post-translational modifications which enables production of complex foreign proteins that are identical or very similar to native proteins (Gellissen, 2000; Valenzuela *et al.*, 1982).

All yeast strains have their own advantages and disadvantages that might favour the expression of certain proteins over others. Therefore, a range of yeast strains has to be considered for expression. In order to select a yeast strain that is better suited for expression of a specific protein, multiple yeasts have to be assessed in parallel and a vector that can be targeted to facilitate expression in different systems would minimise workload. Such wide-range expression vectors are becoming available (e.g CoMed^TM vectors (Steinborn *et al.*, 2006)). A vector needs certain properties to be classified as a wide-range expression vector. Firstly, the element targeted by the vector for insertion into the host genome has to be suitable for all the selected organisms. Secondly, the promoter that drives expression has to be functional in all organisms

and thirdly, the vector needs a selection ladder that can complement the auxotrophy in all organisms (Udem & Warner, 1972).

Some transformed proteins are not stable and give a low or no yield when expressed. Therefore, codon optimization is employed to overcome this and to increases protein expression levels by manipulating the nucleotide sequence of the protein product without changing the amino acid sequence. This tool alters sequences based on the preferred codon usage by the expression host organism to favour efficient transcription, mRNA stability and translation (Nguyen *et al.*, 2004; Burgess-Brown *et al.,* 2008). Gene optimization takes advantage of the degeneracy of the genetic code and gene synthesis.

3.2   Materials and Methods

3.2.1 Enzymes, kits, general chemicals and reagents

Restriction endonucleases, T4 DNA ligase, alkaline phosphatase and DNA molecular weight ladders were all supplied by Fermentas. KAPA HiFi DNA polymerase, dNTPs and associated reagents were supplied by KAPA Biosystems. Sequencing reagents were supplied by Applied Biosystems, Life Technologies. NucleoSpin® gel and PCR clean-up kits and NucleoSpin® plasmid extraction kits for DNA extraction and purification were supplied by Macherey-Nagel, Germany (MN). Oligonucleotide primers were synthesized and supplied by IDT-DNA. Primers were analysed using the 'Oligo Analyzer' tool on the Integrated DNA Technologies (IDT) web page (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) using default settings. Other chemicals were, unless otherwise specified, obtained from Fluka, Merck, or Sigma-Aldrich.

3.2.2 Strains of bacteria and cultivation medium

Plasmid manipulations were performed using *Escherichia coli* strain XL10-Gold [F' proAB laclqZDM15 Tn10 (Tet^r) Amy Cam^r; Stratagene].

Cultivation of *E. coli* was achieved using Luria-Bertani (LB) medium containing per litre, 5 g yeast extract, 10 g sodium chloride, and 10 g tryptone; for plating purposes, agar (15 g per litre) was added. LB medium was supplemented with 100 µg/ml kanamycin or ampicillin for selection of colonies successfully transformed with recombinant vectors. Medium used for preparation of *E.*

coli competent cells were Psi broth (2 g tryptone, 0.5 g yeast extract, 0.5 g MgSO$_4$.7H$_2$O per 100ml dH$_2$O, pH 7.6 with potassium hydroxide (KOH)), TBF1 (100 mM rubidium chloride (RbCl$_2$), 50 mM manganese chloride (MgCl$_2$.4H$_2$O), 30 mM potassium acetate (KOAc), 10 mM calcium chloride (CaCl$_2$.2H$_2$O) and 15% w/v glycerol) and TBF2 (10 Mm MOPS, 10 mM RbCl$_2$, 75 mM CaCl$_2$.2H$_2$O and 15% w/v glycerol).

3.2.3 General methods

Standard genetic techniques were used, as described in Green and Sambrook (2012).

3.2.3.1 Polymerase Chain Reaction (PCR)

For general amplification of DNA or single colonies, the KAPA Taq DNA polymerase kit (KAPA Biosystems) was used. Reactions contained 1X KAPA Taq buffer, dNTPs to a final concentration of 300 µM, 0.3 µM of the forward and reverse primers, 1 U KAPA Taq DNA Polymerase and 50 ng of template DNA and filled to 50 µl with distilled water (dH$_2$O). A negative control was included which contained nuclease-free water (NFW) instead of DNA template.

PCRs were subjected to 30 cycles using the G-Storm Thermal System. Initial DNA denaturation at 95°C for 2 minutes and then cycled at 98°C for 30 seconds for denaturation, primer annealing at melting temperature (T$_M$) of the lowest primer minus 5°C for 30 seconds, extension at 72°C for 3 minutes, and final extension at 72°C for 2 minutes followed by holding at 4°C.

3.2.3.2 Agarose gel electrophoresis

General agarose gel electrophoresis was performed using 1% agarose gel containing 1 g agarose in 100ml of 1X Tris Acetic EDTA electrophoresis buffer (40 mM Tris, 2 mM EDTA, 20mM glacial acetic acid; pH 8.5). A final concentration of 0.6 mg/ml ethidium bromide was used to stain the gel for visualization under UV light using a ChemiDoc XRS (Bio-Rad Laboratories) for documentation purposes, or a DarkReader^{TM} transilluminator (Clare Chemicals, United States) for excision of fragments from agarose gels for purification. Target bands were compared to a 10 000 bp GeneRuler DNA Ladder Mix (Thermo Scientific) to determine size of each band. The sizes presented by the 10 000 bp GeneRuler DNA Ladder Mix are given in Fig. 14 in Appendix D.

3.2.3.3 DNA extraction and purification

Plasmid DNA was extracted from the overnight culture (16 hours) using NucleoSpin® plasmid extraction kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. An overnight culture was centrifuged at full speed for 2 minutes and the supernatant was discarded and the pellet re-suspend in 250 µl re-suspension buffer. A volume of 250 µl lysis buffer was added. The mixture was inverted 5 times, 250 µl of neutralization buffer was added and mixture inverted 5 times again. The tubes were centrifuged for 10 minutes at a speed of 12 000 x g and supernatant transferred to spin column. Tubes were centrifuged at 6 000 x g for 1 minute and flow-through discarded. A 650 µl wash buffer containing ethanol was added and tubes centrifuged at 12 000 x g for 1 minute. The flow-through was discarded before centrifuging for additional 1 minute at 12 000 x g and spin column transferred to 1.5ml microcentrifuge tube. A volume of 20-50 µl Elution buffer (pre-heated to 50 °C) added. Tubes were then incubated at room temperature for 1 minute at 50°C for 10 minutes and centrifuged at 12 000 x g for 1 minute.

For purification, target DNA bands were excised and purified using NucleoSpin® gel and PCR clean-up kit according to manufacturer's instructions. Excised gel bands were weighed and two volumes of extraction buffer added and incubated at 50°C with vortexing every 3 minutes until the gel dissolved. The sample was transferred to a spin column and centrifuged at 6 000 x g for 1 minute and 750 µl wash buffer with ethanol was added. The mixture was centrifuged at 12 000 x g for 1 minute. The samples were centrifuged for an additional 1 minute at 12 000 x g and the spin column transferred to a 1.5ml microcentrifuge tube and 15-30 µl elution buffer (pre-heated to 50 °C) added. The tubes were then incubated at room temperature for 1 minute and at 50°C for 10 minutes before centrifuging at 12 000 x g for 1 minute. DNA was stored at -20 °C.

3.2.3.4 Competent cells preparation

Competent *E. coli* cells were prepared and transformed according to the methods described by Inoue *et al.*, (1990). To prepare competent cells, 5 µl of *E. coli* cells (stored in 30% glycerol) were inoculated in 5ml LB medium as a pre-inoculum. A volume of 1ml pre-inoculum was inoculated in 100ml Psi broth and grown to an OD_{600} of 0.4-0.6. The flask containing *E. coli*

culture was placed on ice for 15 minutes and the culture was centrifuged for 5 minutes at 4000 x g at 4 °C. Supernatant was discarded and pellet re-suspended with 40ml of TBF1 and incubated for 15 minutes on ice. The culture was again centrifuged for 5 minutes at 4000 x g at 4°C and the pellet re-suspended with 2ml TBF2. The culture was incubated on ice for 30 minutes and 50 µl aliquots were prepared in microcentrifuge tubes which were snap-frozen and stored at -70°C until use.

3.2.3.5 Cloning and transformation in *E. coli*

Construction of recombinant vectors was achieved by first digesting vectors and insert DNA with restriction endonucleases that will produce cohesive ends to facilitate ligation. Digestions were performed according to manufacturer's instructions which are specific for each restriction enzyme. Generally, 20 µl digestion reactions were prepared containing 500-1 000 ng DNA, buffer, 1 U restriction enzyme and filled with dH_2O. For selection of the best buffer when performing double restriction digests, a web program (http://www.thermoscientificbio.com/webtools/doubledigest/) was used. Digested DNA samples were electrophoresed on a 1% agarose gel and DNA bands excised for purification using NucleoSpin® gel and PCR clean-up kit according to manufacturer's instructions.

Purified DNA was ligated with a vector:insert ratio of 3:1 as calculated using an online program ligation calculator web (http://www.insilico.uni-duesseldorf.de/Lig_Input.html). The standard vector concentration was chosen as 100 ng. Ligation mixture contained vector and insert DNA, T4 ligase enzyme, buffer and dH_2O. The ligation mixture was incubated at 16°C for 18 hours.

A volume of 10 µl ligation mixture was used to transform competent *E. coli* cells using the rubidium chloride ($RbCl_2$) transformation method (Inoue *et al.*, (1990)). Briefly, $RbCl_2$ cells were thawed on ice, 10 µl DNA added into the $RbCl_2$ cells and incubated on ice for 60 minutes, cells were heated at 42°C for 35 seconds and immediately put in ice-water slurry for 2 minutes and then 1ml LB medium was added and incubated at 37°C for 60 minutes shaking. After the incubation, cells were centrifuged for 35 seconds at 1 000 x g and most of the supernatant removed. The remaining cells were re-suspended and plated on LB plates with appropriate antibiotic (either kanamycin or ampicillin) and incubated at 37°C overnight.

3.2.4 Vectors and rotavirus strain

A vector pair, pKM173 and pKM177 (Figure 3.1) were used which, upon fusion allowed a dual insert vector. The one gene of interest (the downstream gene) is cloned into pKM173. The other gene (the upstream gene) is cloned using the normal cloning sites into pKM177. pKM173 is linearized with the homing endonuclease *I-SceI* and pKM177 is then digested with *I-SceI*, liberating the VP6 expression cassette flanked by the *I-SceI* sites. The liberated cassette is then ligated into the linear dephosphorylated pKM173, yielding a dual insert wide-range yeast expression vector (Albertyn *et al.*, 2011).
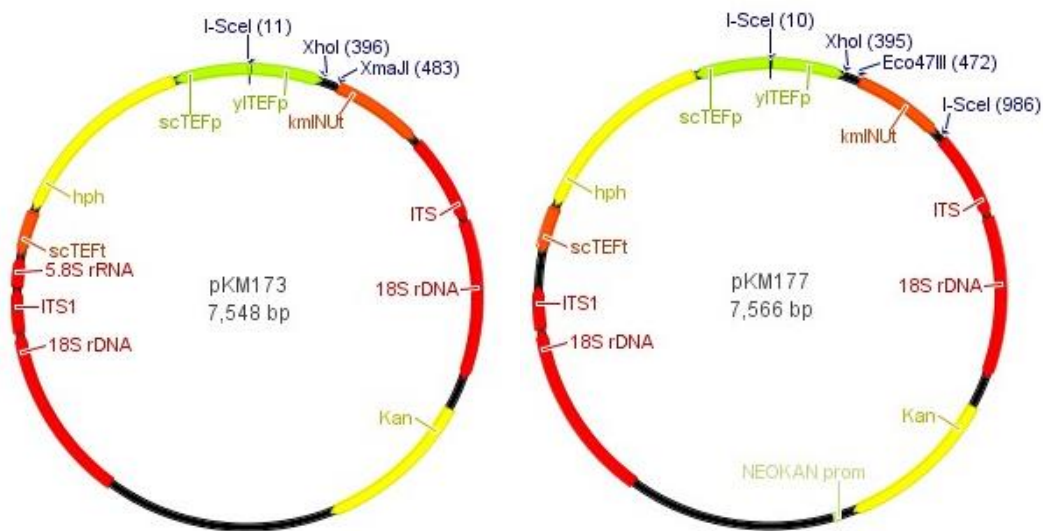
Figure 3.1: Wide-range expression vectors, pKM173 (left) and pKM177 (right) used in this study. Restriction enzymes indicated (*XhoI*, *XmaJI* and *Eco47III*) were used to allow construction of singular expression vectors. Restriction enzyme *I-SceI* was used to further construct dual vectors. scTEFp indicates *S. cerevisiae* TEF promoter and ylTEFp indicates *Y. lipolytica* TEF promoter, hph indicates hygromycin B resistance gene and Kan indicates kanamycin resistance gene, 18S rDNA indicates and ITS RNA sequence of *Kluyveromyces marxianus* and kmINUt indicates *Kluyveromyces marxianus* inulase terminator region.

This wide-range yeast expression system contains the 18S rDNA target sequences of *Kluyveromyces marxianus* to allow genomic integration into the yeast cells. The selective marker is the *hph* gene conferring hygromycin B resistance. The vectors contain the *Yarrowia*

lipolytica TEF promoter and the *Kluyveromyces marxianus* inulinase terminator. The vectors contain the kanamycin resistance gene to facilitate sub-cloning in *E. coli* (Albertyn *et al.*, 2011).

Sequences encoding VP2 and VP6 rotavirus structural proteins used in this study were obtained from rotavirus strain RVA/Human-wt/ZAF/GR10924/1999/G9P[6], with genome constellation G9-P[6]-I2-R2-C2-M2-A2-N2-T2-E2-H2 which was obtained from a stool sample of a neonate experiencing severe diarrhoea at Dr. George Mukhari Hospital, Medunsa Campus, University of Limpopo, South Africa (Jere *et al.*, 2011). The whole genome sequence of this strain was obtained using GS20/FLX technology (Potgieter *et al.*, 2009). Wild-type sequences of this strain were modified (codon optimization) to favour expression in specific yeast strains namely, *Kluyveromyces lactis*, *Arxula adeninivorans* and *Pichia pastoris/Hansenula polymorpha*. *Kluyveromyces lactis* and *Arxula adeninivorans* optimized open reading frames (ORFs) with desired restriction sites were purchased from GenScript and *Pichia pastoris/Hansenula polymorpha* optimized ORFs were provided by Prof. J. Görgens from Stellenbosch University under a material transfer agreement and was previously purchased from GeneArt[TM].

3.2.5 Yeast codon-optimized ORFs

3.2.5.3 *Pichia pastoris* optimized ORFs

Pichia pastoris/Hansenula polymorpha optimized (PO) ORFs were obtained from Prof. J. Görgens (Stellenbosch University) in transformed *E. coli* cells containing the respective recombinant plasmids on LB agar plates containing the appropriate antibiotic (kanamycin for VP6 ORF in pGA15 (kan[R]) and ampicillin for VP2 ORF in pGA4 (amp[R]). An overnight culture was prepared by inoculating colonies in LB broth containing 100 µg/µl of the appropriate antibiotic. Plasmid DNA was extracted using NucleoSpin® plasmid extraction kit as described in section 3.2.3.2.

PCR was performed as described in section 3.2.3.1. For VP2 ORF amplification, primer annealing was carried out at 60°C for 30 seconds, extension at 72°C for 3 minutes, and final extension at 72°C for 2 minutes followed by holding at 4°C. Primer annealing for VP6 ORFs was carried out at 58°C for 30 seconds and extension at 72°C for 1.5 minutes. Primers for amplification of VP2 ORF (VP2 yeast Forward 5'-CTCA<u>CTCGAG</u>ATGGCGTACAG–3' and VP2 yeast Reverse 5'-

GCGT<u>CCTAGG</u>CTACAATTCGT-3') and VP6 ORF (VP6 yeast Forward 5'-CAAC<u>CTCGAG</u>ATGGATGTCCT-3' and VP6 yeast Reverse 5'-GTCC<u>AGCGCT</u>TTATTTGACAA-3') used in this study were designed in a related study by Aliza Naudé (North-West University). Primers introduced appropriate restriction sites (indicated by underlining) at the 5' and 3' ends, respectively, of the ORFs. Forward primers introduced a *XhoI* site at the 5'-end for both VP2 and VP6 ORFs and reverse primers introduced *XmaJI* site at the 3'-end of VP2 ORF and *Eco47III* site at the 3'-end of VP6 ORF, respectively. Primers designated VP2 Forward and VP2 Reverse amplified a 2 631 bp region encoding VP2 and primers designated VP6 Forward and VP6 Reverse amplified a 1 207 bp region encoding VP6.

After amplification, 50 µl amplicons mixed with 7 µl 6X loading dye (Thermo Scientific) were separated by electrophoresis using 1% agarose gel as described in section 3.2.3.1. Target DNA bands were excised, purified using NucleoSpin® gel and PCR clean-up kit according to manufacturer's instructions and DNA stored at -20 °C.

3.2.5.4 *Kluyveromyces lactis and Arxula adeninivorans* optimized ORFs

K. lactis (KO) and *A. adeninivorans* (AO) optimized ORFs, cloned into pUC57 with desired restriction sites to facilitate cloning into the yeast expression vectors, were obtained from GenScript in lyophilised form. Vials were centrifuged at 6 000 x g for 1 minute and 20 µl of 10 mM Tris buffer was added. Vortex briefly and incubated at 50°C for 15 minutes. 1 µl DNA was transformed using rubidium chloride (RbCl$_2$) transformation method as described in section 3.2.3.4.

Single colonies from overnight transformations were inoculated in LB broth containing 100 µg/ml ampicillin and incubated at 37°C overnight. Plasmid DNA was extracted using NucleoSpin® plasmid extraction kit and digested with restriction enzymes to confirm positive colonies containing VP2 and VP6 yeast optimized ORFs. Plasmid DNA of vectors pUC57 containing VP2 were digested with 1 U *XhoI* restriction enzyme (*SalI* for *A. adeninivorans* optimized VP2 ORF) and 2 U *XmaJI* using 2X Tango buffer. Plasmid DNA of vectors pUC57 containing VP6 were digested with 1 U *XhoI* restriction enzyme and 2 U *Eco47III* using 2X Tango buffer. DNA was stored at -20 °C.

3.2.6 Cloning strategy

After purification of the digested DNA of yeast optimized ORFs, wide-range expression vectors containing VP2 and VP6 ORFs were constructed (Figure 3.2). Briefly, 750 ng of pKM173 vector and VP2 ORFs were digested with *XhoI* and *XmaJI* to yield cohesive ends; *Arxula adeninivorans* optimized VP2 ORF was digested with *SalI* instead of *XhoI* because its sequence contains multiple *XhoI* recognition sites. It was possible to use *SalI* (recognition sequence G/TCGAC) enzyme because it has similar recognition sequence to *XhoI* (recognition sequence C/TCGAG) and both enzymes yield compatible cohesive ends so that the overhang produced by one can ligate with the overhang produced by the other. Approximately 750 ng pKM177 vector and VP6 ORFs were digested with *XhoI* and *Eco47III* to yield cohesive ends. Digested vectors, pKM173 and pKM177 were ligated with digested ORFs, VP2 and VP6, respectively using T4 ligase enzyme as described in section 3.2.3.4.

Competent cells were transformed with ligated sequences using rubidium chloride (RbCl$_2$) transformation method as described in section 3.2.3.4. Single colonies were selected from transformed plates and inoculated into 5ml LB medium containing 100 ng/µl kanamycin and incubated overnight at 37 °C. Plasmid DNA was extracted from the overnight culture using NucleoSpin® plasmid extraction kit according to manufacturer's instruction. Recombinant plasmid DNA was digested with cloning restriction enzymes as well as various other restriction enzymes to confirm positive transformants (colonies with pKM173_VP2 and pKM177_VP6).

On completion of the construction of single vectors, *I-SceI* restriction enzyme was chosen for use in construction of dual expression vectors. Concentrations of 750 ng pKM173_VP2 and pKM177_VP6 plasmids were digested with *I-SceI* to yield linearized pKM173_VP2 vector and to liberate the expression cassette flanked by the *I-SceI* sites on pKM177_VP6 (Fig. 3.2). Digested plasmid DNA and the liberated VP6 cassette band were electrophoresed on a 1% agarose gel. The linearized pKM173_VP2 bands were excised and DNA extracted using NucleoSpin® gel and PCR clean-up kit according to manufacturer's instructions. The linear pKM173_VP2 was dephosphorylated using 1U FastAP Alkaline Phosphatase (Thermo Scientific) for 10 minutes at 37°C followed by inactivation for 10 minutes at 75 °C. The VP6 cassette was then ligated into

the linear dephosphorylated pKM173_VP2, yielding a dual insert wide-range yeast expression vector (pKM173_VP2/6). Ligation of dephosphorylated pKM173_VP2 vectors was performed as a negative control to assess the efficiency of dephosphorylation based on the ability of the vectors to self-ligate.

Rubidium chloride (RbCl$_2$) transformation was again used to transform ligated vectors to enable the construction of dual expression vectors. Single colonies picked from transformed plates were inoculated into 5ml LB medium containing 100 ng/µl kanamycin and incubated overnight at 37 °C. Plasmid DNA was extracted from the overnight cultures using NucleoSpin® plasmid extraction kit according to the manufacturer's instruction. Plasmid DNA was digested with the cloning restriction enzyme (*I-SceI*) and confirmed with various restriction enzymes. Extracted DNA from positive colonies was stored at -20 °C.
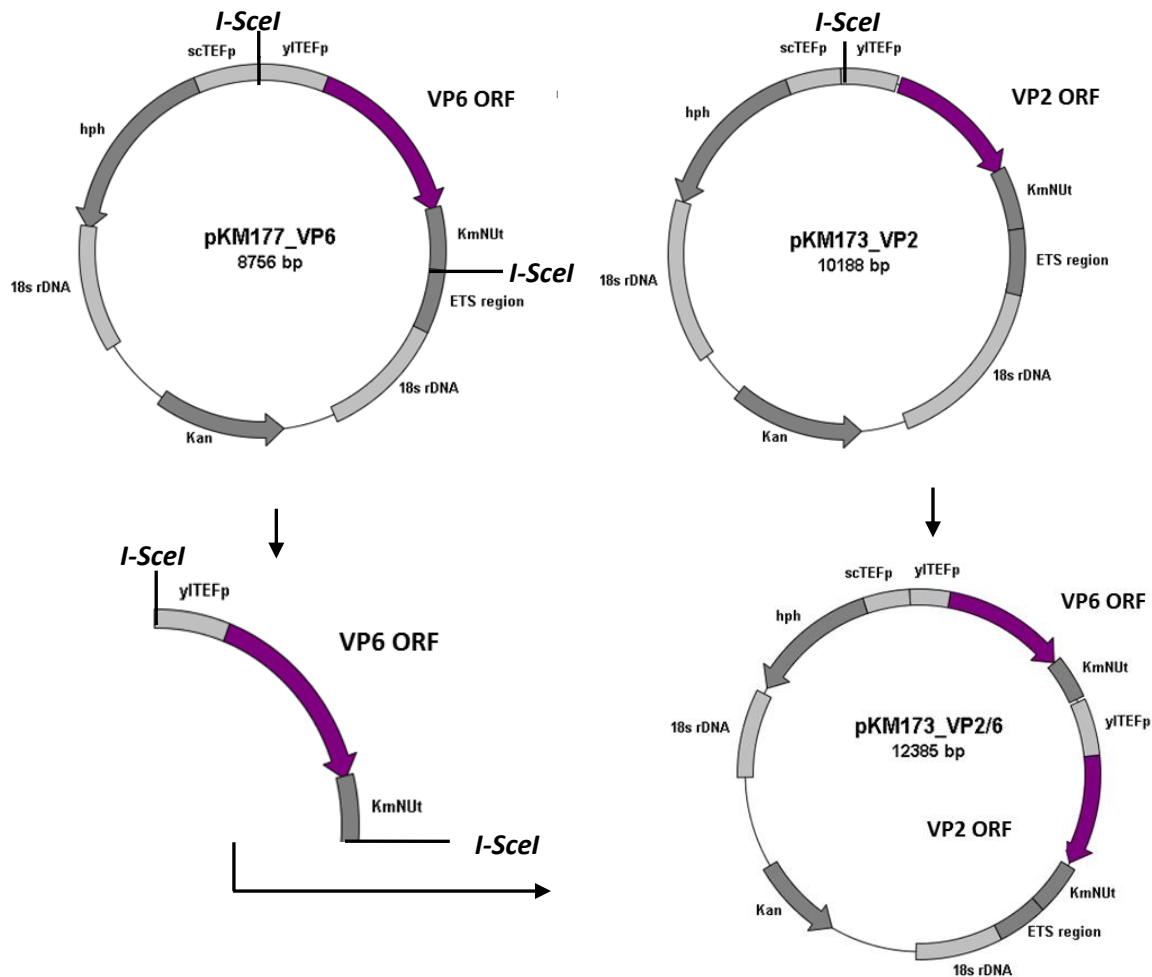
Figure 3.2: An overview of the cloning strategy for construction of dual expression vectors comprising VP2 and VP6 ORFs.

3.2.6.3 DNA Sequencing

After successful construction of single and dual vectors, Sanger sequencing was performed to confirm both successful construction of the recombinant vectors as well as to obtain the nucleotide sequences of cloned ORFs to ensure that no single nucleotide changes occurred during the process of amplification and cloning. DNA was sequenced using the BigDye terminator v.3.1 kit. Table 3.1 show lists of primers that were designed for sequencing. Primer properties were determined using Integrated DNA Technology (IDT) online program.

DNA was sequenced using BigDye^{TM} Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems). Each sequencing reaction contained a final volume of 10 μl with 0.5 μl of premix, sequencing buffer to a final concentration of 1X, 3.2 pmol of oligonucleotide primer, approximately 450 ng of DNA and dH$_2$O to a final volume of 10 μl. DNA was denatured at 96°C for 1 minute. Thermal cycling was performed for 25 cycles at 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 4 minutes, and holding at 4°C.

Sequencing clean-up was done using an ethanol-EDTA precipitation method. Each sequencing reaction was adjusted to 20 μl by adding 10 μl dH$_2$O and transferred to 1.5ml microcentrifuge tube containing 5 μl 125 mM EDTA and 60 μl of 100% ethanol. The reaction was vortexed for 2 seconds, incubated for 15 minutes at room temperature and centrifuged at 4°C for 20 minutes at 14 000 x g. The reaction was completely aspirated and 500 μl of 70% ethanol added before the reaction was centrifuged at 4°C for 10 minutes at 14 000 x g. The reaction was again completely aspirated and dried in concentrator plus speed-vac (Eppendorf) for 5 minutes. The reaction was then sent for electrophoresis at the Department of Biochemistry and Microbiology, University of Free State.

Geneious 6.1.2 (Biomatters) was used for analysis of nucleotide sequences obtained using Sanger sequencing in comparison with *in silico* clones. Only raw sequencing reads with more than 50% quality were used. Sequencing reads obtained using reverse primers were reverse complemented and *de novo* assembled with reads acquired using forward primes. The assembly was then manually edited and aligned with *in silico* clones for comparison. Geneious 6.1.2 was also used to perform *in silico* restriction digests.

Table 3.1: Primers used for the construction of the constructed vectors and their properties.

	Primer name	Nucleotide sequence (5'-3')	Length (bp)	Melting temperature
	pKM173/177_F	GGTATAAAAGACCACCGTCC	20	52.7°C
	pKM173/177_R	GAACAGCTAGAGTGCGTT	18	52.2°C
K. lactis optimized VP2 ORF	KluyVP2 F2	CTGCATCAATTTGTGATGCTAT	22	51.6°C
	KluyVP2 F3	CATTGACAACCGAAAAGTTG	20	50.2°C
	KluyVP2 R1	GTGATTTGAGCGTAATCTCC	20	51°C
A. adeninivorans optimized VP2 ORF	ArxVP2F2	CTGCCTCTATTTGCGATGCTAT	22	55.1°C
	ArxVP2F	GACTGTGGTCCCTAACGATAT	21	53.6°C
	ArxVP2F3	CCTTGACTACCGAGAAGCTC	20	54.7°C
	ArxVP2R1	GTTGAGAACTCCATCAATGACG	22	53.7°C
	ArxVP2R3	GTAATCTGGGCGTAATCTCC	20	52.6°C
P. pastoris and *H. polymorpha* optimized VP2 ORF	VP2_Int_F1	CCGCTTCTATCTGTGACGCTAT	22	56.5°C
	VP2_Int_F2	CAGAGAATCTTTGGTTGCTTG	21	51.9°C
	VP2_Int_F3	GACGTTCCAGAGTTCCAGAC	21	55.4°C
	VP2_Int_R1	AGTTAACAACGTTTCTGTCAG	21	50.7°C
	VP2_Int_R2	GACCGTTTCTGATGTTGTCG	20	53.8°C
K. lactis optimized VP6 ORF	VP6Seq KF1	GGATGAAATGGTTAGAGAATCAC	23	51.3°C
	VP6Seq KF2	CCAAGCAAGATTCGGTAC	18	50.7°C
	VP6Seq KF3	GAGTGTTCACTGTCGCTTC	19	53.4°C
	VP6Seq KR1	CTCAATGAATACAATACATCCATGCTCG	28	56.0°C

	VP6Seq KR2	GTGAAAGATGCAGAGTAAGGGAAG	24	55.4°C
A. *adeninivorans* optimized VP6 ORF	VP6Seq AF1	GGATGAGATGGTCCGAGA	18	53.2°C
	VP6Seq AF2	ACTTATCAGGCCCGATTCG	19	54.5°C
	VP6Seq AF3	TCTATGCTGGTCAAGTAGAGC	21	53.8°C
	VP6Seq AR1	CTTATCTCGGGCATCCTTCAG	21	54.9°C
	VP6Seq AR2	CGGTTAAGAGTAAACGAGGC	20	53.1°C
P. pastoris and *H. polymorpha* optimized VP6 ORF	VP6Seq PF1	GGACGAAATGGTTAGAGAATCTC	23	53.0°C
	VP6Seq PF2	CAACACCTACCAAGCTAGA	19	51.3°C
	VP6Seq PF3	GAGTTTTCACCGTTGCTTC	19	51.9°C
	VP6Seq PR1	GACAAAGAGTACAAAACGTCC	21	51.5°C
	VP6Seq PR2	CTGTTCAAGGTGAAAGAAGC	20	51.6°C

3.3 Results and Discussion

3.3.1 Cloning Approach

Production of rotavirus-like particles is one of the strategies in development of rotavirus vaccines. Rotavirus-like particles are produced by heterologous expression of rotavirus structural proteins VP2, VP6 and/or VP7/4 in a heterologous system. Yeast expression system has been identified as cost-effective for production of rotavirus-like particles (Rodriguez-Limas *et al*., 2011). To investigate which yeast strain best express rotavirus proteins, we need to express rotavirus proteins in a range of different yeast strains. In this study, we used wide-range yeast expression vectors previously developed at UFS (Albertyn *et al*., 2011) to minimise workload involved in expressing in different yeast strains. It is also important that rotavirus proteins are co-expressed in the same cell for successful formation of rotavirus-like particles. Therefore, a dual expression vector is required for simultaneous expression of proteins.

In this study, we used the genome sequence of a rotavirus field strain RVA/Human-wt/ZAF/GR10924/1999/G9P[6] which was fully characterized directly from stool sample (Jere *et al.*, 2011), thus eliminating possibilities of base changes during adaptation to cell culture. The VP2 and VP6 genome segment sequences to be used in this study were optimized for expression in different yeast strains to increase yield of expression and stability in yeast.

The flow diagram below (Figure 3.3) illustrates the approach to successful construction of dual expression vectors containing yeast codon-optimized rotavirus VP2 and VP6 ORFs.
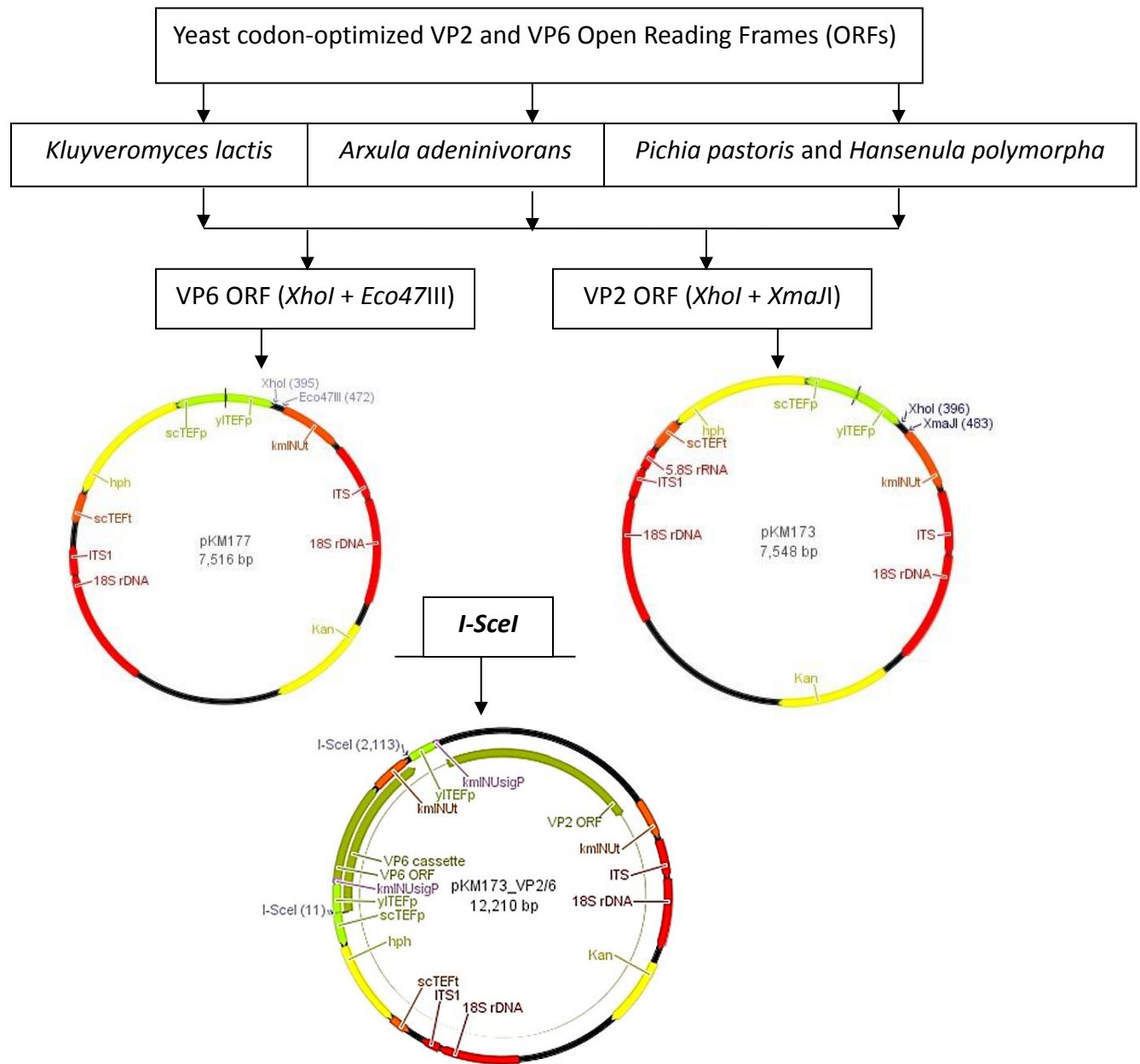
Figure 3.3: Cloning strategy for successful construction of dual vectors. Sequences of VP2 and VP6 open reading frames of rotavirus strain GR10924 were optimized for yeasts *K. lactis*, *A. adeninivorans* and *P. pastoris/H. polymorpha*. Optimized VP2 ORFs were ligated into pKM173 using *XhoI* and *XmaJI* and VP6 ORFs ligated with pKM177 using *XhoI* and *Eco47III*. The pKM177 vectors containing VP6 ORFs were digested with *I-SceI* to liberate the VP6 ORF containing cassettes and pKM173 vectors containing VP2 ORFs were linearized with *I-SceI* and the VP6 ORF containing cassettes were ligated with linearized pKM173 vectors containing VP2 to yield dual vectors containing both VP2 and VP6 ORFs.

3.3.3 Yeast codon-optimized ORFs

3.3.3.3 *Pichia pastoris* optimized ORFs

P. pastoris/H. polymorpha optimized ORFs encoding VP2 and VP6 were successfully amplified using KAPA HiFi DNA polymerase (Figure 3.4). The agarose gel revealed intense DNA bands at the expected sizes of 2548 bp for VP2 ORFs and 1207 bp for VP6 ORFs. There were no non-specific amplification of non-desired amplicons and no band was visible in the negative control which indicates that no contamination was present. The primers used for amplification, inserted restriction recognition sites to allow insertion of these ORFs in appropriate vectors. The bands were excised and purified using NucleoSpin® plasmid extraction kit.
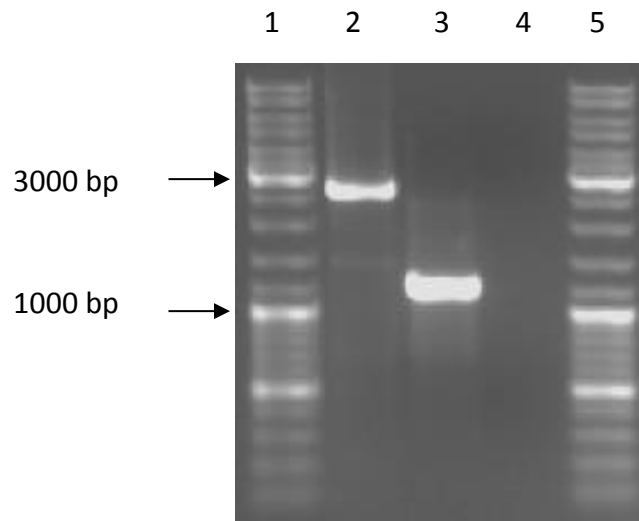
Figure 3.4: PCR amplification of *P. pastoris/H. polymorpha* optimized ORFs. Lanes 1 and 5 represent GeneRuler DNA Ladder Mix (0.5 µg) ladder, lane 2 VP2 ORF, lane 3 VP6 ORF and lane 4 negative control.

3.3.3.4 *Kluyveromyces lactis* and *Arxula adeninivorans* optimized ORFs

Kluyveromyces lactis (KO) and *Arxula adeninivorans* (AO) optimized ORFs were successfully recovered from pUC57 plasmids. This was done through transformation of plasmids and selecting single colonies. Plasmid DNA from transformed colonies was digested with enzymes that cut the plasmids twice and liberate VP2 (*XhoI/SalI* and *XmaJI*) and VP6 (*XhoI* and *Eco47III*) ORFs respectively (Figures 3.5 and 3.6). An *in silico* digest was performed using Geneious 6.1.2

(Biomatters) for comparison of bands observed. Successful recovery of the ORFs allows continuation to the next step which was to clone these ORFs into the wide-range expression vectors.
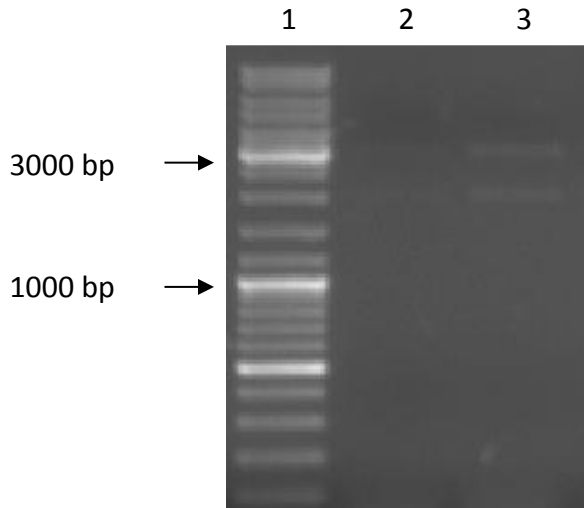
Figure 3.5: **Digestion of pUC57 plasmid containing KO and AO VP2 ORFs. (A) In lane 1 is GeneRuler DNA Ladder Mix, lane 2 VP2 *XhoI* and *XmaJI* digested pUC57_VP2 KO and lane 3 VP2 *SalI* and *XmaJI* digested pUC57_VP2 AO. (B) *In silico* digest of pUC57_VP2 with *XhoI* and *XmaJI*.**
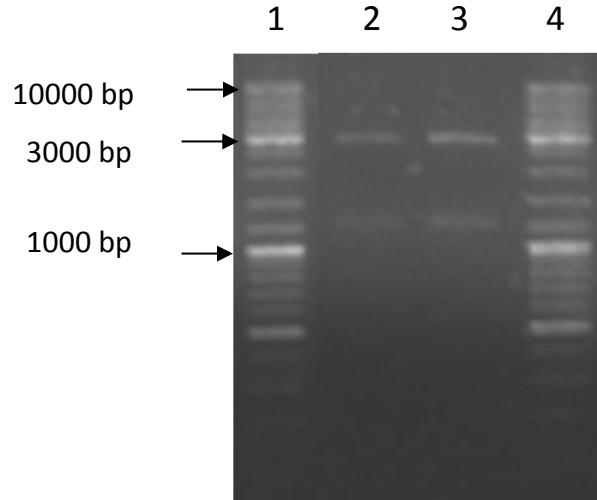
Figure 3.6: **Digestion of pUC57 plasmid containing KO and AO optimized VP6 ORFs. (A) Lanes 1 and 4 represent GeneRuler DNA Ladder Mix. Lane 2 *XhoI* and *Eco47III* digested pUC57_VP6 *K. lactis* optimized and lane 3 *XhoI* and *Eco47III* digested pUC57_VP6 AO.**

3.3.4 Construction of expression vectors containing yeast optimized rotavirus VP2 and VP6 ORFs

Wide-range expression vectors pKM173 and pKM177 along with different yeast codon-optimized VP2 and VP6 ORFs were digested with restriction enzymes. pKM173 and VP2 ORFs were digested with *XhoI/SalI* and *XmaJI*, pKM177 and VP6 ORFs were digested with *XhoI* and *Eco47III.* Digestion of vectors and ORFs DNA were performed in duplicate to increase the yield of recovered after bands were excised and purified using the NucleoSpin® plasmid extraction kit. Purified DNA was electrophoresed in order to analyse DNA recovery (Figure 3.7).
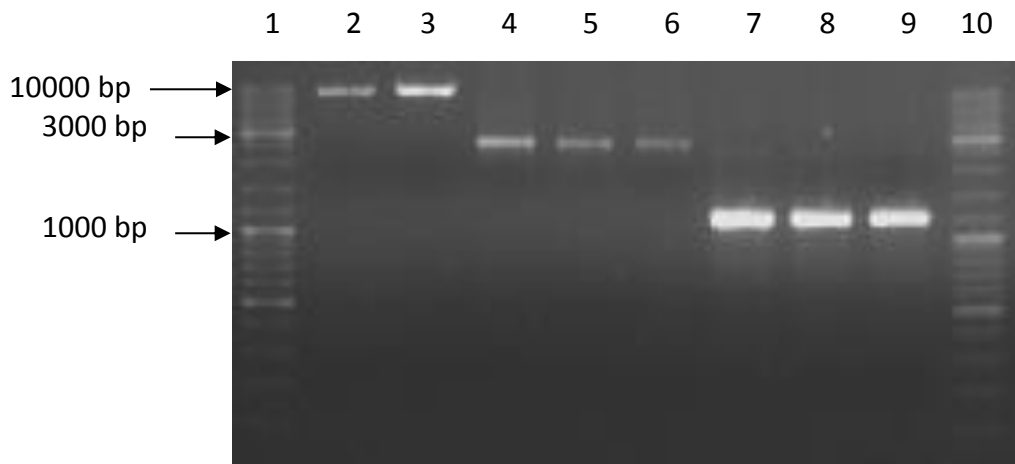
Figure 3.7: Digestion of pKM173 and pKM177 vectors along with VP2 and VP6 ORFs with cloning restriction enzymes. Lanes 1 and 10 is GeneRuler DNA Ladder Mix , lane 2 digested pKM173, lane 3 digested pKM177, lane 4 VP2 PO, lane 5 VP2 KO, lane 6 VP2 AO, lane 7 VP6 PO, lane 8 VP6 KO, and lane 9 VP6 AO.

Ligation reactions were prepared to insert VP2 ORFs into pKM173 vector and VP6 ORFs into pKM177 vector. These reactions were transformed and positively transformed colonies were obtained as confirmed by restriction analysis using cloning enzymes (*XhoI/SalI* and *XmaJI* for pKM173 and VP2 ORFs and *XhoI* and *Eco47III* for pKM177 and VP6 ORFs) (Figure 3.7). Digestion of recombinant vectors with cloning enzymes was expected to liberate VP2 ORFs from pKM173 vectors and VP6 ORFs from pKM177 vectors as also indicated by the *in silico* analysis. Recombinant vectors were also digested with other enzymes for confirmation of insertion and compared to *in silico* prepared digests (Figures 3.8, 3.9 and 3.10). Figure 3.8 presents digestion

of recombinant vectors containing VP2 ORFs. Vectors containing KO VP2 ORFs were digested with *BglII*. Those containing AO VP2 ORFs were digested with *PstI* and vectors containing PO VP2 ORFs were digested with *BstXI*. Figure 3.9 indicates digestion of recombinant vectors containing VP6 ORFs. Vectors containing KO VP6 ORFs were digested with *Bsp119I*, vectors containing AO VP6 ORFs were digested with *BstXI* and vectors containing PO VP ORFs were digested with *Bsp119I*. Digestion yielded band sizes similar to the prediction. Five bands were observed for AO pKM173_VP2 vectors digested with *PstI*. Bands close to one another look like one band or undigested but are two because if it was undigested the bands will appear over 10 000 bp in size and the sizes on the marker support those bands predicted to be there. Four bands were observed for KO pKM173_VP2 vectors digested with *BglII* on the gel and on the prediction. Three bands were observed for PO pKM173_VP2 digested *BstXI*.
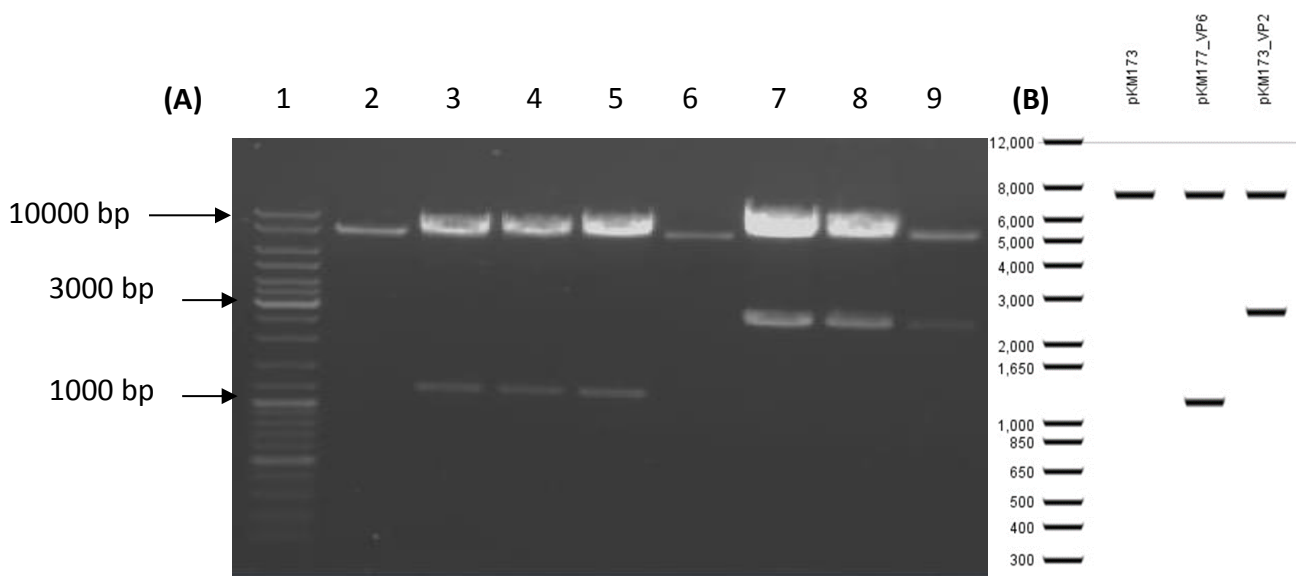
Figure 3.8: Digestion of ligated pKM173_VP2 and pKM177_VP6 to confirm successful construction of singular expression vectors. (A) Lanes 2-5 digested with *XhoI* and *Eco47III*) and lanes 6-9 digested with *XhoI* and *XmaJI*.   Lane 1 is GeneRuler DNA Ladder Mix, lane 2 pKM177, lane 3 pKM177_VP6 KO, lane 4 pKM177_VP6 AO, lane 5 pKM177_VP6 PO, lane 6 pKM173, lane 7 pKM173_VP2 KO, lane 8 pKM173_VP2 AO, and lane 9 pKM173_VP2 PO. (B) *In silico* digest of pKM173_VP2 and pKM177_VP6 with *XhoI*, *XmaJI* and *Eco47III*.
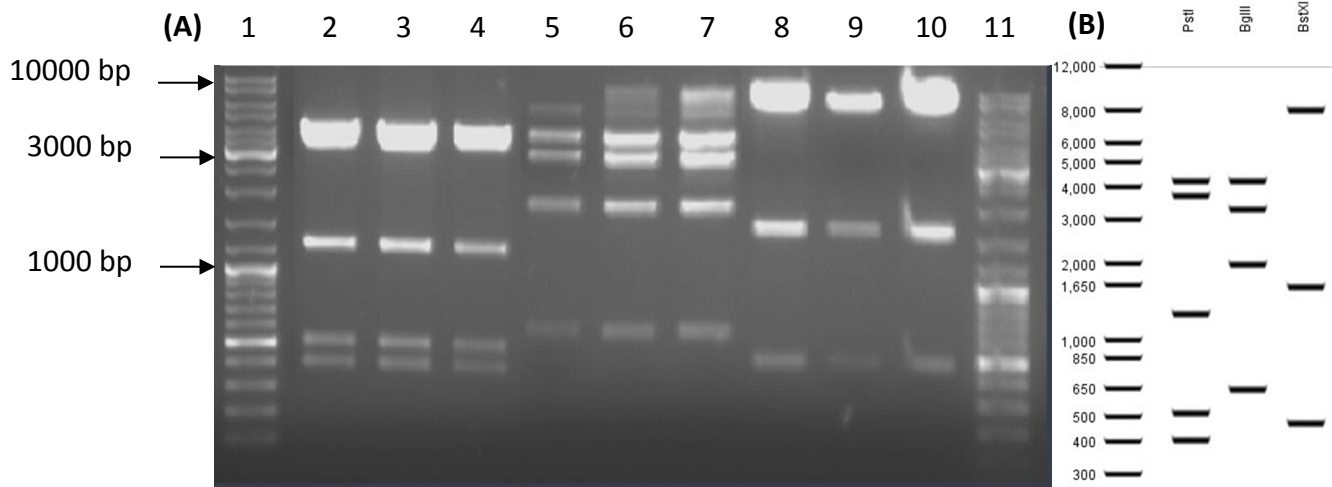
Figure 3.9: Digestion of recombinant vectors containing VP2 ORFs. (A) Lanes 1 and 11 indicate GeneRuler DNA Ladder Mix, lanes 2-4 *PstI* digested AO pKM173_VP2 vectors, lanes 5-7 *BglII* digested KO pKM173_VP2 vectors and lanes 8-10 *BstXI* digested PO pKM173_VP2. (B) *In silico* digest of pKM173_VP2 with *PstI, BglII* and *BstXI.*
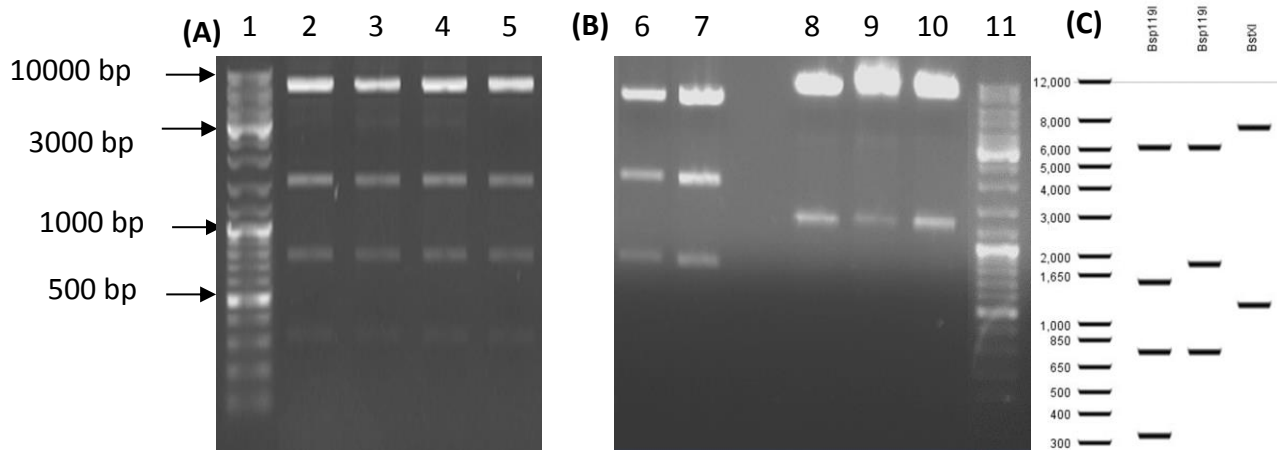
Figure 3.10: Digestion of recombinant vectors containing VP6 ORFs. (A and B) Lane 1 and 11 indicate GeneRuler DNA Ladder Mix, lanes 2-5 *Bsp119I* digested PO pKM177_VP6 vectors, lanes 6 and 7 *Bsp119I* digested KO pKM177_VP6 vectors and lanes 8-10 *BstXI* digested AO pKM177_VP6 vectors. (C) *In silico* digest of pKM177_VP6 with *Bsp119I* and *BstXI.*

3.3.5 Construction of dual expression vectors containing both VP2 and VP6 ORFs in single vectors

Expression vectors constructed to contain either yeast codon-optimized VP2 and VP6 ORFs were used in construction of dual vectors comprising of both VP2 and VP6 ORFs to allow simultaneous expression of proteins encoded by these ORFs. To achieve this, pKM173_VP2 and pKM177_VP6 vectors were digested with a restriction enzyme *I-SceI* to yield linearized pKM173_VP2 and liberating the VP6 expression cassette flanked by the *I-SceI* sites on pKM177_VP6. Dephosphorylation of pKM173_VP2 using Fast AP was efficient as no colonies grew on the self-ligation transformation. The expression cassette was successfully cloned into the linear dephosphorylated pKM173_VP2 (Figure 3.11 and 3.12). Vectors were digested with *I-SceI* enzyme to liberate the VP6 ORF expression cassette (2 103 bp).
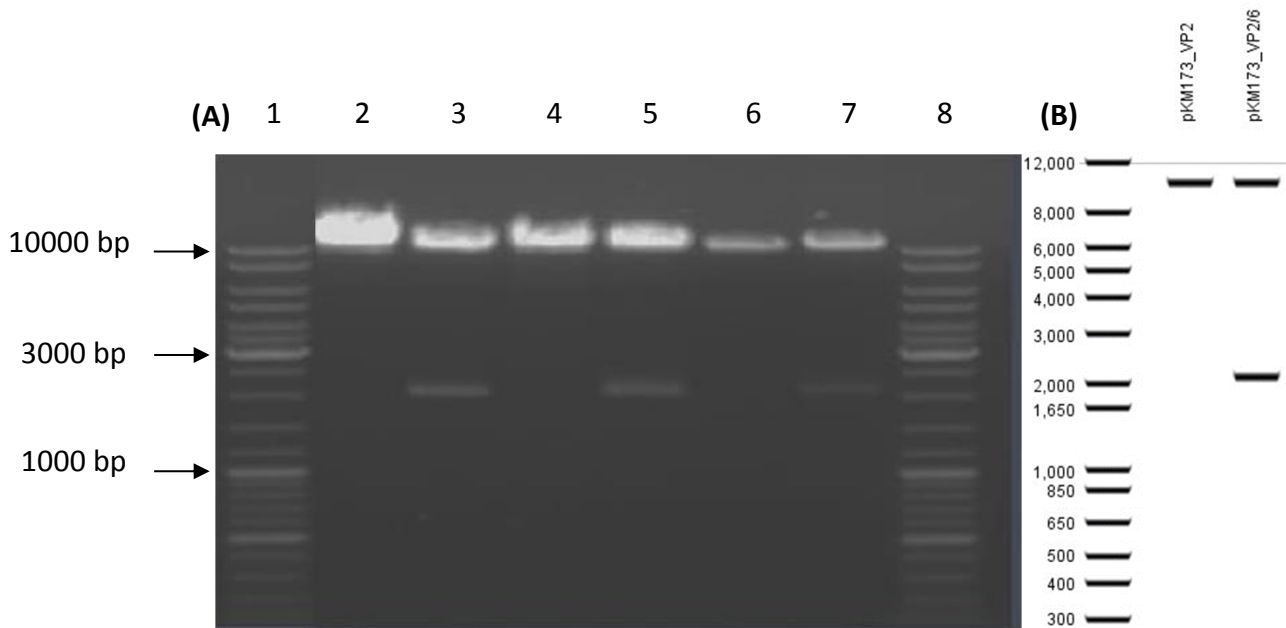
Figure 3.11: Digestion of ligated pKM173_VP2/6 with *I-SceI* to confirm successful construction of dual expression vector. (A) Lanes 1 and 8 is GeneRuler DNA Ladder Mix ladder, lane 2 pKM173_VP2 KO, lane 3 pKM173_VP2/6 KO, lane 4 pKM173_VP2 AO, lane 5 pKM173_VP2/6 AO, lane 6 pKM173_VP2 PO and lane 7 pKM173_VP26 PO. (B) *In silico* digest of pKM173_VP2/6 and pKM173_VP2 with *I-SceI*.
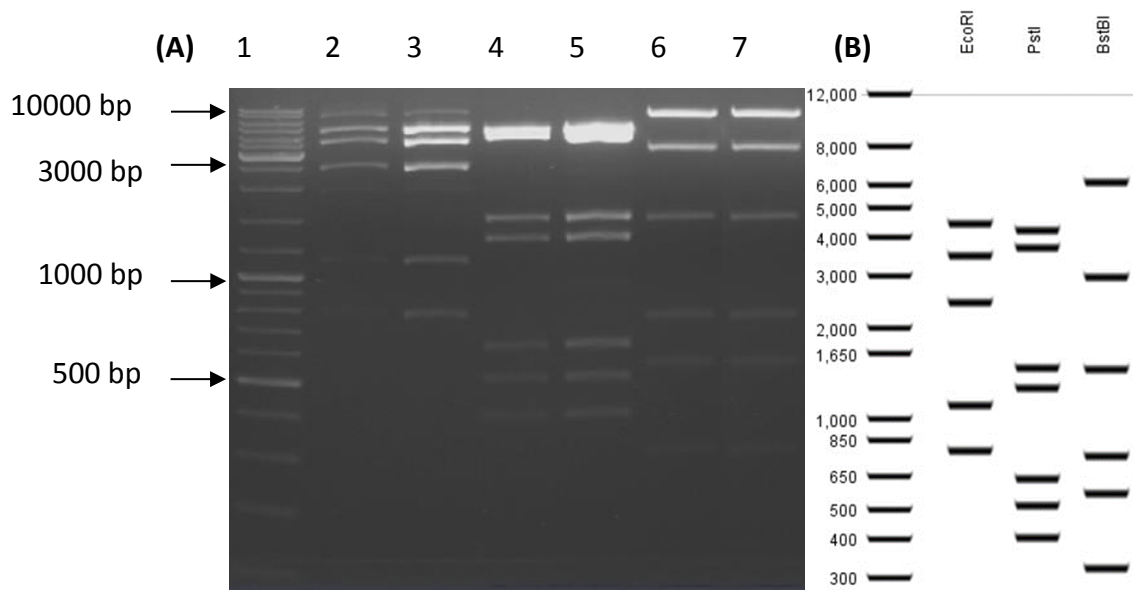
Figure 3.12: Digestion of ligated pKM173_VP2/6 to confirm successful construction of dual expression vector. (A) Lane 1 indicates GeneRuler DNA Ladder Mix, lanes 2 and 3 *EcoRI* digested KO pKM173_VP2/6, lanes 4 and 5 *PstI* digested AO pKM173_VP2/6 and lanes 6 and 7 *Bsp119I* digested AO pKM173_VP2/6. (B) *In silico* digest of pKM173_VP2/6 with *EcoRI, PstI* and Bsp119I.

Geneious 6.1.2 was used for analysis of nucleotide sequences obtained using Sanger sequencing in comparison with *in silico* clones. The *in silico* sequences were obtained from GenScript for KO and Ao and GeneArt for PO.

A total of three clones for each yeast codon-optimized ORF were obtained but only results from one clone are indicated in this report to represent other identical clones. Nucleotide sequences obtained through sequencing were aligned to the *in silico* sequences using Clustal W. Analysis of nucleotide sequences using Geneious 6.1.2 program indicated that no single nucleotide substitutions have occurred during amplification of ORFs and construction of recombinant vectors (Appendix A, Figures 1-9).

3.4   Summary

Recombinant DNA technology allows scientists to express foreign proteins in expression systems. To achieve this, expression vectors should have appropriate selection markers and have promoters to drive the expression in the appropriate host. In this study, we used wide-range yeast expression vectors to construct recombinant vectors containing the open reading frames encoding rotavirus structural proteins VP2 and VP6 (Table 3.2). These vectors will allow for the evaluation of expression of these proteins in different yeast strains in parallel. Dual vectors will also allow for simultaneous expression of VP2 and VP6 proteins to enable the possible formation of double-layered rotavirus-like particles.

Table 3.2: A summary of recombinant vectors constructed using three different yeast optimized VP2 and VP6 ORFs.

Vector	Insert gene	Final product	Yeast optimized for
pKM177	VP6 ORF	pKM177+ VP6	*A. adeninivorans*
pKM177	VP6 ORF	pKM177+ VP6	*K. lactis*
pKM177	VP6 ORF	pKM177+ VP6	*P. pastoris/H. polymorpha*
pKM173	VP2 ORF	pKM173+ VP2	*A. adeninivorans*
pKM173	VP2 ORF	pKM173+ VP2	*K. lactis*
pKM173	VP2 ORF	pKM173+ VP2	*P. pastoris/H. polymorpha*
pKM173	VP2/6 ORF	pKM173+ VP2/6	*A. adeninivorans*
pKM173	VP2/6 ORF	pKM173+ VP2/6	*K. lactis*
pKM173	VP2/6 ORF	pKM173+ VP2/6	*P. pastoris/H. polymorpha*

The successful production of recombinant expression vectors described in this chapter enabled us to further evaluate whether the recombinant wide expression vectors constructed in this study can be used to engineer recombinant yeast strains.

CHAPTER 4

Engineering recombinant yeast strains and expression of rotavirus VP6 protein in yeast

4.1 Introduction

The wide-range yeast expression vectors containing yeast optimized VP6 ORF and dual plasmids containing yeast optimized VP2 and VP6 ORFs were constructed as described in chapter 3. The wide-range yeast expression system contains the 18S rDNA target sequences of *Kluyveromyces marxianus* to allow genomic integration into the yeast cells. The selective marker is the *hph* gene conferring hygromycin B resistance. The vectors contain the *Yarrowia lipolytica* TEF promoter and the *Kluyveromyces marxianus* inulinase terminator (Albertyn *et al.*, 2011).

Codon optimization is a tool that increases protein expression level by manipulating the nucleotide sequence of the protein product without changing the amino acid sequence. This tool alters sequences based on codon usage preferred by the expression host organism to favour efficient transcription, mRNA stability and translation. Gene optimization takes advantage of degeneracy of the genetic code and gene synthesis. Codon-optimized human short-chain dehydrogenase/reductase genes (SDRs) were shown to express and be purified more readily than the native genes in *E. coli* BL21 strain (Burgess-Brown *et al.,* 2008). Codon optimization also made it possible to independently express two HIV-1 accessory proteins (Vpu and Vif) without viral Tat and Rev proteins. This was similar to the finding that optimization of the vpu gene stabilizes the *vphu* mRNA in the nucleus and enhances its export to the cytoplasm (Nguyen *et al.*, 2004).

Rotavirus structural proteins (VP2, VP6 and VP7/4) have been successfully expressed in insect cells (Crawford *et al.*, 1994). This was followed by production of VP2/6 VLPs by a single (genome segments VP2 and VP6 from rotavirus simian SA11) or dual (genome segment VP2 from bovine RF strain and genome segment VP6 from simian SA11 strain) baculovirus expression vectors (Bertolotti-Ciarlet *et al.*, 2003). Chimaeric RLPs were produced from the consensus sequences of African rotaviruses (G2, G8, G9 or G12 strains associated with either P[4], P[6] or P[8]

genotypes). Here sequences encoding rotavirus proteins VP2, VP4, VP6 and VP7 were codon-optimized for expression in insect cells (Jere *et al.*, 2014).

Triple-layered rotavirus virus-like particles (tlRLPs) have been successfully produced in yeast (*Saccharomyces cerevisiae*) (Rodríguez-Limas *et al.*; 2011). The approach followed for expression of the three proteins was based on the cloning of the coding regions of the three proteins into three individual plasmids for expression in yeast. This approach was, however, shown to result in inadequate co-expression. Cloning all three ORFs into one plasmid resulted in sufficient co-expression (Rodríguez-Limas *et al.*; 2011). *Saccharomyces cerevisiae* extracts containing rotavirus-like particles (RLP) were used as a vaccine candidate in an adult mouse model. Two doses of 1 mg of yeast extract containing rotavirus proteins resulted in an immunological response capable of reducing rotavirus replication after infection (Rodríguez-limas *et al.,* 2014). These results support the use of RLPs as alternative rotavirus vaccine and also indicated that yeast expression can be successfully used for production of RLPs.

4.2 Materials and Methods

4.2.1 Enzymes, kits, general chemicals and reagents

Hygromycin B was supplied by Calbiochem, while G418/Geneticin® and kanamycin sulphate were supplied by Sigma-Aldrich. Monoclonal VP6 group-specific antibody and secondary goat anti-mouse IgG antibody were supplied by Santa Cruz Biotechnology and polyclonal NCDV antibody and donkey anti-goat secondary antibody were from Abcam. Other chemicals were, unless otherwise specified, obtained from Fluka, Merck, or Sigma-Aldrich.

4.2.2 Yeasts strains and mammalian cells and cultivation medium

Bacterial gene expressions were performed using *E. coli*, strain BL21 (DE3) [F–, ompT, hsdSB (rB–, mB–), dcm, gal, λ(DE3); Stratagene].

All yeast strains (Table 4.1) were obtained from storage under liquid nitrogen in the MIRCEN yeast culture collection at the Department of Microbial, Biochemical and Food Biotechnology, UFS, South Africa. For short-term storage the cultures were frozen in YPD broth containing glycerol (7% v/v final concentration). The yeasts were cultured on yeast extract, peptone,

dextrose (YPD) (10 g yeast extract, 20 g peptone and 20 g glucose per litre in dH$_2$O) agar (15 g per litre) plates.

Table 4.1: Yeast strains for transformation

Yeast strain	Reference	Hygromycin B concentration
Kluyveromyces marxinaus	Isolate 2.1	600mg/l
Kluyveromyces lactis	UOFS Y-1167	600mg/l
Arxula adeninivorans	UOFS Y-1220	600mg/l
Candida deformans	UOFS Y-2356	600mg/l
Debaromyces hansenii	UOFS Y-0610	800mg/l
Hansenula polymorpha	UOFS Y-0915	800mg/l
Saccharomyces cerevisiae	CEN PK42	600mg/l
Yarrowia lipolytica	Polf	600mg/l
Pichia pastoris	UOFS Y-1167	800mg/l

Mammalian COS-7 cells were obtained from Prof. Sharon Prince, Department of Human Biology, Faculty of Health Sciences, University of Cape Town. Cells were revived in DMEM+Glutamax^{TM}-1 Dulbecco's modified eagle medium (DMEM) (Sigma Aldrich) supplemented with 1% non-essential amino acids (NEAA) (Lonza), 20% fetal bovine serum (FBS) superior (BiochromAG, Berlin) in 1% Penicillin-Streptomycin-Amphotericin B mixture (Lonza) in a 25 cm^3 flask (Sigma Aldrich). Cells were maintained in DMEM with 1% non-essential amino acids (NEAA) from Lonza and 5% fetal bovine serum (FBS) superior (BiochromAG, Berlin) in 1% Penicillin Streptomycin (Pen/Strep) in a 75 cm^3 flask.

COS-7 cells were detached from the flask using 1ml Trypsin/EDTA (Lonza) per 75 cm^3 flask for five minutes in 37 °C.

4.2.3 General methods

Standard genetic techniques were used, as described in Green and Sambrook (2012).

4.2.3.1 Competent cells preparation

Competent *E. coli* cells were prepared and transformed according to the methods described by Inoue and colleagues (1990) as described in chapter 3 (section 3.2.3.3).

Competent cells for yeast transformation were prepared using the method described by Klebe and colleagues (1983). Pre-inoculum was prepared by inoculating colonies of revived yeast strains in 5ml YPD medium and incubated at 30°C shaking for 24 hours. A volume of 1ml pre-inoculum was inoculated in 100ml YPD and grown to an OD_{600} of 0.6-0.8. Cells were centrifuged at 1 000 x g for 5 minutes and washed once with 50 ml solution 1 (1 M sorbitol, 10 mM bicine-NaOH (pH 8.35), 3% ethylene glycol and 5% DMSO). Cells were then resuspended in 2 ml of solution 1 and 200 µl aliquots were prepared and stored at -70°C until use.

4.2.3.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis

Expression of rotavirus VP6 protein was examined using SDS-PAGE and western blot analysis. Proteins were analysed using a 12% SDS-PAGE resolving gel (per 10 ml, 2.2 ml dH_2O, 4 ml 30% polyacrylamide gel, 2.5 ml 1.5 M Tris (pH 8.8), 100 µl 10% SDS, 100 µl 10% ammonium persulfate (APS) and 4 µl TEMED). A glass plate sandwich was assembled and the resolving gel added leaving approximately 1/4 of the space free which was covered with isopropanol to level the resolving gel and waited until the resolving gel polymerized (30 minutes). The isopropanol was discarded and stacking gel (per 3 ml, 1.72 ml dH_2O, 0.67 30% polyacrylamide gel, 1 ml 1.5 M Tris (pH 6.8), 30 µl 10% SDS, 30 µl 10% APS and 3 µl TEMED) was poured over the resolving gel and combs inserted and gel allowed to set for 1 hour before loading samples.

Samples were mixed with sample buffer (50 µl Beta-mercaptoethanol and 950 µl Laemmli sample buffer (Bio-Rad)) in a 1:1 ratio then heated at 95°C for 10 minutes for final denaturation. The gel was placed into the electrophoresis tank, and the tank filled with fresh 1 X Tris Glycine SDS (TGS) buffer. Combs were carefully removed from gel and samples loaded

together with molecular weight marker. The polyacrylamide gel was electrophoresed for 2 hours at 90 volts (V).

For western blot analysis, SDS-PAGE was performed as described above, proteins were transferred to a BioTrace™ NT Nitrocellulose Transfer Membrane (Pall Corporation) for 1 hour at 100 V in a transfer buffer containing 0.025 M Tris, 0.2 M glycine and 20% methanol (pH 8.4). Protein transfer was verified by staining nitrocellulose membrane with Ponceau S (Sigma) solution (0.1% and 1% acetic acid) for 10 minutes. The membrane was blocked using 5% skimmed milk (Nestle) in 1 X Tris-buffered saline with 1% Tween-20™ (TBST; CIS Bio) for 1 hour at room temperature with shaking to prevent nonspecific antibody binding. The membrane was washed twice with TBST and incubated at 37°C for 2 hours with primary antibody (Rotavirus group-specific antigen Antibody (3C10), mouse monoclonal IgG$_{2a}$ (Santa Cruz)) diluted 1:500 in TBST). The membrane was washed twice with TBST and incubated at 37°C for 1 hour with 1:2 000 secondary antibody (goat-anti mouse antibody IgG-AP : sc-2008 (Santa Cruz)). Detection was implemented by adding 10 ml of (5-bromo-4-chloro-3-indolyl phosphate, p-toluidine salt) (BCIP-T) (Thermo Scientific)/ Nitroblue Tetrazolium (NBT) (Sigma-Aldrich) alkaline substrate (33 µl 50mg/ml BCIP-T and 44 µl 75mg/ml NBT to 10ml alkaline phosphatase buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl and 10 mM MgCl$_2$)) and incubated at room temperature until bands were visible.

A different set of antibodies were also used for western blot analysis since protein detection with the group-specific antigen was problematic. After the membrane was blocked using TBST with 5% skimmed milk for 1 hour at room temperature, it was incubated with the primary (Anti-Rotavirus NCDV antibody (Biotin) (Abcam)) polyclonal antibody raised against the Nebraska calf diarrhoea virus (NCDV) strain at a 1:200 dilution for 3 hours followed by washing three times with TBST. The membrane was incubated with a donkey anti-goat secondary antibody (Abcam) which was diluted to 1:500 for 1 hour with shaking. To develop protein bands on the nitrocellulose membrane, a 4-chloro-1-naphthol peroxidase substrate tablet (Sigma) was dissolved in 10 ml ice-cold methanol followed by adding 2 ml of the dissolved substrate to 10 ml PBS (pH 7.4). Five microlitres of hydrogen peroxide (Sigma) were added to the mixture and the membrane development solution was applied onto the nitrocellulose membrane for 5

minutes or until protein bands had developed sufficiently. To stop further development, the membrane was transferred into deionised water.

4.2.4   Engineering of recombinant yeast strains

4.2.4.1 Transformation of yeast strains

A total of three pKM177 vectors containing three different yeasts codon-optimized VP6 ORFs and pKM173 dual vectors also containing three different yeast codon-optimized VP2 and VP6 ORFs that were constructed as discussed in chapter 3 were used to transform nine yeast strains (Table 4.1).

Transformation in all yeast strains except *K. marxianus* was carried out using a method described by Chen and colleagues 1997. Approximately 1 µg of plasmid DNA was linearized with *Not*I restriction endonuclease and added to the yeast competent cells along with 50 µg of salmon sperm carrier DNA (Life Technologies). A volume of 1.4 ml solution 2 (40% w/v PEG-1000; 200 mM bicine-NaOH, pH 8.4) was added and the mixture was vortexed for 1 minute before incubating at $30^{o}C$ for 1 hour. Cells were heat shocked at $37^{o}C$ for 10 minutes and harvested at 2 000 x g for 5 minutes. The pellet was resuspended in 500 µl solution 3 (150 mM NaCl; 10 mM bicine-NaOH, pH 8.4) and 500 µl YPD, followed by recovery at $30^{o}C$ for 2 hours. Cells were again harvested at 2000 x g for 5 minutes. Most of the supernatant (approximately 800 µl) was poured off and the pellet resuspended in the remaining supernatant and plated on selective YPD plates containing hygromycin B until colonies formed. The concentrations of hygromycin B for plating each yeast strains are indicated in Table 4.1.

K. marxianus was transformed using the one-step method (Chen *et al*., 1997). An overnight culture was prepared and 0.8ml centrifuged at 16 000 x g for 4 seconds. Approximately 1 µg *Not*I digested DNA and 50 µg of salmon sperm DNA was added to the pellet and resuspended in 350 µl one step buffer (40% PEG in dH_2O) pre-heated to 45°C. The cells were then incubated for 30 minutes at 30°C and heat shocked for 20 minutes at 42°C. Cells were centrifuged at 4 000 x g for 2 minutes and the supernatant aspirated. The pellet was resuspended in 200 µl YPD and a volume of 150 µl was plated onto YPD with 600 µg/ml hygromycin B.

4.2.4.2 Evaluation of integration of VP2 and VP6 ORFs into yeast genomes

Following successful transformation, colonies were streaked out on YPD plates containing 800 mg/l hygromycin B and colony PCR was conducted to evaluate the integration of rotavirus ORFs (VP2 and/or VP6) into the various yeast genomes. VP2 and VP6 ORFs were amplified using KAPA Taq DNA polymerase (KAPA Biosystems) according to manufacturer's instructions. Reactions contained 1 X KAPA Taq buffer, dNTPs to a final concentration of 300 µM, 0.3 µM each of the forward and reverse primers, 1 U KAPA Taq DNA Polymerase and colony (each colony was touched with 10 µl tip and dispersed into dH$_2$O and heated for 10 minutes at 98°C) for template DNA. The reaction was filled to 50 µl with distilled water (dH$_2$O). A negative control was included which did not contain any DNA template. Amplification of colonies transformed with pKM177_VP6 vectors was carried out using VP6 yeast Forward primer 5'-CAACCTCGAGATGGATGTCCT-3' and VP6 yeast Reverse primer 5'-GTCCAGCGCTTTATTTGACAA-3' which amplify the whole VP6 ORF and yield 1207 bp. Other primers for partial amplification of VP6 ORF were used. VP6 forward (5'-AACGTTGAAGTGGAGTTTC-3') and pKM173/177_Reverse (5'-GAACAGCTAGAGTGCGTT-3') primers were used to yield approximately 733 bp amplicons. Reactions were subjected to 30 cycles using G-Storm Thermal System. Initial DNA denaturation at 95°C for 2 minutes and then cycled at 98°C for 30 seconds for denaturation, primer annealing at 54°C for 30 seconds and extension at 72°C for 2 minutes. A final extension step of 72°C for 2 minutes was included followed by holding at 4°C.

Each colony transformed with a dual expression vector pKM173_VP2/6 was amplified for both VP2 and VP6 ORFs. Primers used for amplification of colonies transformed with dual vectors are shown in Tables 4.3 and 4.4. Reactions were subjected to 30 cycles using G-Storm Thermal System. Initial DNA denaturation was carried out at 95°C for 2 minutes and then cycled at 98°C for 30 seconds for denaturation, primer annealing at 46°C for 30 seconds and extension at 72°C for 1.5 minutes. A final extension step of 72°C for 2 minutes was included followed by holding at 4°C.

Agarose gel electrophoresis was performed using 1% agarose gel as described in section 3.2.3.1.

Table 4.2: Primers used for partial amplification of VP2 ORF in colonies transformed with dual vectors.

VP2 ORF optimized	Primer name	Nucleotide sequence (5' - 3')	Amplicon size (bp)
K. lactis	KluyVP2_F2	CTGCATCAATTTGTGATGCTAT	891
	VP2_Int_R2	GACCGTTTCTGATGTTGTCG	
A. adeninivorans	ArxVP2F	GACTGTGGTCCCTAACGATAT	860
	ArxVP2R1	GTTGAGAACTCCATCAATGACG	
P. pastoris and *H. polymorpha*	VP2_Int_F1	CCGCTTCTATCTGTGACGCTAT	891
	VP2_Int_R2	GACCGTTTCTGATGTTGTCG	

Table 4.3: Primers used for partial amplification of VP6 ORF in colonies transformed with dual vectors

VP6 ORF optimized	Primer name	Nucleotide sequence (5' -3')	Amplicon size (bp)
K. lactis	VP6Seq KF3	GAGTGTTCACTGTCGCTTC	330
	pKM173/177_R	GAACAGCTAGAGTGCGTT	
A. adeninivorans	VP6Seq AF3	TCTATGCTGGTCAAGTAGAGC	304
	pKM173/177_R	GAACAGCTAGAGTGCGTT	
P. pastoris and *H. polymorpha*	VP6Seq PF3	GAGTTTTCACCGTTGCTTC	336
	pKM173/177_R	GAACAGCTAGAGTGCGTT	

4.2.5 Rotavirus VP6 protein expression

Pierce™ BCA protein assay kit (Thermo Scientific Pierce™) was used to determine the total protein to be used for western blot analysis. Albumin (BSA) standards were diluted as shown in Table 4.4. BCA Working Reagent (WR) was prepared by mixing 50 parts of BCA Reagent A with 1

part of BCA Reagent B (50:1, Reagent A:B). Absorbance was measured as described by the manufacturer's instruction using the microplate procedure. Briefly, 48-well plates (Thermo Scientific Pierce™) were used and 25 µl of each standard or unknown sample replicate was added into a microplate well. A volume of 200 µl of the WR was added to each well and mixed thoroughly on a plate shaker for 30 seconds. The plate was covered and incubated at 37°C for 30 minutes and then allowed to cool to room temperature. The absorbance was measured at 562nm on a plate reader The SpectraMax® M2 (Molecular Devices). Averaged replicate blank absorbance measurements were subtracted from the 562 nm measurements of all other individual standard and unknown sample replicates and a standard curve was drawn by plotting the average blank-corrected 562 nm measurement for each BSA standard vs. its concentration in µg/ml. The standard curve was used to determine the protein concentration of each unknown sample. Fifty microgram of total protein was used to evaluate protein expression by western blot analysis.

Table 4.4: Dilutions of Albumin (BSA) Standards

Vial	Volume of Diluent (µl)	Volume and Source of BSA (µl)	Final BSA Concentration (µg/ml)
A	0	300 of Stock	2000
B	125	375 of Stock	1500
C	325	325 of Stock	1000
D	325	325 of vial C dilution	500
E	325	325 of vial E dilution	250
F	325	325 of vial F dilution	125
G	400	0	0 = Blank

4.2.5.1 Production of rotavirus VP6 control in mammalian cells

pcDNA™3.1(–)(Invirogen™) vector contains the neomycin resistance gene for selection of stable cell lines using neomycin (Geneticin®). Antibiotic sensitivity was determined by testing a range of concentrations to determine the minimum concentration necessary for the cell line used. A 75 cm^3 tissue culture flask of confluent COS-7 cells was trypsinised and 30 ml cell suspension was prepared in DMEM medium (section 4.2.2). A volume of 1 ml cell suspension was diluted in 25 ml in DMEM medium (section 4.2.2) and added into 6 well plate in 4ml aliquots then incubated at 37°C in a Forma™ Steri-Cycle™ CO_2 incubator (Thermo Scientific) for 24 hours. After 24 hours, different concentrations of Geneticin® (Sigma-Aldrich) were added to each well: 0, 125, 250, 500, 750, and 1000 μg/ml. The selective medium was replenished every 2 days and the percentage of surviving cells was observed using Nikon Eclipse ts100 inverted microscope (Nikon) each day.

Once the appropriate Geneticin® concentration to use for selection for host cell line was determined, a cell line expressing the gene of interest was generated. Cells were transfected using TurboFect Transfection reagent (Thermo Scientific). pCDNA3.1_VP6, previously constructed by Valerie Oberhardt in our laboratory, was linearized with *BglII* to decrease the likelihood of the vector integrating into the genome in a way that disrupts the gene of interest or other elements required for expression in mammalian cells. *BglII* restriction enzyme was chosen because it cuts the vector upstream of CMV promoter and there are no *BglII* recognition sites in the insert VP6 ORF sequence. Approximately 3 μg *BglII* linearized pCDNA3.1_VP6 was diluted in 400 μl per well of serum-free DMEM growth medium (section 4.2.2). A volume of 6 μl per well TurboFect transformation reagent was added to the diluted DNA and mixed. The mixture was incubated for 1 hour at room temperature. A volume of 100 μl of the transfection reagent/DNA mixture was added drop-wise to each well and then incubated at 37°C in a CO_2 incubator. After 24 hours, medium was replenished with selective medium containing 500 ng/μl Geneticin®. Cells were analysed each day until cells without recombinant vector (negative control) died and then transfected cells were recovered in 5% FCS medium in 25 cm^3 flask. When cells were confluent in 25 cm^3 flask, they were transferred to 75 cm^3 flask with new medium. Confluent cells in 75 cm^3 flask were washed twice with ice-cold PBS and rocking

gently. Approximately 1ml for a 100 mm^3 tissue culture dish 1% SDS in TE lysis buffer (containing one tablet of protease inhibitors cocktail (Roche)), was added to the flask and incubate for 20 minutes on ice. Cells were scraped from the surface using a rubber spatula and transferred to a microcentrifuge tube. The lysate was clarified by centrifugation for 10 minutes at 16 000 x g at 4 °C. The supernatant was transferred to a new tube and stored at -20°C until use.

Western blot analysis was performed using the rotavirus group-specific antibody as described (section 4.2.3.2) to examine VP6 protein expression using the untranfected COS-7 cells as a negative control. Rotavirus proteins were probed using the primary antibody rotavirus group-specific antigen antibody (3C10), mouse monoclonal IgG$_{2a}$ (Santa Cruz)).

The mammalian recombinant expression vector was designed to include a C-terminal His tag. Therefore, western blot analysis was also performed for detection of His-tag linked to VP6 using Anti-His (C-Term)-HRP (Invitrogen). Purified Baeyer–Villiger monooxygynase (BVMO) protein obtained from Mr. Mthethwa (M.Sc student, UFS) was used as a positive control. Protein bands on the nitrocellulose membrane were developed by dissolving 4-chloro-1-naphthol peroxidase substrate tablet (Sigma) in 10 ml ice-cold methanol followed by adding 2 ml of the dissolved substrate to 10 ml PBS (pH 7.4). Five microlitres of hydrogen peroxide (Sigma) were added to the mixture and the membrane development solution was applied onto the nitrocellulose membrane for 5 minutes or until protein bands had developed sufficiently. To stop further development, the membrane was transferred into deionised water.

4.2.5.1 Rotavirus VP6 control in bacterial cells

A recombinant vector pCOLD_VP6 was provided by Prof A. Van Dijk, North West University. *E. coli* strain BL21 was transformed with 1 µl pCold^{TM} VP6 His-Tag using RbCl$_2$ transformation method as described (chapter 3 section 3.2.3.4). Transformed colonies were used to prepare a pre-inoculum into 5ml LB with 100 µg/ml of ampicillin and incubated overnight at 37℃ shaking. Expression was carried out using the pCold^{TM} Cold Shock Expression System (TAKARA) protocol. Briefly, 500 µl of pre-inoculum was inoculated in 50 ml LB medium including 100 µg/ml of ampicillin and the culture was incubated at 37°C with shaking until cells reached OD$_{600}$= 0.4 -

0.5 (0.473). The culture was refrigerated at 20℃ for 30 minutes. A final concentration of 1.0 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma) was added and the culture incubated with shaking at 20℃ for 24 hours. Cells were collected at 10 000 x g for 10 minutes at 4℃. Lysis of cells was evaluated using three different methods. This was because soluble rotavirus VP6 could not be detected when the first and second methods were used. Firstly by mechanical cell disruption by the use of French Press constant cell disruption systems one shot model (United Kingdom) at 30 KPsi in Tris buffer (100 mM NaCl and 20 mM Tris (pH 7.4). Secondly, cells were incubated for 20 minutes shaking in a commercially available BugBuster™ Protein Extraction Reagent (Novagen). A volume of 3ml Lysis buffer was added per 1 g cells. The third method was a manual lysis buffer containing 1% SDS and 1X TE and cells incubated for 20 minutes. Insoluble cell debris were removed by centrifugation at 16 000 × g for 20 minutes at 4°C and supernatant transferred to a fresh tube and stored at −20°C until use.

Western blot analysis was carried out for detection of bacterial expression of VP6 using group-specific antibodies as well as the polyclonal RV antibody (section 4.2.3.2).

4.2.5.2 Evaluation of VP6 protein expression in yeast

All nine yeast strains were revived by streaking on YPD plates until colonies formed. Colonies were then inoculated into 5 ml broth (pre-inoculum culture) and incubated at 30°C for 24 hours. Volumes of 500 μl of 24 hours cultures were inoculated in 50 ml (1/100) YPD broth in a 500 ml flask and incubated overnight.

Yeast cells were then lysed using French Press at 35 kPsi in 3ml 50 mM Tris buffer pH 7.4 per 1 g cells. Lysed and unlysed yeast cells were visualized under light microscope to ensure cell breakage. Insoluble cell debris were removed by centrifugation at 16 000 × g for 20 minutes at 4°C and supernatant transferred to a fresh tube and stored at −20°C until use. Western blot analysis was performed on soluble total fraction of expressed yeast cells for expression of VP6 using the polyclonal RV antibody (section 4.2.3.2).

4.3   Results and Discussion

4.3.1 Engineering of recombinant yeast strains for expression of rotavirus VP6 and VP2/6

Vectors containing yeast codon-optimized VP6 ORFs and a dual vector containing yeast codon-optimized VP2 and VP6 ORFs that were constructed as discussed in chapter 3 were used to transform nine yeast strains (Table 4.5).

Table 4.5: A summary of recombinant vectors used to transform yeast strains (Recombinant vectors described in Chapter 3)

Vector	Insert gene	Final product	Yeast codon-optimized	Yeast transformed
pKM177	VP6 ORF	pKM177+ VP6	*K. lactis*	All
pKM177	VP6 ORF	pKM177+ VP6	*A. adeninivorans*	
pKM177	VP6 ORF	pKM177+ VP6	*P. pastoris/H.polymorpha*	
pKM173	VP2/6 ORF	pKM173+ VP2/6	*K. lactis*	
pKM173	VP2/6 ORF	pKM173+ VP2/6	*A. adeninivorans*	
pKM173	VP2/6 ORF	pKM173+ VP2/6	*P. pastoris/H.polymorpha*	

Different concentrations of hygromycin B were used to optimize antibiotic resistance for all yeast strains (Table 4.1). Amplification of VP6 ORF in colonies transformed with pKM177 vector containing yeast codon-optimized VP6 ORFs yielded expected bands of approximately 1 207 bp (Figure 4.1) and 692 bp (Figures 4.2 and 4.3). Figure 4.1 shows PCR amplicons from yeast genomic DNA and figures 4.2 and 4.3 show colony PCR which explain intense primer dimer bands at the bottom of the gels.
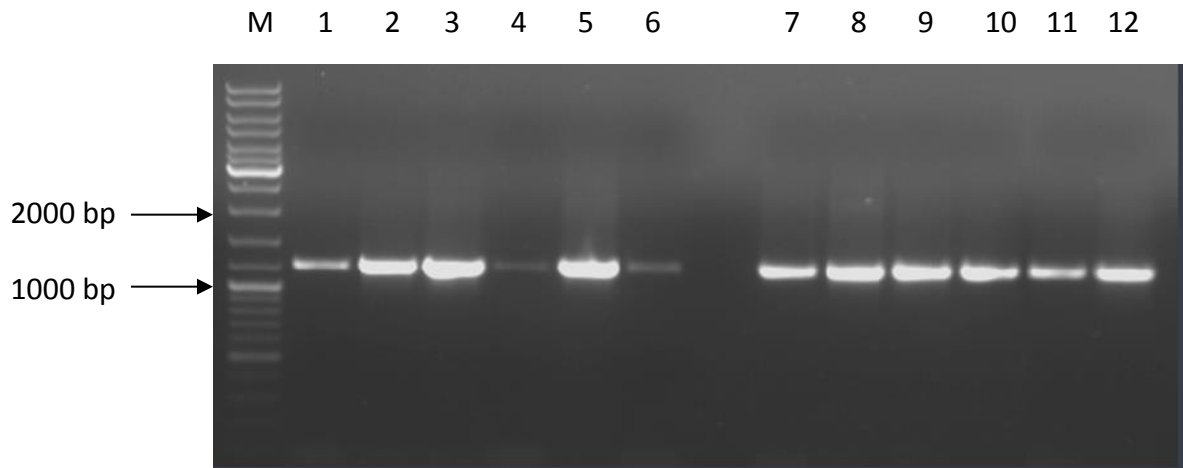
Figure 4.1: Amplification of colonies transformed with pKM177 vector containing different yeast codon-optimized VP6 ORFs. Lane M is GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-6 different yeast codon-optimized VP6 ORFs in yeast *S. cerevisiae* (lanes 1 and 2 KO, lanes 3 and 4 AO and lanes 5 and 6 PO). Lanes 7-12 different yeast codon-optimized ORFs in yeast *A. adeninivorans* (lanes 7 and 8 KO, lane 9 and 10 AO and lanes 11 and 12 PO).
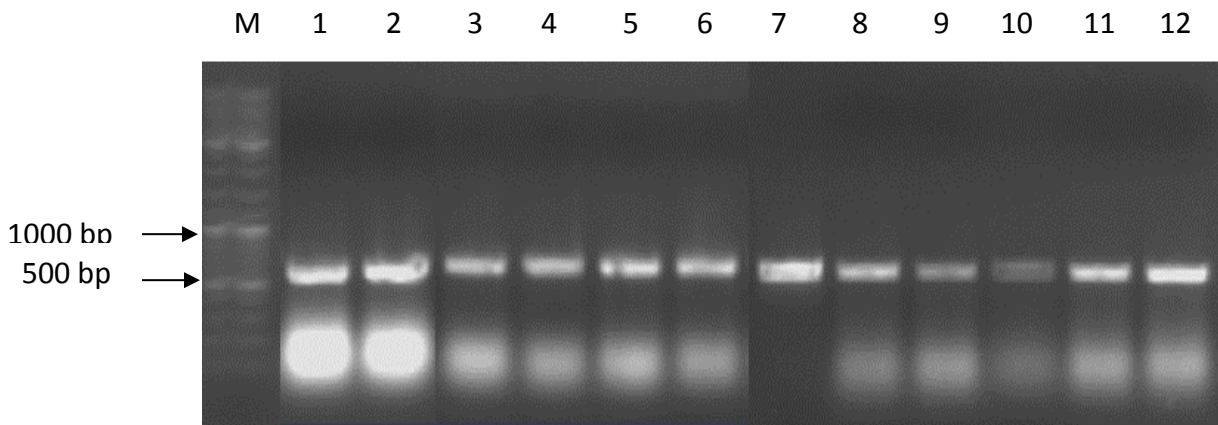
Figure 4.2: Partial amplification of VP6 from colonies transformed with pKM177 vector containing different yeast codon-optimized VP6 ORFs. Lane M is GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-6 different yeast codon-optimized VP6 ORFs in yeast *P. pastoris* (lanes 1 and 2 KO, lanes 3 and 4 AO and lanes 5 and 6 PO). Lanes 7-12 different yeast codon-optimized ORFs in yeast *H. polymorpha* (lanes 7 and 8 KO, lanes 9 and 10 AO and lanes 11 and 12 PO).
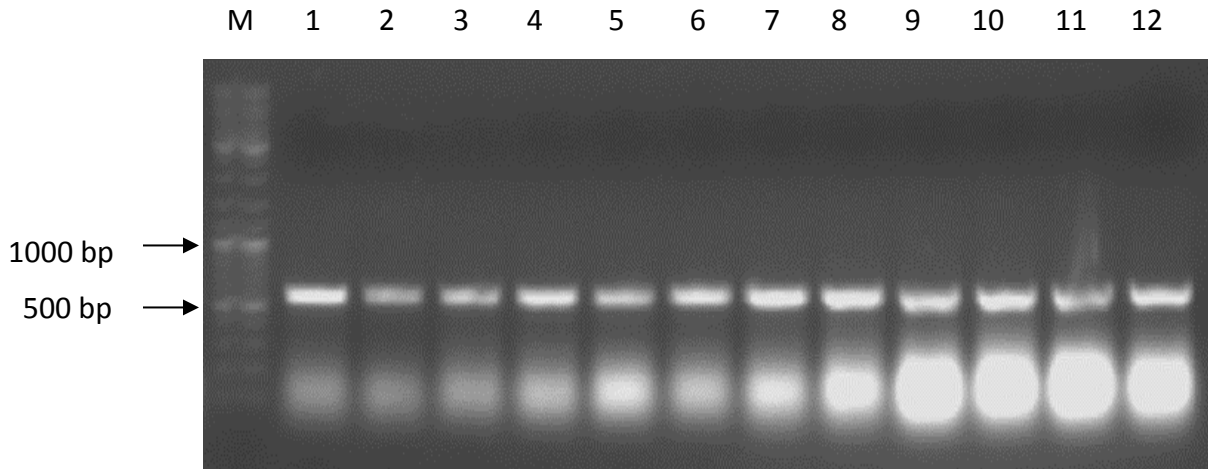
Figure 4.3: Partial amplification of VP6 from colonies transformed with pKM177 vector containing different yeast codon-optimized VP6 ORFs. Lane M is GeneRuler DNA Ladder Mix (Thermo Scientific). Lanes 1-6 different yeast codon-optimized VP6 ORFs in yeast _K. marxianus_ (lanes 1 and 2 KO, lanes 3 and 4 AO and lanes 5 and 6 PO). Lanes 7-12 different yeast codon-optimized ORFs in yeast _K. lactis_ (lanes 7 and 8 KO, lanes 9 and 10 AO and lanes 11 and 12 PO).

The yeast strains with positive VP6 ORF integration in colonies transformed with pKM177 vector containing different yeast codon-optimized VP6 ORFs are summarized in Table 4.6. _K. marxianus_ has the highest integration rate for KO (7/8) and AO (8/8) VP6 ORFs and _H. polymorpha_ had the highest integration for PO (7/8) VP6 ORF. Yeast _P. pastoris_ had the lowest rate of VP6 integration (4/12 for PO VP6 ORF and 4/12 for KO and AO VP6 ORFs) for all yeast codon-optimized VP6 ORFs.

Yeast strains that did not show positive integration of VP6 ORF did, however indicate positive transformation as colonies formed in transformed YPD plates not on the negative control. Therefore, these colonies were then amplified for the resistance gene hygromycin B (1026 bp) and all these colonies indicated positive integration for hygromycin B resistance gene (Figure 4.4). Yeast strains showing no integration for VP6 ORF but integration of hygromycin B (1026 bp) are indicated in Table 4.7.

Table 4.6: Number of colonies positive for different yeast codon-optimized VP6 ORF integration in different yeast strains

Optimized VP6 ORFs	Yeast cells					
	P. pastoris	*A. adeninivorans*	*S. cerevisiae*	*H. polymorpha*	*K. marxianus*	*K. lactis*
PO VP6	6/12	4/6	4/8	7/8	6/8	4/8
KO VP6	4/12	6/8	4/8	6/8	7/8	5/8
AO VP6	4/12	4/8	4/8	5/8	8/8	5/8
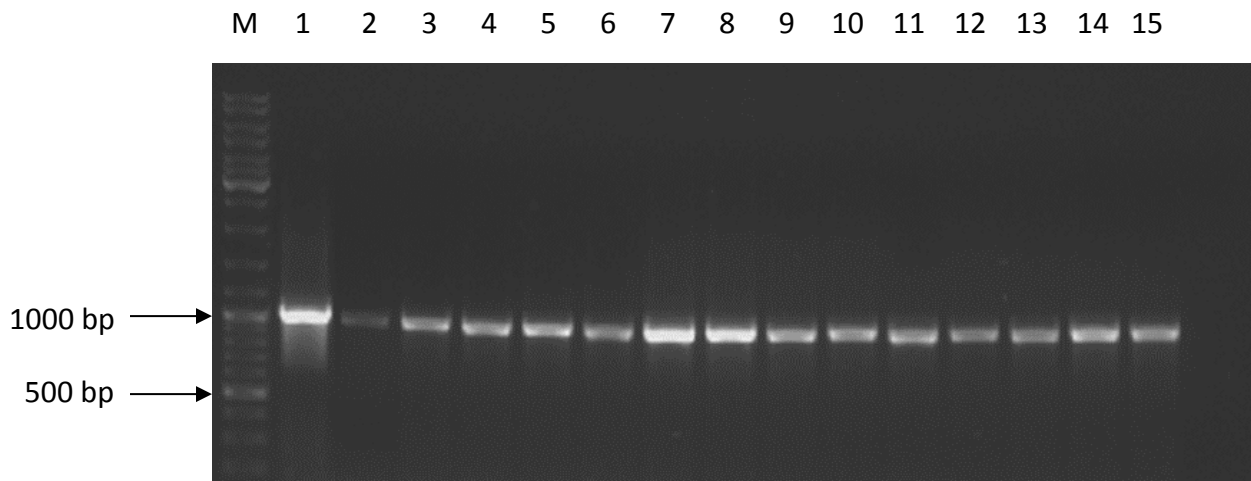
Figure 4.4: Amplification of hygromycin B gene in colonies transformed with pKM177 vector containing different yeast codon-optimized VP6 ORFs. Lane M is GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-5 *Y. lipolytica* transformed colonies, lanes 6-10 *C. deformans* transformed colonies and lanes 11-15 *D. hansenii* transformed colonies.

Table 4.7: Number of colonies positive for amplification of hygromycin B gene but not for VP6 ORF integration

Hygromycin B ORFs	Yeast cells		
	Y. lipolytica	*C. deformans*	*D. hansenii*
KO VP6	3/4	4/4	4/4
AO VP6	4/4	4/4	4/4
PO VP6	4/4	4/4	4/4

Amplification of colonies for yeast transformed with pKM173 dual vectors containing both VP2 and VP6 yeast codon-optimized ORFs indicated that all yeast strains have positive integration of both ORFs in a single colony. An illustration of VP2/6 amplification results is given in figure 4.5-4.8 Amplification indicated expected size bands of VP2 (836 bp) ORF and VP6 (326 bp).
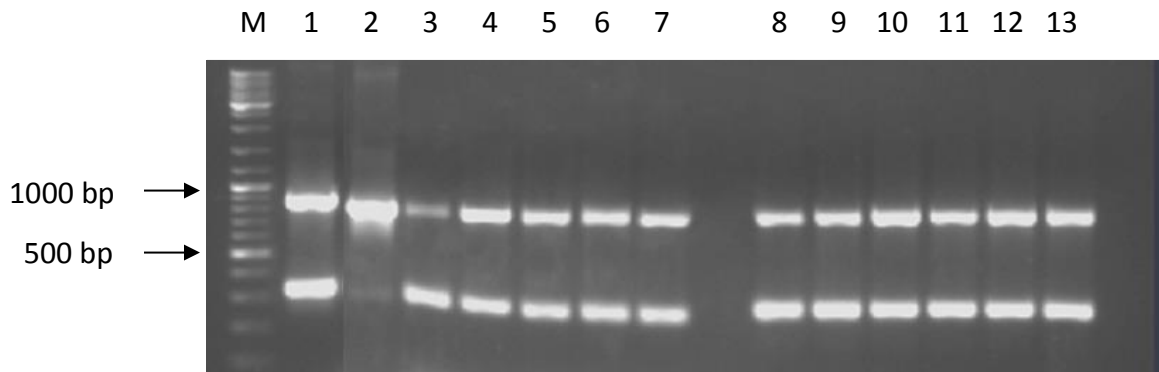
Figure 4.5: Amplification of partial VP2 (836 bp) and VP6 (326 bp) ORFs from single colonies of yeast strains transformed with dual vectors. Lane M contains GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-7 different yeast codon-optimized VP2/6 ORFs in *S. cerevisiae* (lanes 1, 2 and 3 KO, lanes 4 and 5 AO and lanes 6 and 7 PO). Lanes 8-13 different yeast codon-optimized ORFs in *A. adeninivorans* (lanes 8 and 9 KO, lanes 10 and 11 AO and lanes 12 and 13 PO).
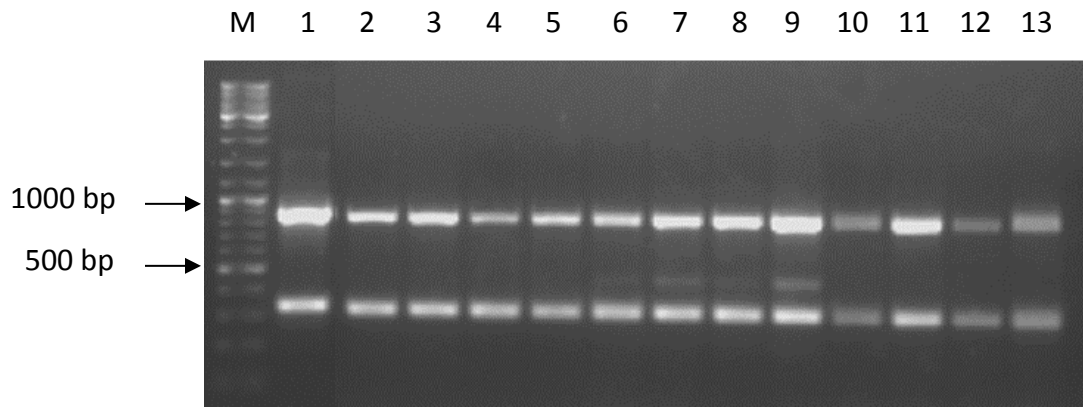
Figure 4.6: Amplification of partial VP2 (836 bp) and VP6 (326 bp) ORFs from single colonies of yeast strains transformed with dual vectors. Lane M contains GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-7 different yeast codon-optimized VP2/6 ORFs in *P. pastoris* (lanes 1, 2 and 3 KO, lanes 4 and 5 AO and lanes 6 and 7 PO). Lanes 8-13 different yeast codon-optimized ORFs in *H. polymorpha* (lanes 8 and 9 KO, lanes 10 and 11 AO and lanes 12 and 13 PO)
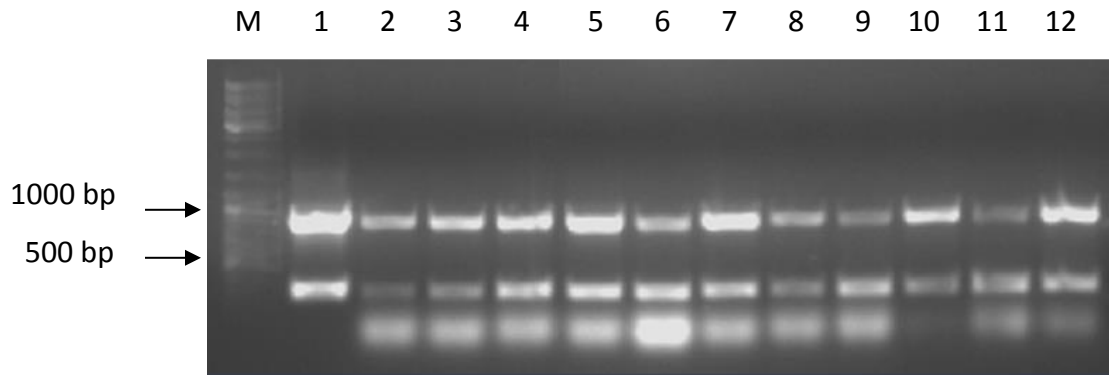
Figure 4.7: Amplification of partial VP2 (836 bp) and VP6 (326 bp) ORFs from single colonies of yeast strains transformed with dual vectors. Lane M contains GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-6 different yeast codon-optimized VP2/6 ORFs in *K. marxianus* (lanes 1 and 2 KO, lanes 3 and 4 AO and lanes 5 and 6 PO). Lanes 7-12 different yeast codon-optimized ORFs in *K. lactis* (lanes 7 and 8 KO, lanes 9 and 10 AO and lanes 11 and 12 PO).
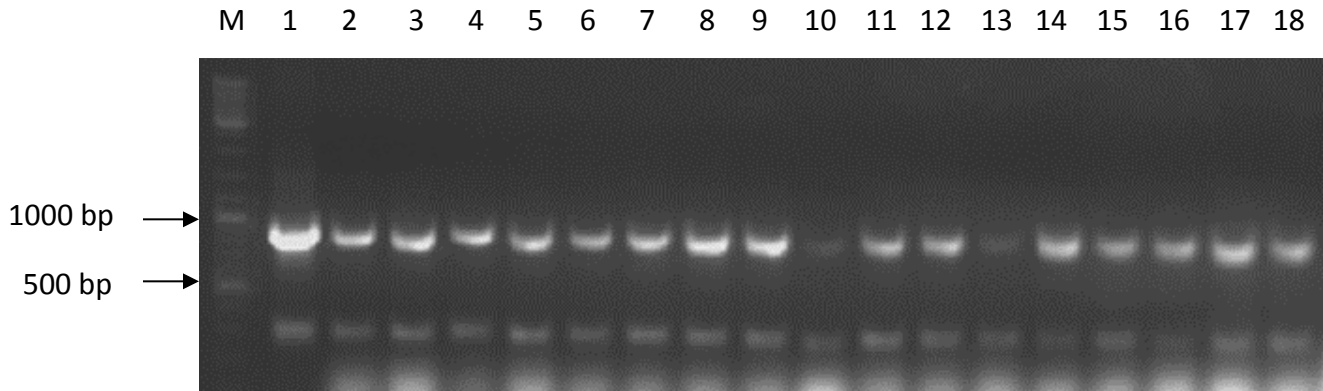
Figure 4.8: Amplification of partial VP2 (836 bp) and VP6 (326 bp) ORFs from single colonies of yeast strains transformed with dual vectors. Lane M contains GeneRuler DNA Ladder Mix (Thermo Scientific), lanes 1-6 different yeast codon-optimized VP2/6 ORFs in *Y. lipolytica* (lanes 1 and 2 KO, lanes 3 and 4 AO and lanes 5 and 6 PO). Lanes 7-12 different yeast codon-optimized ORFs in *C. deformans* (lanes 7 and 8 KO, lanes 9 and 10 AO and lanes 11 and 12 PO). Lanes 13-18 different yeast codon-optimized ORFs in *D. hansenii* (lanes 13 and 14 KO, lanes 15 and 16 AO and lanes 17 and 18 PO).

The yeast strains with positive VP6 ORF integration in colonies transformed with pKM177 vector containing different yeast codon-optimized VP2/6 ORFs are summarized in Table 4.8. *Y. lipolytica* and *K. lactis* has a 100% integration rate for KO VP2/6 ORFs (6/6). Of the colonies tested for *A. adeninivorans*, *Y. lipolytica*, *P. pastoris*, and *H. polymorpha* 100% integration for AO VP2/6 ORF were obtained. *K. marxianus*, *P. pastoris, C. deformans* and *D. hansenii* had the 100% rate of VP2/6 PO ORF integration. *S. cerevisiae* had 100% rate of VP2/6 integration (3/3 for AO VP2/6 ORF and 6/6 for KO and PO VP2/6 ORFs) for all yeast codon-optimized VP2/6 ORFs. *K. marxianus* had the lowest integration rate of VP2/6 for KO VP2/6 ORF (2/6), *K. marxianus* and *K. lactis* had the lowest integration rate of VP2/6 for AO VP2/6 ORF (4/6), and *Y. lypolytica* had the lowest integration rate of VP2/6 for PO VP2/6 ORF (3/6). Integration of different yeast codon-optimized VP2/6 ORFs was better than integration of different yeast codon-optimized VP6 ORFs as all yeast strains had VP2/6 integration.

Table 4.8: Number of colonies positive for VP2 and VP6 ORF integration in different yeast strains transformed with dual vectors

Optimized ORFs	Yeast Strains								
	A. adeninivorans	*Y. lypolytica*	*K. marxianus*	*P. pastoris*	*S. cerevisiae*	*K. lactis*	*H. polymorpha*	*C. deformans*	*D. hansenii*
KO VP2/6	3/7	6/6	2/6	5/6	6/6	6/6	3/6	4/6	5/6
AO VP2/6	7/7	6/6	4/6	5/5	3/3	4/6	5/5	5/6	5/6
PO VP2/6	5/7	3/6	6/6	6/6	6/6	4/6	5/6	6/6	6/6

4.3.2 Expression of rotavirus VP6 protein

Pierce^{TM} BCA protein assay kit (Pierce^{TM}, Biotechnology, Inc) was used for estimating the total protein to be used for western blot analysis. Absorbance readings at a wavelength of 562 nm of albumin standards were measured and standard curve was drawn (Figure 4.9) using known concentrations of albumin standard ampules.
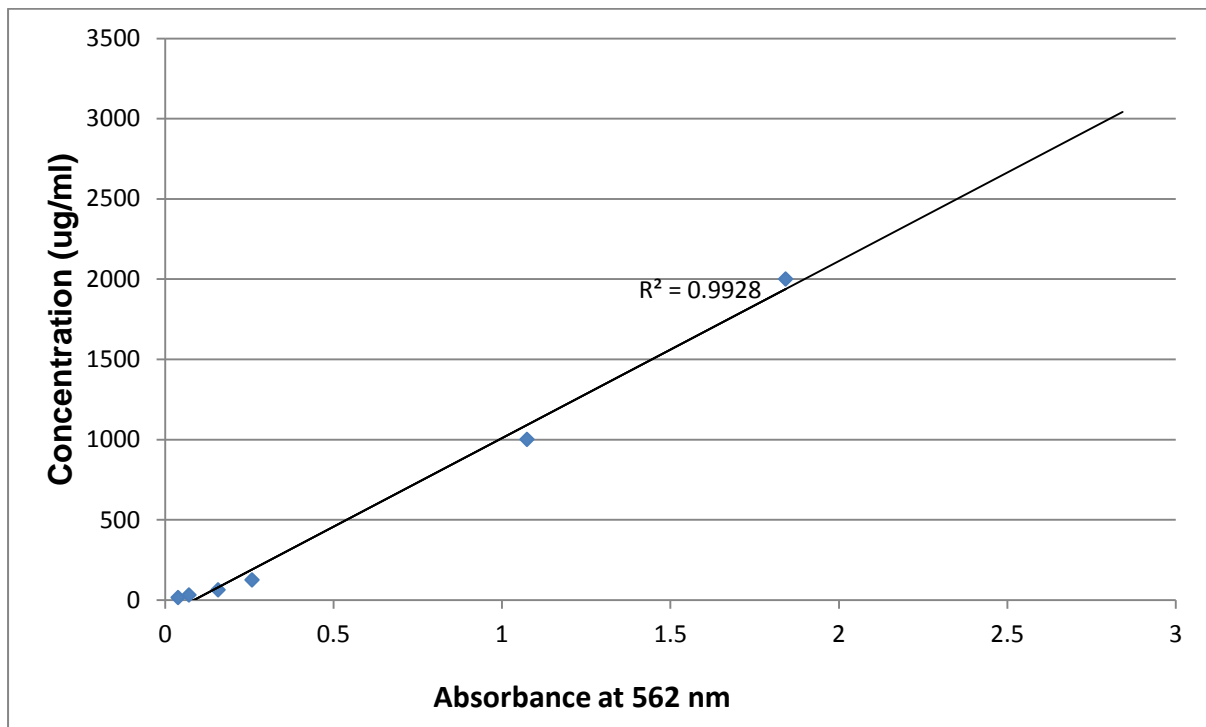
Figure 4.9: Standard curve of absorbance reading at a wavelength of 562 nm to known concentrations (µg/ml) of albumin standard ampules.

4.3.2.1 Rotavirus VP6 expression in mammalian cells

Western blot analysis of VP6 expressed in mammalian cells using rotavirus group-specific mouse monoclonal antibody as primary antibody and goat-anti mouse antibody as secondary antibody produced a non-specific reaction against both transfected and non-transfected cells (Figure 4.10). A size of 48.16 kDa (Prasad *et al.*, 1988) was expected but a band of approximately 62 kDa was observed. A 4-chloro-1-naphthol peroxidase substrate method was used for detection of VP6 and Baeyer–Villiger monooxygynase (BVMO) protein (positive control) both containing a His-tag. Specific reaction was obtained for the positive control (~63 kDa) but no reaction was obtained for VP6 expressed in COS-7 cells (Figure 4.11) where a 48.16 kDa (Prasad *et al.*, 1988) band size was expected. It is possible that the

secondary IgG monoclonal antibody reacted non-specifically with albumin present in the serum added. No such secondary antibody was used for the His-tag-detection and therefore the non-specific 62 kDa band was not detected in recombinant COS-7 cells. The absence of the VP6 protein band indicates that the mammalian cells were not successfully transformed for with the recombinant vector containing VP6 ORF.
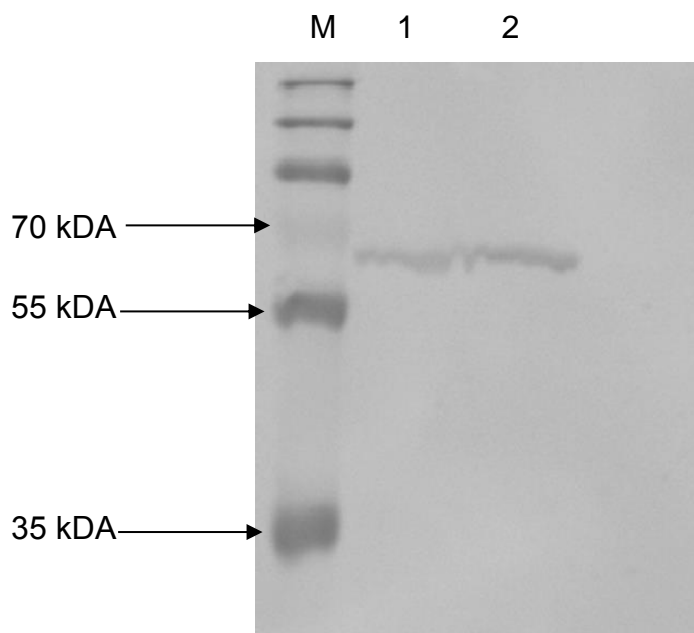
Figure 4.10: Western blot analysis of rotavirus VP6 using monoclonal rotavirus-specific antibody. Lane M is PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane 1 COS-7 cells transfected with pcDNA3.1_VP6 (His-tag) and lane 2 non-transfected COS-7 cells.
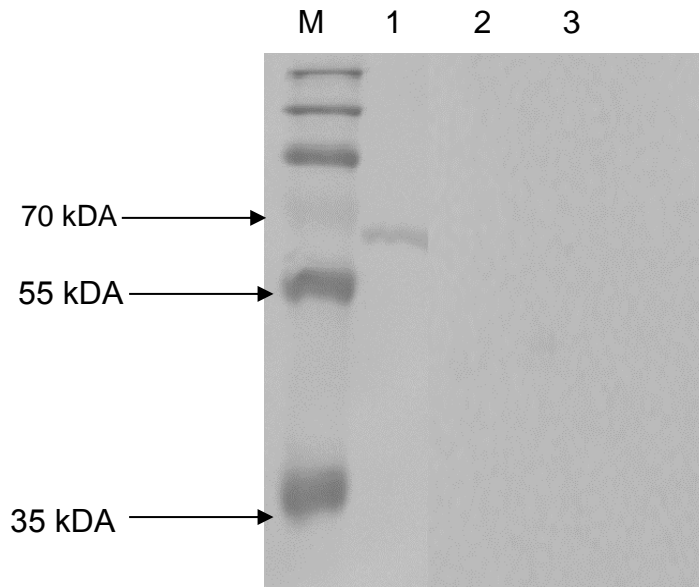
Figure 4.11: Western blot analysis against His-tag using anti-His (C-Term)-HRP antibody. Lane M is PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane 1: BVMO, lane 2: COS-7 cells transfected with pcDNA3.1_VP6 (His-tag) and lane 3: non-transfected COS-7 cells.

4.3.2.2   Rotavirus VP6 expression in bacterial cells

Western blot analysis of rotavirus VP6 expressed in BL21 cells resulted in no reaction (Figure 4.12) for BL21 cells lysed with all three lysis buffers (1% SDS in 1 x TE, BugBuster™ Protein Extraction Reagent and French Press) when rotavirus proteins were probed using the rotavirus group-specific antibody. Detection was achieved using BCIP-T/NBT alkaline substrate method and also by 4-chloro-1-naphthol peroxidase substrate method.

When rotavirus VP6 expressed in BL21 cells was probed by primary antibody rotavirus group-specific antibody, no reaction was obtained using 4-chloro-1-naphthol peroxidase substrate detection method. Detection using 4-chloro-1-naphthol peroxidase substrate method revealed expression of VP6 at the size of 48.18 kDa (Prasad *et al*., 1988) in both total and soluble fractions (Figure 4.13) when BL21 cells were lysed with 1% SDS 1 x TE buffer only when polyclonal primary antibody raised against NCDV was used. The polyclonal antibody raised against NCDV was previously successfully used to detect expression of rotavirus structural proteins for the same strain used in this study (Jere *et al.,* 2014). VP6 expression could also be detected in the total fraction of cells lysed with French Press but no reaction was observed when cells were lysed with BugBuster™ Protein Extraction Reagent.
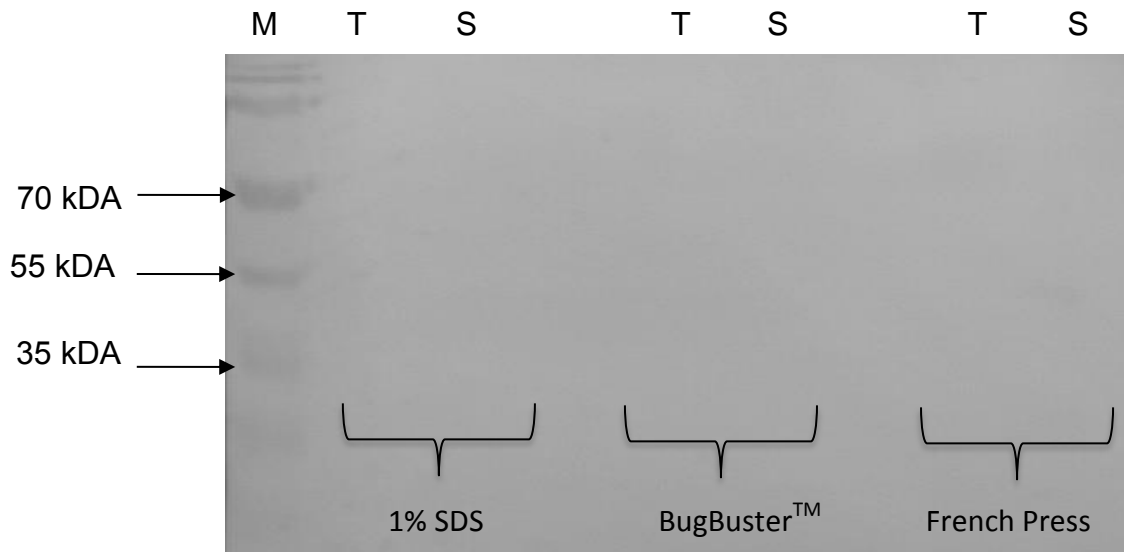
Figure 4.12: Western blot analysis of rotavirus VP6 expressed in BL21 cells lysed with all three cell disruption methods in total (T) and soluble (S) fractions of a total protein detected with BCIP-T/NBT alkaline substrate method. M indicates PageRuler™ plus Prestained Protein Ladder (Thermo Scientific)
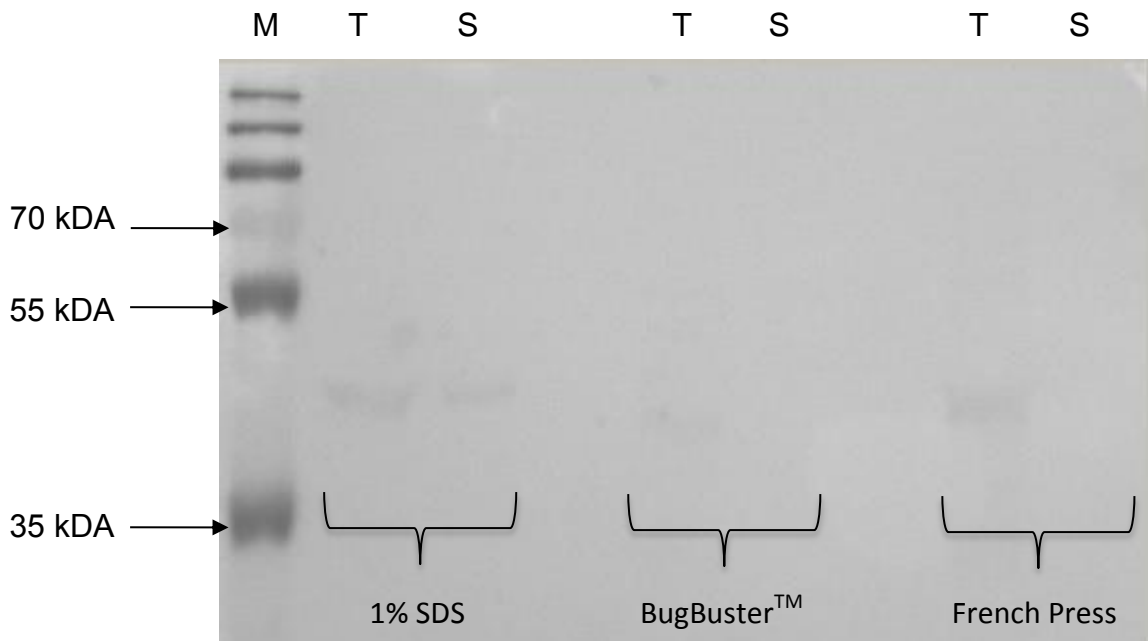
Figure 4.13: Western blot analysis of rotavirus VP6 expressed in BL21 cells lysed with all three cell disruption methods in total (T) and soluble (S) fractions of a total protein detected using 4-chloro-1-naphthol peroxidase substrate method. M indicates PageRuler™ plus Prestained Protein Ladder (Thermo Scientific).

4.3.2.3 Rotavirus VP6 expression in yeast cells

Yeast cells were lysed using the French Press at 35 kPsi and cells were visualized under light microscope to evaluate the structure of lysed and unlysed cells (Figure 4.15). This was done to confirm that at least 70% of yeast cells were lysed in order to progress to protein analysis using western blot analysis.
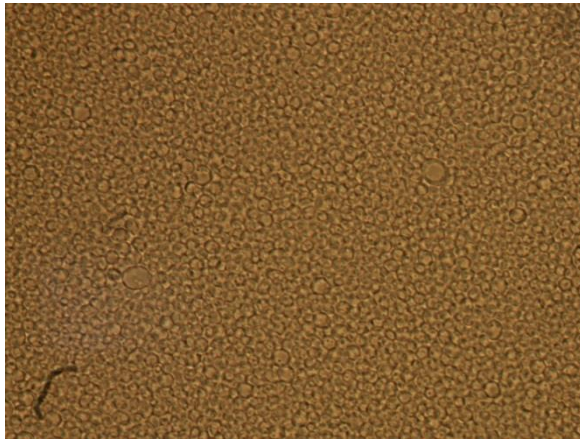
P. pastoris unlysed
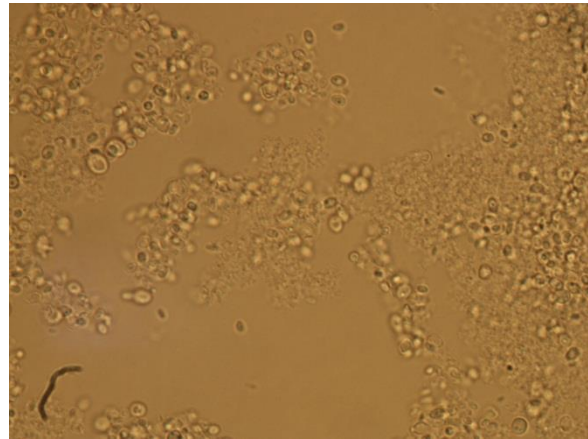
P. pastoris lysed
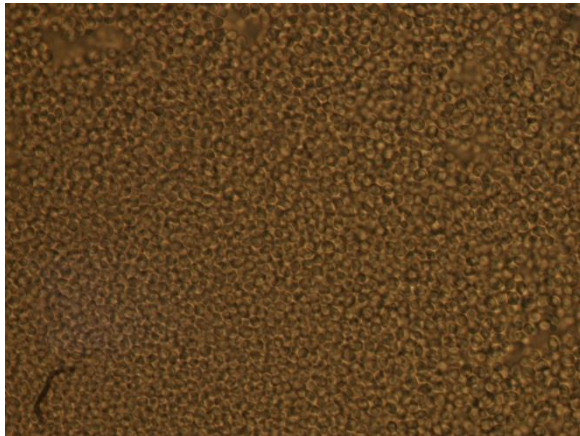
H. polymorpha unlysed
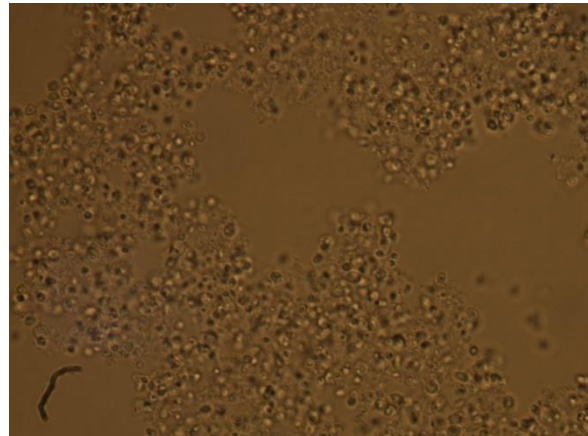
H. polymorpha lysed
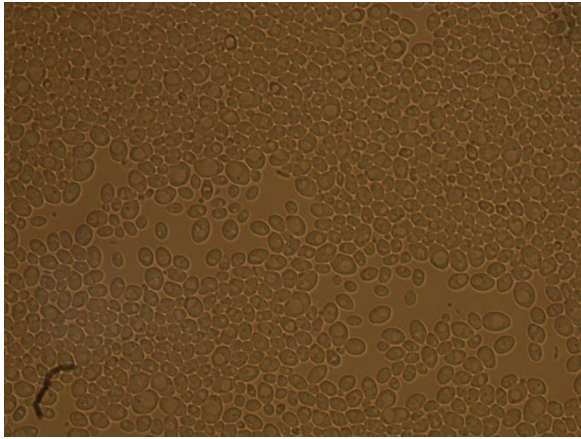
S. cerevisiae unlysed
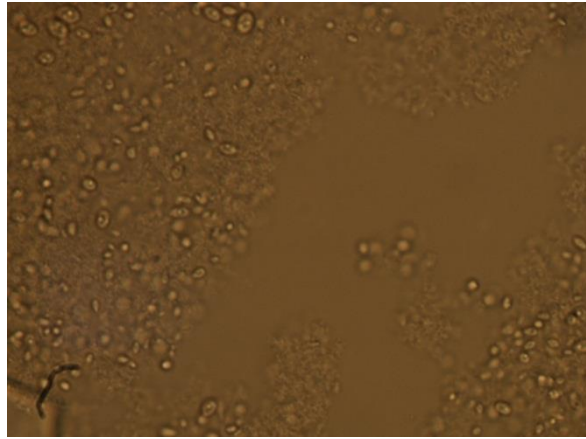
S. cerevisiae lysed

A. adeninivorans unlysed
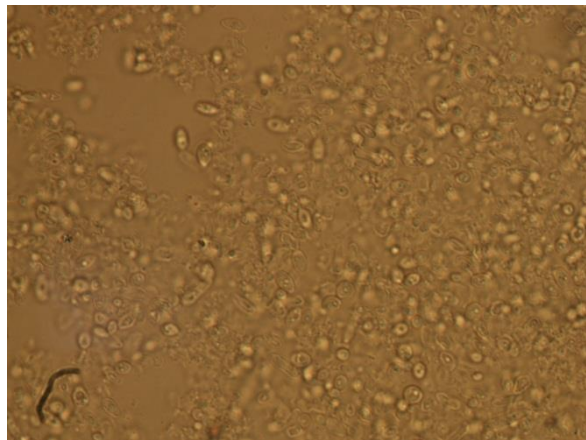
A. adeninivorans lysed

K. marxianus unlysed
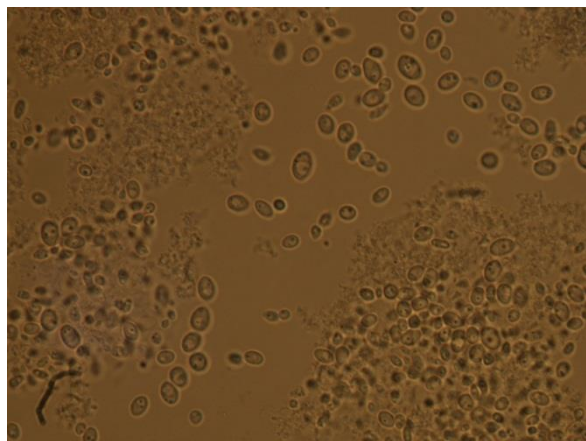
K. marxianus lysed
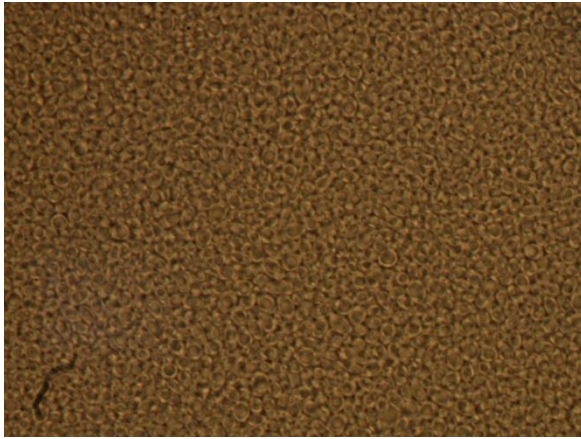
K. lactis unlysed
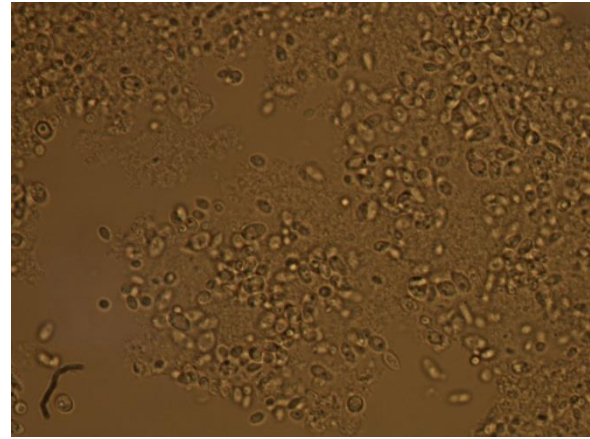
K. lactis lysed
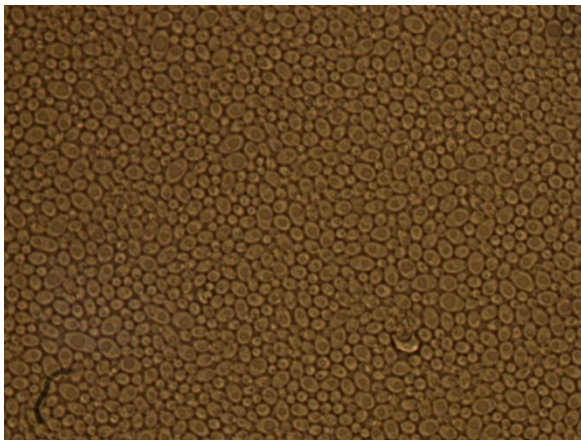
Y. lipolytica unlysed
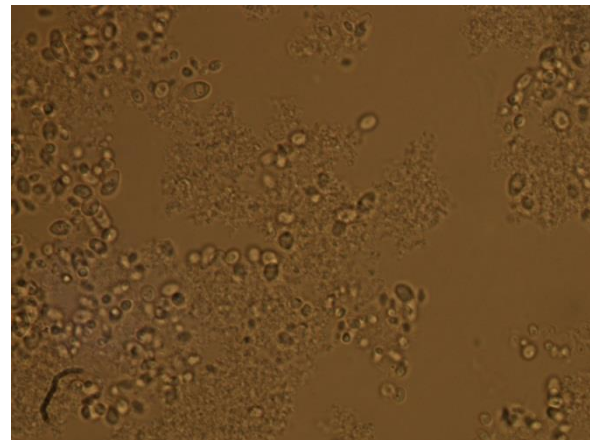
Y. lipolytica lysed
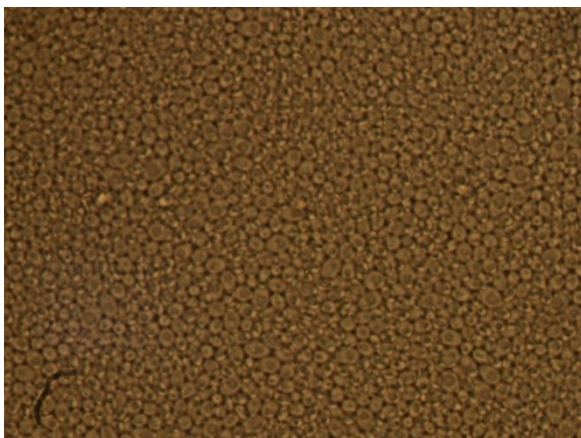
D. hansenii unlysed

D. hansenii lysed
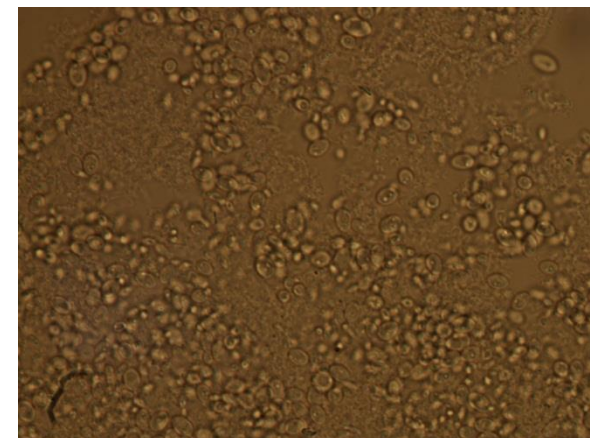
<p style="text-align: center;">C. deformans unlysed   C. deformans lysed</p>
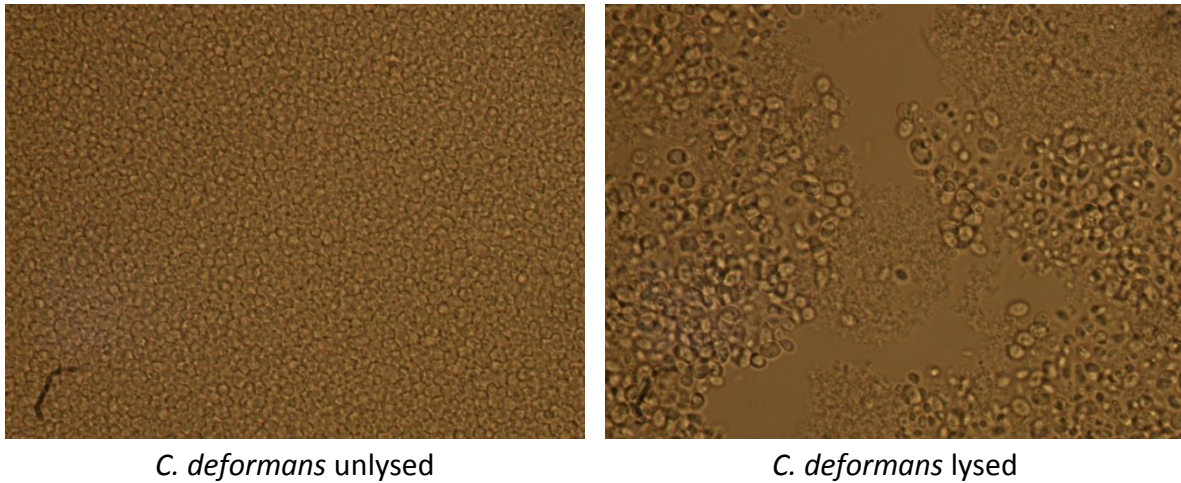
Figure 4.14: Comparison of unlysed and lysed yeast cells with French Press as evaluated using an Axioskop light microscope at 100X magnification (Zeiss, Germany).

A total of two clones for each yeast optimized ORF in each yeast showing VP6 integration were analysed for VP6 expression. Western blot analysis for VP6 expression in yeast strains showed no reactions for *K. lactis* and *A. adeninivorans* codon-optimized VP6 ORF in all six yeast strains that tested positive for intergration of VP6 (Figures 4.16-4.21). All yeast strains except *K. marxianus* showed positive reaction for VP6 expression for *P.pastoris/H. polymorpha* codon-optimized VP6 ORF. However, VP6 band occurred at a slightly lower position than expected. The expected band was at 48.16 kDa (Prasad *et al.,* 1988) which was evident for rotavirus VP6 expressed in BL21 *E. coli* strain but an approximate 40 kDa protein was observed for rotavirus VP6 expressed in yeast strains. Untransformed yeast strains were included that did not react with antibody and therefore the result does not seem to be non-specific.
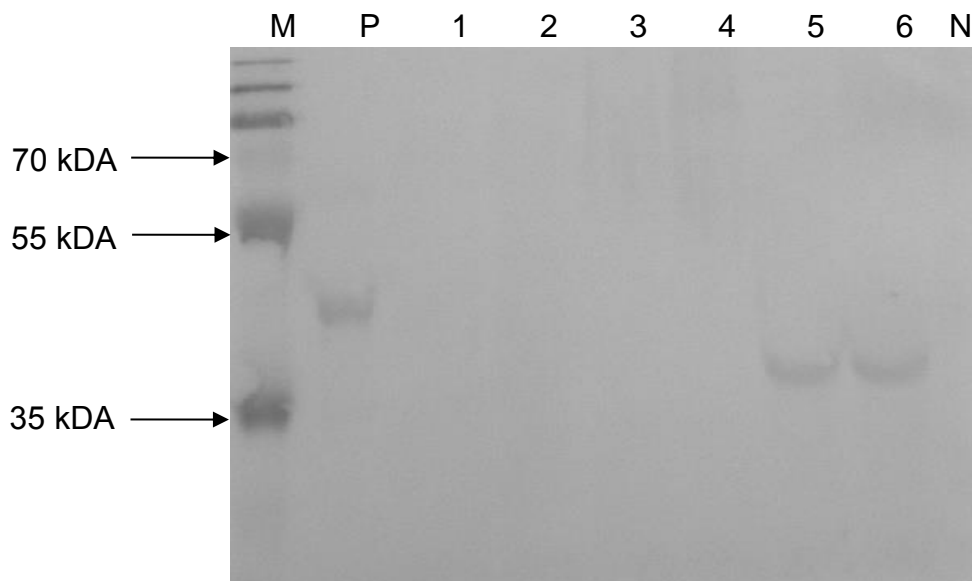
Figure 4.15: **Western blot analysis of rotavirus VP6 expressed in** *P. pastoris.* **Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.**
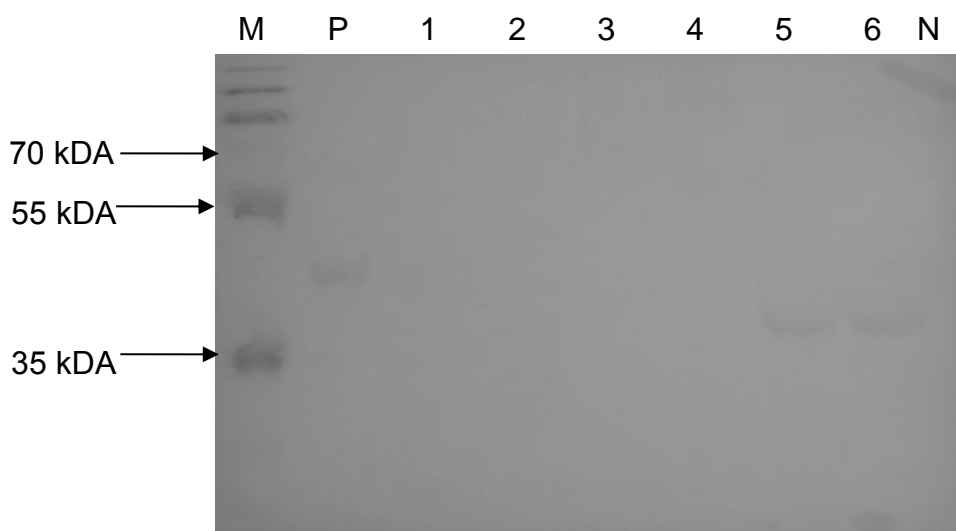
Figure 4.16: **Western blot analysis of rotavirus VP6 expressed in** *H. polymorpha.* **Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.**
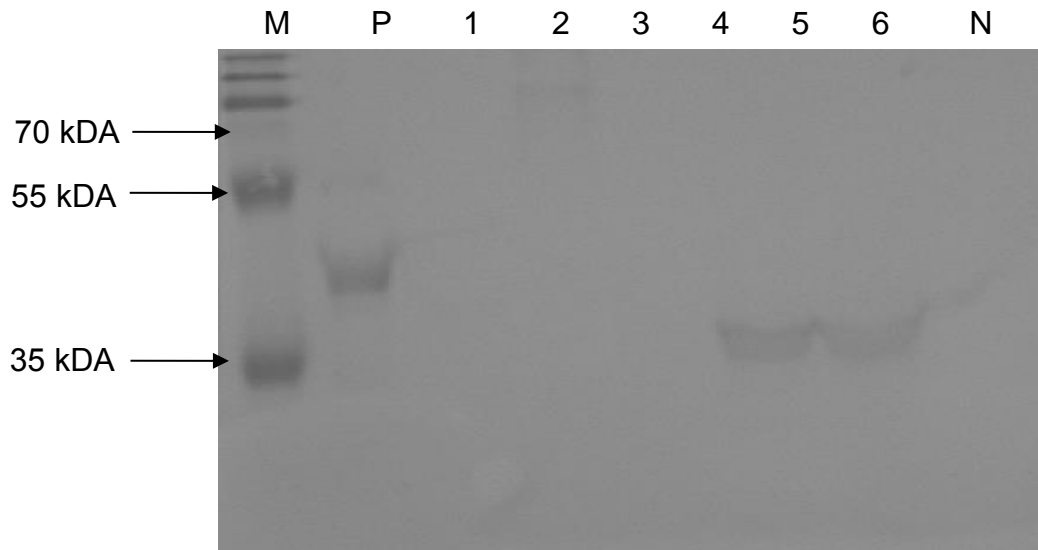
Figure 4.17: Western blot analysis of rotavirus VP6 expressed in *A. adeninivorans.* Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.
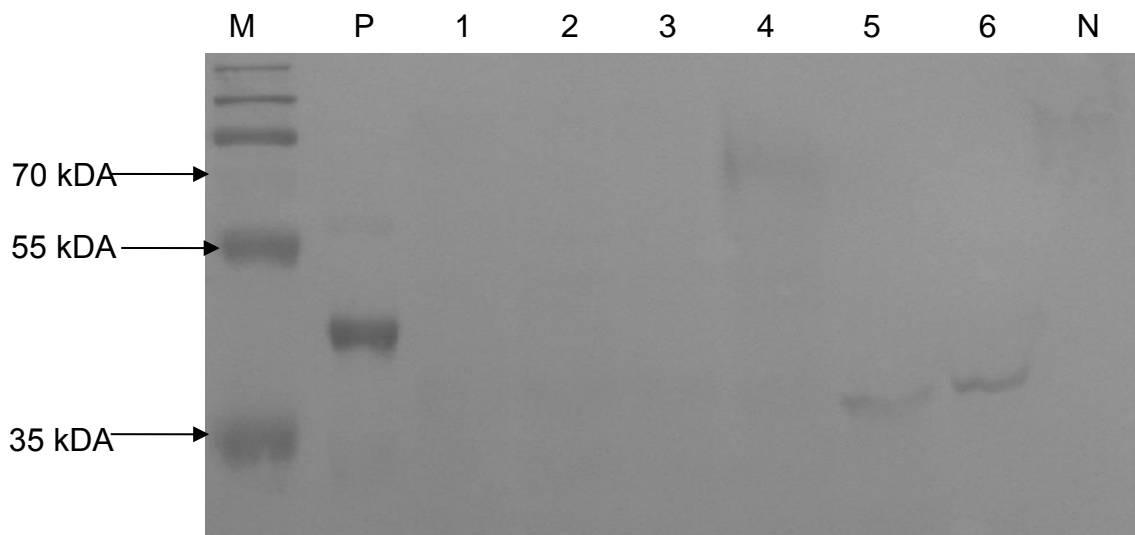
Figure 4.18: Western blot analysis of rotavirus VP6 expressed in *S. cerevisiae.* Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.
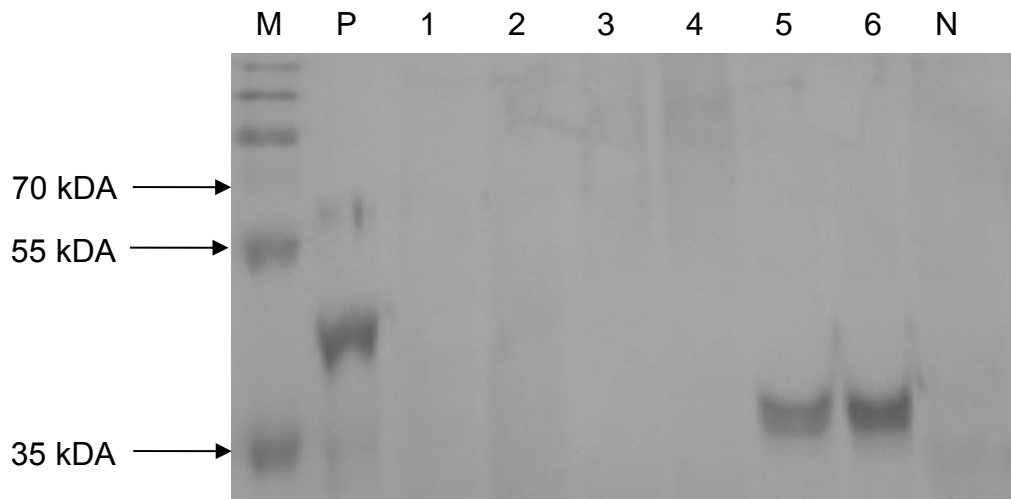
Figure 4.19: Western blot analysis of rotavirus VP6 expressed in *K. lactis*. Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.
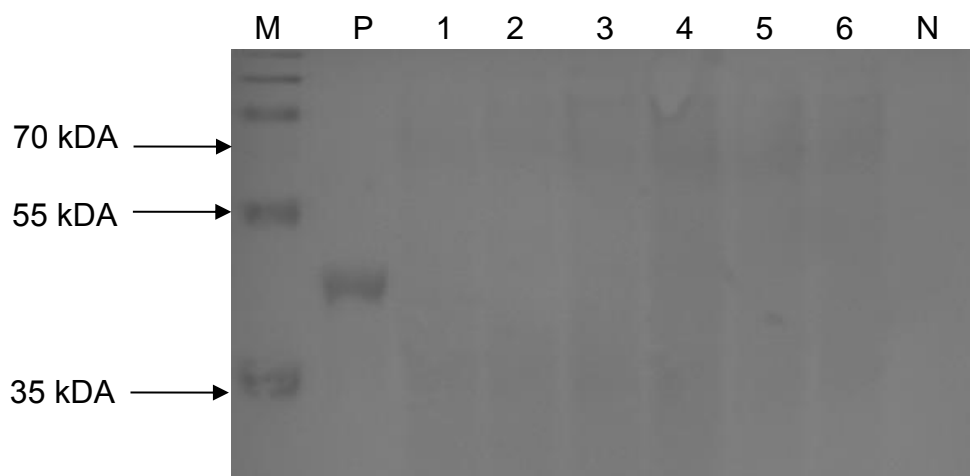
Figure 4.20: Western blot analysis of rotavirus VP6 expressed in *K. marxianus*. Lane M: PageRuler™ plus Prestained Protein Ladder (Thermo Scientific), lane P: bacterial-expressed VP6, lanes 1 and 2: Two clones of KO VP6 ORFs, lanes 3 and 4: Two clones of AO VP6 ORFs, lanes 5 and 6: Two clones of PO optimized VP6 ORFs and lane N is non-transformed yeast cells.

4.4 Summary

In this chapter, nine yeast strains were transformed with pKM177 vectors containing yeast codon-optimized VP6 ORFs. However, a total of six yeast strains indicated positive integration for VP6 ORF as detected by PCR. The other three yeast strains indicated no VP6 ORF integration but hygromycin B resistance gene integration suggesting that the vectors successfully integrated in these yeast strains but VP6 ORF was lost. Integration of both rotavirus VP2/6 ORFs was evident in all yeast strains transformed with dual expression vectors. Using different yeast codon-optimized ORFs did not show any significant difference in integration of ORFs in yeast strains in that integration was not favoured in yeast strains for which ORFs were optimized for.

An attempt to prepare a control for VP6 expression in mammalian cells was unsuccessful. It was however possible to prepare a rotavirus VP6 protein expression control in bacterial cells. VP6 expression in bacteria was shown to be more evident when using 1% SDS buffer as compared to a commercially available BugBuster™ and French Press. The SDS buffer was previously used to solubilize rotavirus proteins expressed in insect cells (Labbe *et al.*, 1992). The positive control indicated specific reaction when polyclonal rotavirus antibody raised against NCDV rotavirus strain was used but no reaction was observed when rotavirus group-specific mouse monoclonal antibody was used.

Western blot analysis for VP6 expression in yeast strains showed no reactions for *K. lactis* and *A. adeninivorans* codon-optimized VP6 ORF in all six yeast strains that tested positive for intergration of VP6. All yeast strains except *K. marxianus* showed positive reaction for VP6 expression for *P. pastoris/H. polymorpha* codon-optimized VP6 ORF. However, VP6 band occurred at a slightly lower position than expected.

CHAPTER 5

Concluding remarks and future prospects

In this study, we used the wide-range yeast expression system to generate recombinant yeast strains with the potential to express VP2 and VP6. These vectors allow the screening of a variety of yeasts in parallel, hence minimising the workload. Dual expression vectors allow for simultaneous expression of VP2 and VP6 proteins which enables the formation of double-layered rotavirus-like particles. Rotavirus VP2 and VP6 ORFs were codon-optimized to increase the level of expression and also protein stability in yeasts, *K. lactis*, *A. adeninivorans* and *P. pastoris/H. polymorpha*. These were done because in a previous related study performed by A. Naude, wild-type VP2 and VP6 ORFs were used for expression in yeast strains and although integration of these ORFs was observed, no protein expression was obtain when using western blot analysis. Therefore, codon optimization was used to favour expression in yeast strains. The optimized ORFs differ from wild type in nucleotide sequences but have the same amino acid sequences (APPENDICES A and B, Figures 10-13).

Wide-range expression vectors used in this study were previously used to express a cytochrome P450 monooxygenase (P450) in different yeasts namely: *Saccharomyces cerevisiae*, *Kluyveromyces marxianus*, *Yarrowia lipolytica* and *Arxula adeninivorans* (Theron *et al*., 2014). In this study, these vectors are also proven to be appropriate in transformation of rotavirus VP2 and VP6 ORFs in all nine yeasts strains used in this study namely: *Kluyveromyces marxianus, Kluyveromyces lactis, Debaryomyces hansenii, Yarrowia lipolytica, Hansenula polymorpha, Pichia pastoris, Candida deformans, Arxula adeninivorans* and *Saccharomyces cerevisiae*.

Transformation of yeast strains in this study with the pKM177 vector containing codon optimised rotavirus VP6 ORF was successful for 6 yeast strains (*Kluyveromyces marxianus, Kluyveromyces lactis, Hansenula polymorpha, Pichia pastoris, Arxula adeninivorans* and *Saccharomyces cerevisiae*). For three yeast strains (*Y. lipolytica*, *D. hansenii* and *C. deformans*) no VP6 ORF integration was obtained but hygromycin B integration suggesting that the vectors successfully integrated in these yeast strains but the VP6 ORF was lost. Surprisingly, when VP2 and VP6 ORFs were transformed simultaneously in different yeast strains, integration of both ORFs was evident in all yeast strains even those yeast strains

where VP6 ORF integration alone was not observed. A similar trend was observed by Rodrieguez-Limas and colleagues based on level of VP6 expression that was lower when rotavirus VP6 is expressed alone than when it was expressed simultaneously with rotavirus VP2 and VP7 (Rodrieguez-Limas *et al*., 2011).

Using different yeast codon-optimized ORFs did not show any significant difference in integration of ORFs in yeast strains in that integration was not favoured in specific yeast strains for which ORFs were optimized for. Integration of ORFs was evenly distributed among all yeast strains with different yeast codon-optimized ORFs. *K. marxianus* had high efficiency of integration of VP6 ORFs while *P. pastoris* had the lowest rate of VP6 integration. *S. cerevisiae* had high efficiency of integration for VP2/6 ORFs simultaneously while *Y. lipolytica* had the lowest integration rate for VP2/6 ORFs simultaneously. This could mean that codon optimization has no effect on integration as even wild-type rotavirus VP2 and VP6 ORFs successfully integrated in a previous related study by Aliza Naudé although the integration rate was not determined.

Rotavirus VP6 protein was successfully expressed in bacterial cells for use as a positive control. VP6 expression in bacteria was detected when cells were lysed using 1% SDS buffer as compared to a commercially available BugBuster™ and physical lysis using French Press where no expression could be detected. The SDS buffer contains salts (TE) to regulate the acidity and osmolarity of the lysate and contains SDS which is a detergent that breaks up membrane structures. SDS is ionic detergent and is regarded to be harsh on proteins, this result in high yield of soluble protein (Green & Sambrook, 2012). SDS buffer has previously been used to solubilize rotavirus proteins expressed in insect cells (Labbe *et al*., 1992). Lysing cells with 1% SDS buffer resulted in some of the protein being present in soluble form which was able to react with polyclonal rotavirus antibody raised against NCDV rotavirus strain. However, no reaction was observed when rotavirus group-specific mouse monoclonal antibody was used. The NCDV antibody that was previously successfully used to detect RV proteins from the same strain in insect cells (Jere *et al.,* 2014).

Western blot analysis for VP6 expression in yeast strains indicated no reactions for *K. lactis* and *A. adeninivorans* codon-optimized VP6 proteins in all six yeast strains that tested positive for intergration of VP6 namely: *Kluyveromyces marxianus, Kluyveromyces lactis, Hansenula polymorpha, Pichia pastoris, Arxula adeninivorans* and *Saccharomyces cerevisiae*.

Five yeast strains with the exception of *K. marxianus* showed possible reaction for VP6 expression for *P.pastoris/H. polymorpha* codon-optimized VP6 ORF. This indicates that codon optimization does have an effect on protein expression. However, expression of VP6 did not favour the specific yeast strains for which codons were optimized. In contrast *P.pastoris/H. polymorpha* codon-optimized VP6 ORF was the only one showing expression in all yeast strains except *K. marxianus. P.pastoris/H. polymorpha* codon-optimized VP6 ORF were synthesised by GeneArt while *K. lactis* and *A. adeninivorans* codon-optimized VP6 ORF were synthesised by GenScript. The inability of the yeats strains containing the *K. lactis* and *A. adeninivorans* codon-optimized VP6 ORF must be explored further. This may be achieved by screening more transformed colonies and/or obtaining KO and AO ORFs form GeneArt as they were obtained from GenScript.

Evaluation of expression of rotavirus VP6 in yeast using western blot analysis resulted in a smaller size protein band than VP6 expression in bacterial cells of which VP6 sequece is from the same rotavirus strain used in expression of yeast strains. A similar size of VP6 using the same rotavirus strain was obseverd when *P. pastoris/H. polymorpha* codon-optimized VP6 ORF was expressed in *P. pastoris* and *H. polymorpha* (unpublished data, Prof. J Görgens, Stellenbosch University) and also when tobacco plants optimized VP6 ORF were expressed in plants (unpublished data, Dr. Inga Hitzeroth, University of Cape Town). The expressed VP6 protein has to be further identified with techniques such as mass spectrophotometry and/or by the use of electron microscopy, as it has been shown that when VP6 is expressed alone, globular structures are observed under EM (Blazevic *et al*., 2011). If indeed the expression of VP6 can be confirmed, the reason for the difference in size must be established. A possible explanation for the size difference is post-translational modifications in yeast, possibly protease cleavage which would require the identification of the cleavage site and site-directed mutagenesis to remove the probable cleavage site. Simultaneous expression of VP2/6 still has to be evaluated as nine recombinant yeast strains containing both VP2 and VP6 ORFs have been engineered in this study. Simultaneous expression of VP2/6 was not explored in this study due to time constraints. Expression will be followed by evaluation of formation of double-layered rotavirus particles in transformed yeast strains and VP6 and VP2 produced will be tested in animal models to confirm their immunogenicity.

Reference List

Afrikanova, I. & Fabbretti, E. (1998). Rotavirus NSP5 phosphorylation is up-regulated by interaction with NSP2. *J Gen Virol* **2**:2679–86.

Aiyegbo, M. S., Sapparapu, G., Spiller, B. W., Eli, I. M., Williams, D. R., Kim, R *et al*., (2013). Human Rotavirus VP6-Specific Antibodies Mediumte Intracellular Neutralization by Binding to a Quaternary Structure in the Transcriptional Pore. *PLoS One* **8:**e61101.

Albertyn, J., Labuschagne, M., Theron, C. & Smit, M. (2011). Novel expression constructs. South Africa. Patent number 2011/04652. Filed 23 June 2011.

Angel, J., Franco, M. A. & Greenberg, H. B. (2012). Rotavirus immune responses and correlates of protection. *Curr Opin Virol* **2**:1–7.

Anthony, I., Bullivant, S. & Dayal, S. (1991). Rotavirus spike structure and polypeptide composition. *J Virol* **65**:4334-40.

Appaiahgari, M. B., Glass, R., Singh, S., Taneja, S., Rongsen-Chandola, T., Bhandari, N *et al.*, (2014). Transplacental rotavirus IgG interferes with immune response to live oral rotavirus vaccine ORV-116E in Indian infants. *Vaccine* **32**:651–6.

Armah, G. E., Sow, S. O., Breiman, R. F., Dallas, M. J., Tapia, M. D., Feikin, D. R *et al.*, (2010). Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. *Lancet* **376**:606–14.

Armah, G. E., Kapikian, A. Z., Vesikari, T., Cunliffe, N., Jacobson, R. M., Burlington, D. B. *et al.*, (2013). Efficacy, immunogenicity, and safety of two doses of a tetravalent rotavirus vaccine RRV-TV in Ghana with the first dose administered during the neonatal period. *J Infect Dis* **208**:423–31.

Au, K., Chan, W., Burns, J. & Estes, M. (1989). Receptor activity of rotavirus nonstructural glycoprotein NS28. *J Virol* **63**:4553-69

Azevedo, M. & Gonzalez, A. (2010). An Oral versus Intranasal Prime/Boost Regimen Using Attenuated Human Rotavirus or VP2 and VP6 Virus-Like Particles with Immunostimulating Complexes Influences Protection and Antibody-Secreting Cell

Responses to Rotavirus in a Neonatal Gnotobiotic Pig Mode. *Clin Vaccine Immunol* **17**:420–8.

Babji, S. & Kang, G. (2012). Rotavirus vaccination in developing countries. *Curr Opin Virol* **2**:1–6

Bagchi, P., Bhowmick, R., Nandi, S., Kant Nayak, M. & Chawla-Sarkar, M. (2013). Rotavirus NSP1 inhibits interferon induced non-canonical NFκB activation by interacting with TNF receptor associated factor 2. *Virol* **444**:41–4.

Ball, J. M., Tian, P., Zeng, C. Q. Y., Morris, A. P. & Estes, M. K. **(1996).** Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. *Science* **272**:101-104

Barnes, G. L., Lund, J. S., Mitchell, S. V, De Bruyn, L., Piggford, L., Smith, A. L *et al.*, (2002). Early phase II trial of human rotavirus vaccine candidate RV3. *Vaccine* **20**:2950–6.

Bernstein, D. I., Sack, D. A., Rothstein, E., Reisinger, K., Smith, V. E., O'Sullivan, D *et al*., (1999). Efficacy of live, attenuated, human rotavirus vaccine 89-12 in infants: a randomised placebo-controlled trial. *Lancet* **354**:287–90.

Bertolotti-Ciarlet, A., Ciarlet, M., Crawford, S. E., Conner, M. E. & Estes, M. K. (2003). Immunogenicity and protective efficacy of rotavirus 2/6-virus-like particles produced by a dual baculovirus expression vector and administered intramuscularly, intranasally, or orally to mice. *Vaccine* **21**:3885–900.

Bhandari, N., Sharma, P., Glass, R. I., Ray, P., Greenberg, H., Taneja, S *et al*., (2006). Safety and immunogenicity of two live attenuated human rotavirus vaccine candidates, 116E and I321, in infants: results of a randomised controlled trial. *Vaccine* **24**:5817–23.

Bhandari, N., Rongsen-chandola, T., Bavdekar, A., John, J., Antony, K., Taneja, S *et al*., (2014). Efficacy of a monovalent human-bovine (116E) rotavirus vaccine in Indian infants : a randomised , double-blind , placebo-controlled trial. *Lancet* **383**:2136–43.

Blazevica, V., Lappalainena, S., Nurminena, K., Huhtia, L. & Vesikari, T. (2011). Norovirus VLPs and rotavirus VP6 protein as combined vaccine for childhood gastroenteritis. *Vaccine* **29**(45):8126-33

Bhowmick, R., Halder, U. C., Chattopadhyay, S., Nayak, M. K. & Chawla-Sarkar, M. (2013). Rotavirus-encoded nonstructural protein 1 modulates cellular apoptotic machinery by targeting tumor suppressor protein p53. *J Virol* **87**:6840–50.

Bican, P., Cohen, J., Charpilienne, A. & Scherrer, R. (1982). Purification and characterization of bovine rotavirus cores. *J Virol* **43**:1113-7.

Bishop, R. F., Barnes, G. L., Cipriani, E. & Lund, J. S. (1983). Clinical immunity after neonatal rotavirus infection. A prospective longitudinal study in young children. *N Engl J Med* **309**:72–6.

Bishop, R., Davidson, G., Holmes, I. & Ruck, B. (1973). Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *Lancet* **302**:1281–3.

Black, R. E., Cousens, S., Johnson, H. L., Lawn, J. E., Rudan, I., Bassani, D. G *et al*., (2010). Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* **375**:1969–87.

Boyle, J. & Holmes, K. (1986). RNA-binding proteins of bovine rotavirus. *J Virol* **58**:561-8.

Bretthauer, R. K. & Castellino, F. J. (1999). Glycosylation of *Pichia pastoris*-derived proteins. *Biotechnol Appl Biochem* **30**(3):193–200.

Bryce, J., Boschi-Pinto, C., Shibuya, K. & Black, R. E. (2005). WHO estimates of the causes of death in children. *Lancet* **365**:1147–52.

Burgess-Brown, R., Millar, ID., Leroy, BP., Urquhart, JE., Fearon, IM., De Baere, E. *et al*., (2008). Biallelic mutation of BEST1 causes a distinct retinopathy in humans. Am J Hum Genet **82**(1):19-31

Chen, D. C., Beckerich, J. M., & Gaillardin, C. (1997). One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **48**:232–5.

Chirico, G., Marzollo, R., Cortinovis, S., Fonte, C. & Gasparoni, A. (2008). Anti-infective properties of human milk. *J Nutr* **138**:S1801–6.

Christy, C., Madore, H. P., Pichichero, M. E., Gala, C., Pincus, P., Vosefski, D *et al*., (1988). Field trial of rhesus rotavirus vaccine in infants. *Pediatr Infect Dis J* **7**:645-50.

Clark, H. F., Borian, F. E., Bell, L. M., Modesto, K., Gouvea, V. & Plotkin, S. A. (1988). Protective Effect of WC3 Vaccine against Rotavirus Diarrhea in Infants during a Predominantly Serotype 1 Rotavirus Season. *J Infect Dis* **158**:570–87.

Clements-Mann, M. L., Makhene, M. K., Mrukowicz, J., Wright, P. F., Hoshino, Y., Midthun, K et al., (1999). Safety and immunogenicity of live attenuated human ± bovine (UK) reassortant rotavirus vaccines with VP7-specificity for serotypes 1 , 2 , 3 or 4 in adults, children and infants. *Vaccine* **17**:2715–25.

Contreras-Gómez, A, Sánchez-Mirón, A, García-Camacho, F., Molina-Grima, E. & Chisti, Y. (2013). Protein production using the baculovirus-insect cell expression system. *Biotechnol Prog* **30**:1–18.

Crawford, S. E., Labbé, M., Cohen, J., Burroughs, M. H., Zhou, Y. J. & Estes, M. K. (1994). Characterization of virus-like particles produced by the expression of rotavirus capsid proteins in insect cells. *J Virol* **68**:5945–52.

Danchin, M., Kirkwood, C. D., Lee, K. J., Bishop, R. F., Watts, E., Justice, F. A et al., (2013). Phase I trial of RV3-BB rotavirus vaccine: A human neonatal rotavirus vaccine. *Vaccine* **31**:2610–6.

Dennehy, P. (2008). Rotavirus vaccines: an overview. *Clin Microbiol Rev* **21**:198–208.

Desselberger, U. & Huppertz, H. I. (2011). Immune responses to rotavirus infection and vaccination and associated correlates of protection. *J Infect Dis* **203**:188–95.

Desselberger, U., Manktelow, E. & Li, W. (2009). Rotaviruses and rotavirus vaccines. *Br Med Bull* **90**:37–51.

Donato, C. M., Sing, L., Boniface, K. F., Crawford, N. W., Buttery, J. P., Lyon, M et al., (2012). Identification of Strains of RotaTeq Rotavirus Vaccine in Infants With Gastroenteritis Following Routine Vaccination. *J Infect Dis* **206**:377–83.

Eichwald, C., Rodriguez, J. & Burrone, O. (2004). Characterization of rotavirus NSP2/NSP5 interactions and the dynamics of viroplasm formation. *J Gen Virol* **85**:625–34.

Ericson, B. L., Graham, D. Y., Mason, B. B., Hanssen, H. H. & Estes, M. K. (1983). Two types of glycoprotein precursors are produced by the simian rotavirus SA11. *Virol* **127**:320–32.

Estes MK. **(2001).** Rotaviruses and their replication. In *Fields Virol*, 4th edn., pp. 1747–85. Edited by H. P. Knipe DM. Philadelphia: Lippincott Williams & Wilkins.

Estes, M. & Graham, D. (1979). Rotavirus stability and inactivation. *J Gen Virol* **4**:403–9.

Fabbretti, E., Afrikanova, I., Vascotto, F. & Burrone, O. R. (1999). Two non-structural rotavirus proteins , NSP2 and NSP5 , form viroplasm-like structures *in vivo*. *J Gen Virol* **80**:333–9.

Federico, M. (2010). Virus-like particles show promise as candidates for new vaccine strategies. *Future Virol* **5**:371–4.

Fiore, L., Greenberg, H. B. & Mackow, E. R. (1991). The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin. *Virol* **181**:553–63.

Fischer, T. K., Valentiner-Branth, P., Steinsland, H., Perch, M., Santos, G., Aaby, P *et al*.**, (2002).** Protective immunity after natural rotavirus infection: a community cohort study of newborn children in Guinea-Bissau, west Africa. *J Infect Dis* **186**:593–7.

Fleming, F. E., Böhm, R., Dang, V. T., Holloway, G., Haselhorst, T., Madge, P. D *et al*.**, (2014).** Relative roles of GM1 ganglioside, N-acylneuraminic acids, and α2β1 integrin in mediumting rotavirus infection. *J Virol* **88**:4558–71.

Franco, M. A, Feng, N. & Greenberg, H. B. (1996). Molecular determinants of immunity and pathogenicity of rotavirus infection in the mouse model. *J Infect Dis* **174**:S47–50.

Franco, M. A. & Greenberg, H. B. (1997). Immunity to Rotavirus in T Cell Deficient Mice. *Virol* **179**:169–79.

Fu, C., Wang, M., Liang, J., He, T., Wang, D. & Xu, J. (2007). Effectiveness of Lanzhou lamb rotavirus vaccine against rotavirus gastroenteritis requiring hospitalization: a matched case-control study. *Vaccine* **25**:8756–61.

Fu, C., He, Q., Xu, J., Xie, H., Ding, P., Hu, W *et al*.**, (2012).** Effectiveness of the Lanzhou lamb rotavirus vaccine against gastroenteritis among children. *Vaccine* **31**:154–8.

Gellissen, G. (2000). Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* **54**:741–50.

Gerngross, T. U. (2004). Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat Biotechnol* **22**:1409–14.

Gladstone, B. & Ramani, S. (2011). Protective effect of natural rotavirus infection in an Indian birth cohort. *N Engl J Med* **365**:337–46.

Glass, R. I., Bhan, M. K., Ray, P., Bahl, R., Parashar, U. D., Greenberg, H *et al*., (2005). Development of candidate rotavirus vaccines derived from neonatal strains in India. *J Infect Dis* **192:**S30–5.

González, R. A., Torres-Vega, M. A., López, S. & Arias, C. F. (1998). *In vivo* interactions among rotavirus nonstructural proteins. *Arch Virol* **143**:981–96.

Gouvea, V. & Brantly, M. (1995). Is rotavirus a population of reassortants? *Trends Microbiol* **3**:159–62.

Goveia, M. G., Ciarlet, M., Owen, K. E. & Ranucci, C. S. (2011). Development, clinical evaluation, and post-licensure impact of RotaTeq, a pentavalent rotavirus vaccine. *Ann N Y Acad Sci* **1222**:14–8.

Graff, J. W., Mitzel, D. N., Weisend, C. M., Michelle, L., Hardy, M. E. & Flenniken, M. L. (2002). Interferon Regulatory Factor 3 Is a Cellular Partner of Rotavirus NSP1. *J Virol* **76**(18):9545-50.

Grandy, G., Medina, M., Soria, R., Terán, C. G. & Araya, M. (2010). Probiotics in the treatment of acute rotavirus diarrhoea. A randomized, double-blind, controlled trial using two different probiotic preparations in Bolivian children. *BMC Infect Dis* **10**:253-60.

Green, M.R. and Sambrook, J. (2012). Molecular cloning: a laboratory manual - 4th edition

Greenberg, H. & Flores, J. (1983). Gene coding assignments for growth restriction, neutralization and subgroup specificities of the W and DS-1 strains of human rotavirus. *J Gen Virol* **64**(2):313-20.

Greenberg, H. & McAuliffe, V. (1983). Serological analysis of the subgroup protein of rotavirus, using monoclonal antibodies. *Infect Immun* **39**:91-101.

Groome, M. J., Moon, S., Velasquez, D., Jones, S., Koen, A. & Niekerk, N. Van. **(2014).** Effect of breastfeeding on immunogenicity of oral live-attenuated human rotavirus

vaccine : a randomized trial in HIV-uninfected infants in Soweto , South Africa. *Bulletin of the World Health Organization* **92**(4):238-45.

Hanlon, P., Marsh, V., Shenton, F. & Jobe, O. (1987). Trial of an attenuated bovine rotavirus vaccine (RIT 4237) in Gambian infants. *Lancet* **1**:1342–45.

Heaton, P. M., Goveia, M. G., Miller, J. M., Offit, P. & Clark, H. F. (2005). Development of a pentavalent rotavirus vaccine against prevalent serotypes of rotavirus gastroenteritis. *J Infect Dis* **192:**S17–21.

Inoue, H., Nojima, H., & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**:23-8.

Iturriza-Gómara, M. (2001). Reassortment *in vivo*: driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. *J Virol* **75**:3696–705.

Jigami, Y. & Odani, T. (1999). Biochim. Biophys. Acta **1426**: 335–45.

Jayaram, C. & Deusen, R. Van. (1998). Characterization of a monoclonal antibody recognizing a DAG-containing epitope conserved in aphid transmissible potyviruses: evidence that the DAG motif is in a defined conformation. *Virus Res* **58**:1–11.

Jere, K. C., Mlera, L., O'Neill, H. G., Potgieter, A. C., Page, N. A., Seheri, M. L. & Van Dijk, A. A. (2011). Whole Genome Analyses of African G2 , G8 , G9 , and G12 Rotavirus Strains Using Sequence-Independent Amplification and 454 W Pyrosequencing. *J Med Virol* **83**:2018–42.

Jere, K. C., O'Neill, H. G., Potgieter, A. C. & van Dijk, A. A. (2014). Chimaeric virus-like particles derived from consensus genome sequences of human rotavirus strains co-circulating in Africa. *PLoS One* **9**:e105167.

Jiang, B., Gentsch, J. R. & Glass, R. I. (2002). The Role of Serum Antibodies in the Protection against Rotavirus Disease : An Overview **34**(10):1351-61.

Jiang, B., Gentsch, J. R. & Glass, R. I. (2008). Inactivated rotavirus vaccines: a priority for accelerated vaccine development. *Vaccine* **26**:6754–8.

Jigami, Y. & Odani, T. (1999). Mannosylphosphate transfer to yeast mannan. *Biochimica et Biophysica Acta* **1426**:335–45.

Johansson, E., Istrate, C., Charpilienne, A, Cohen, J., Hinkula, J., Poncet, D., Svensson, L. *et al.*, (2008).* Amount of maternal rotavirus-specific antibodies influence the outcome of rotavirus vaccination of newborn mice with virus-like particles. *Vaccine* **26**:778–85.

Kalica, A. R., Flores, J. & Greenberg, H. B. (1983). Identification of the rotaviral gene that codes for hemagglutination and protease-enhanced plaque formation. *Virology* **125**:194–205.

Kapikian, A. Z., Flores, J., Hoshino, Y., Glass, R. I., Midthun, K., Gorziglia, M *et al.*, (1986). Rotavirus: the major etiologic agent of severe infantile diarrhea may be controllable by a Jennerian approach to vaccination. *J Infect Dis* **153**:815–22.

Kim, I. S., Trask, S. D., Babyonyshev, M., Dormitzer, P. R. & Harrison, S. C. (2010). Effect of mutations in VP5 hydrophobic loops on rotavirus cell entry. *J Virol* **84**:6200–7.

Klebe, R. J., Harriss, J. V., Sharp, Z. D. & Douglas, M. G. (1983). A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast. *Gene* **25**:333-41.

Lanata, C. F., Black, R. E., del Aguila, R., Gil, A., Verastegui, H., Gerna, G., *et al.*, (1989). Protection of Peruvian children against rotavirus diarrhea of specific serotypes by one, two, or three doses of the RIT 4237 attenuated bovine rotavirus vaccine. *J Infect Dis* **159**:452–9.

Lappalainen, S., Pastor, A. R., Tamminen, K., López-Guerrero, V., Esquivel-Guadarrama, F., Palomares, L.A. *et al.*, (2014). Immune responses elicited against rotavirus middle layer protein VP6 inhibit viral replication *in vitro* and *in vivo*. *Hum Vaccin Immun* **10**(7):2039-47.

Lawton, J. A., Zeng, C. Q., Mukherjee, S. K., Cohen, J., Estes, M. K. & Prasad, B. V. (1997). Three-dimensional structural analysis of recombinant rotavirus-like particles with intact and amino-terminal-deleted VP2: implications for the architecture of the VP2 capsid layer. *J Virol* **71**:7353–60.

Liu, M., Mattion, N. M. & Estes, M. K. (1992). Rotavirus VP3 expressed in insect cells possesses guanylyltransferase activity. *Virol* **188**:77–84.

Ludwig, C. & Wagner, R. (2007). Virus-like particles — universal molecular toolboxes. *Curr Opin Biotechnol* **18**:537–45.

Madhi, S., Cunliffe, N. & Steele, D. (2010). Effect of human rotavirus vaccine on severe diarrhea in African infants. *N Engl J Med* **362**:289–298.

Malherbe, H. & Harwin, R. (1963). The cytopathic effects of vervet monkey viruses. *Med J* 407-415.

Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S *et al*., (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol* **82**:3204–19.

Matthijnssens, J., Ciarlet, M., McDonald, S., H Attou, I., Bányai, K., Brister, J *et al*., (2012). Uniformity of Rotavirus Strain Nomenclature Proposed by the Rotavirus Classification Working Group (RCWG). *Arch Virol* **156**:1397–1413.

Mattion, N. M., Cohen, J., Aponte, C. & Estes, M. K. (1992). Characterization of an oligomerization domain and RNA-binding properties on rotavirus nonstructural protein NS34. *Virol* **190**:68–83.

Maver, P. J., Poljak, M., Seme, K. & Kocjan, B. J. (2010). Detection and typing of low-risk human papillomavirus genotypes HPV 6, HPV 11, HPV 42, HPV 43 and HPV 44 by PCR and restriction fragment length polymorphism. *J Virol Methods* **169**:215–8.

Midthun, K., Greenberg, H. B., Hoshino, Y., Kapikian, A. Z., Wyatt, R. G. & Chanock, R. M. (1985). Reassortant rotaviruses as potential live rotavirus vaccine candidates. *J Virol* **53**:949–54.

Moon, S., Wang, Y., Shane, A. L., Nguyen, T., Ray, P., Dennehy, P *et al*., (2010). Inhibitory effect of breast milk on infectivity of live oral rotavirus vaccines. *Pediatr Infect Dis J* **29**:919–23.

Murphy, T., Gargiullo, P., Massoudi, M., Nelson, D., Jumaan, A., Okoro, C *et al*., (2001). Intussusception among Infants Given An Oral Rotavirus Vaccine. *N Engl J Med* **344**:64–72.

Mwenda, J. M., Ntoto, K. M., Abebe, A., Enweronu-Laryea, C., Amina, I., Mchomvu, J *et al.*, (2010). Burden and epidemiology of rotavirus diarrhea in selected African countries: preliminary results from the African Rotavirus Surveillance Network. *J Infect Dis* **202:**S5–S11.

Nandi, S., Chanda, S., Bagchi, P., Nayak, M. K., Bhowmick, R. & Chawla-Sarkar, M. (2014). MAVS protein is attenuated by rotavirus nonstructural protein 1. *PLoS One* **9**(3):e92126.

Newburg, D. S. (2009). Neonatal protection by an innate immune system of human milk consisting of oligosaccharides and glycans. *J Anim Sci* **87**:26–34.

Nguyen, K., Llano, M., Akari, H., Miyagi, E., Poeschla, E. M., Strebel, K *et al.*, (2004). Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression *Virol* **319**:163–75.

Padilla-Noriega, L., Paniagua, O., Guzman-Leon, S., (2002). Rotavirus protein NSP3 shuts off host cell protein synthesis. *Virol* **298**:1-7.

Palomares, L. & Ramírez, O. (2009). Challenges for the production of virus-like particles in insect cells: The case of rotavirus-like particles. *Biochem Eng J* **45**:158–67.

Parashar, U., Bresee, J., Gentsch, J. & Glass, R. (1998). Rotavirus. *Emerg Infect Dis* **4**:561–71.

Parashar, U., Burton, A., Lanata, C., Boschi-Pinto, C., Shibuya, K., Steele, D *et al.*, (2009). Global mortality associated with rotavirus disease among children in 2004. *J Infect Dis* **200**:9–15.

Parashar, U. D., Hummelman, E. G., Bresee, J. S., Miller, M. A. & Glass, R. I. (2003). Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis* **9**:565–72.

Patel, M. M., López-collada, V. R., Bulhões, M. M., Oliveira, L. H. De, Márquez, A. B., Flannery, B *et al.*, (2011). Intussusception Risk and Health Benefits of Rotavirus Vaccination in Mexico and Brazil. *N Engl J Med* **364**:2283–92.

Patel, M., Pitzer, V. & Alonso, W. (2012). Global Seasonality of Rotavirus Disease. *Pediatr Infect Dis J* **32**:134–147.

Payne, D. C., Edwards, K. M., Bowen, M. D., Keckley, E., Peters, J., Esona, M. D *et al.,* (2010). Sibling transmission of vaccine-derived rotavirus (RotaTeq) associated with rotavirus gastroenteritis. *Pediatrics* **125**:438–41.

Pesavento, J. B., Crawford, S. E., Estes, M. K. & Prasad, B. V. V. (2006). Rotavirus proteins: structure and assembly. *Curr Top Microbiol Immunol* **309**:189–219.

Phua, K. B., Lim, F. S., Lau, Y. L., Nelson, E. S., Huang, L. M., Quak, S. H *et al.,* (2009). Safety and efficacy of human rotavirus vaccine during the first 2 years of life in Asian infants: randomised, double-blind, controlled study. *Vaccine* **27**:5936–41.

Pizarro, J., Sandino, A. & Pizarro, J. (1991). Characterization of rotavirus guanylyltransferase activity associated with polypeptide VP3. *J Gen Virol* **72**:325–332.

Plenge-Bönig, A., Soto-Ramírez, N., Karmaus, W., Petersen, G., Davis, S. & Forster, J. (2010). Breastfeeding protects against acute gastroenteritis due to rotavirus in infants. *Eur J Pediatr* **169**:1471–6.

Potgieter, A. C., Page, N. A, Liebenberg, J., Wright, I. M., Landt, O. & van Dijk, A. A. (2009). Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *J Gen Virol* **90**:1423–32.

Prasad, B. V. V., Wang, G. J., Clerx, J. P. M. & Chiu, W. (1988). Three-dimensional structure of rotavirus. *J Mol Biol* **199**:269–75.

Rodríguez-Limas, W. A., Tyo, K.E., Nielsen, J., Ramírez O. T. & Palomares L. A. (2011). Molecular and process design for rotavirus-like particle production in *Saccharomyces cerevisiae*. *Microb Cell Fact*. **10**:33-43.

Rodríguez-limas, W. A., Ruth, A., Esquivel-soto, E., Esquivel-guadarrama, F., Ramírez, O. T. & Palomares, L. A. (2014). Immunogenicity and protective efficacy of yeast extracts containing rotavirus-like particles: A potential veterinary vaccine. *Vaccine* **32**(24):2794-8.

Romanos, M., Scorer, C. & Clare, J. (1992). Foreign gene expression in yeast: a review. *Yeast* **8**:423–488.

Romanowski, B. (2011). Long term protection against cervical infection with the human papillomavirus: Review of currently available vaccines. *Hum Vaccin* **7**:161–169.

Ruiz-Palacios, G. M., Pérez-Schael, I., Velázquez, F. R., Abate, H., Breuer, T., Clemens, S. C et al., (2006). Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *N Engl J Med* **354**:11–22.

Santos, N. & Hoshino, Y. (2005). Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev Med Virol* **15**:29–56.

Santosham, M., Keenan, E. M., Tulloch, J., Broun, D. & Glass, R. (1997). Oral Rehydration Therapy for Diarrhea: An Example of Reverse Transfer of Technology. *Pediatrics* **100**(5):E10.

Sarker, S. A., Casswall, T. H., Mahalanabis, D. M., Alam, N. H., Albert, M. J., Brüssow, H et al., (1998). Successful treatment of rotavirus diarrhea in children with immonoglobulin from immunized bovine colostrum. *Pediatr Infect Dis J* **17**:1149–54.

Seheri, M., Nemarude, L., Peenze, I., Netshifhefhe, L., Nyaga, M. M., Ngobeni, H. G et al., (2014). Update of rotavirus strains circulating in Africa from 2007 through 2011. *Pediatr Infect Dis J* **33**(1):76–84.

Sen, A., Pruijssers, A. J., Dermody, T. S., García-Sastre, A. & Greenberg, H. B. (2011). The early interferon response to rotavirus is regulated by PKR and depends on MAVS/IPS-1, RIG-I, MDA-5, and IRF3. *J Virol* **85**:3717–32.

Settembre, E. C., Chen, J. Z., Dormitzer, P. R., Grigorieff, N. & Harrison, S. C. (2011). Atomic model of an infectious rotavirus particle. *EMBO J* **30**:408–16.

Shu, Q., Qu, F. & Gill, H. S. (2001). Probiotic treatment using *Bifidobacterium lactis* HN019 reduces weanling diarrhea associated with rotavirus and *Escherichia coli* infection in a piglet model. *J Pediatr Gastroenterol Nutr* **33**:171–7.

Silvestri, L. S., Taraporewala, Z. F. & Patton, J. T. (2004). Rotavirus Replication : Plus-Sense Templates for Double-Stranded RNA Synthesis Are Made in Viroplasms. *J Virol* **78**:7763–74.

Sørensen, H. P. (2010). Towards universal systems for recombinant gene expression. *Microb Cell Fact* **9**:27-31.

Steele, D., Peenze, I., de Beer, M., Pager, C., Yeats, J., Potgieter, N *et al.*, **(2003).** Anticipating rotavirus vaccines: epidemiology and surveillance of rotavirus in South Africa. *Vaccine* **21**:354–360.

Steinborn, G., Böer, E., Scholz, A., Tag, K., Kunze, G. & Gellissen, G. (2006). Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic *Arxula adeninivorans*, methylotrophic *Hansenula polymorpha* and other yeasts. *Microb. Cell Fact* **5**:33.

Stoll, B. J., Glass, R. I., Huq, M. I., Khan, M. U. & Holt, J. E. (1982). Outside Europe Surveillance of patients attending a diarrhoeal disease hospital in Bangladesh. *J Infect* **285**(6349):1185-8.

Svensson, L., Sheshberadaran, H., Vesikari, T., Norrby, E. & Wadell, G. (1987). Immune response to rotavirus polypeptides after vaccination with heterologous rotavirus vaccines (RIT 4237, RRV-1). *J Gen Virol* **68:**1993–9.

Tanaka, G., Faruque, A. S. G., Luby, S. P., Malek, M. A., Glass, R. I. & Parashar, U. D. (2007). Deaths from rotavirus disease in Bangladeshi children: estimates from hospital-based surveillance. *Pediatr Infect Dis J* **26**:1014–8.

Taraporewala, Z., Chen, D. & Patton, J. T. (1999). Multimers Formed by the Rotavirus Nonstructural Protein NSP2 Bind to RNA and Have Nucleoside Triphosphatase Activity J. *Virol* **73**(12):9934-43.

Tate, J. E., Burton, A. H., Boschi-Pinto, C., Steele, a D., Duque, J. & Parashar, U. D. (2012). 2008 Estimate of Worldwide Rotavirus-Associated Mortality in Children Younger Than 5 Years Before the Introduction of Universal Rotavirus Vaccination Programmes: a Systematic Review and Meta-Analysis. *Lancet Infect Dis* **12**, 136–41.

Taylor J A, O'Brien J A, & Yeager M.(1996). The cytoplasmic tail of NSP4, the endoplasmic reticulum-localised non-structural glycoprotein of rotavirus, contains distinct virus binding and coiled coil domains. *EMBO J*. **15**:4469–76.

Theron, C. W., Labuschagn, M., Gudiminchi, R., Albertyn, J. & Smit M. S. (2014). A broad-range yeast expression system reveals *Arxula adeninivorans* expressing a fungal self-

sufficient cytochrome P450 monooxygenase as an excellent whole-cell biocatalyst. FEMS Yeast Res **14**:556–66

Trojnar, E., Sachsenröder, J., Twardziok, S., Reetz, J., Otto, P. H. & Johne, R. (2013). Identification of an avian group A rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses. *J Gen Virol* **94**:136–42.

Udem, S. & Warner, J. (1972). Ribosomal RNA synthesis in *Saccharomyces cerevisiae*. *J Mol Biol* **65**:227–42.

Uhnoo, I., Olding-Stenkvist, E. & Kreuger, A. (1986). Clinical features of acute gastroenteritis associated with rotavirus, enteric adenoviruses, and bacteria. *Arch Dis Child* **61**:732-38

Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G. & Hall, B. D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**:347–50.

Vega, C. G., Bok, M., Vlasova, A. N., Chattha, K. S., Gómez-Sebastián, S., Nuñez, C *et al*., (2013). Recombinant monovalent llama-derived antibody fragments (VHH) to rotavirus VP6 protect neonatal gnotobiotic piglets against human rotavirus-induced diarrhea. *PLoS Pathog* **9**:e1003334.

Velázquez, F. R., Matson, D. O., Guerrero, M. L., Shults, J., Calva, J. J., Morrow, A. L *et al*., (2000). Serum antibody as a marker of protection against natural rotavirus infection and disease. *J Infect Dis* **182**:1602–9.

Velázquez, F., Matson, D., Calva, J., Guerrero, M., Morrow, A., Carter-Campbell, S *et al*., (1996). Rotavirus infection in infants as protection against subsequent infections. *N Engl J Med* **335**:1022–8.

Vesikari, T., Isolauri, E., D'Hondt, E., Delem, A., André, F. E. & Zissis, G. (1984). Protection of infants against rotavirus diarrhoea by RIT 4237 attenuated bovine rotavirus strain vaccine. *Lancet* **1**:977–81.

Vesikari, T., Karvonen, A., Prymula, R., Schuster, V., Tejedor, J., Cohen, R *et al*., (2007). Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, double-blind controlled study. *Lancet* **370**:1757–63.

Vesikari, T., Matson, D. O., Dennehy, P., Van Damme, P., Santosham, M., Rodriguez, Z *et al.*, **(2006).** Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. *N Engl J Med* **354**:23–33.

Victora, C. G., Bryce, J., Fontaine, O. & Monasch, R. (2000). Reducing deaths from diarrhoea through oral rehydration therapy. *Bull World Health Organ* **78**:1246–55.

Wang, Y., Azevedo, M., Saif, L. J., Gentsch, J. R., Glass, R. I. & Jiang, B. (2010). Inactivated rotavirus vaccine induces protective immunity in gnotobiotic piglets. *Vaccine* **28**:5432–6.

World Health Organization (2000). Report of the meeting on future directions for rotavirus vaccine research in developing countries. *Geneva.*

Zaman, K., Dang, D. A., Victor, J. C., Shin, S., Yunus, M., Dallas, M. J *et al.*, **(2010).** Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. *Lancet* **376**:615–23.

Zanardi, L., Haber, P., Mootrey, G., Niu, M. & Wharton, M. (2001). Intussusception among recipients of rotavirus vaccine: reports to the vaccine adverse event reporting system. *Pediatrics* **107**:E97.

Zeng, C. Q., Wentz, M. J., Cohen, J. & Estes, M. K. (1996). Characterization and Replicase Activity of Double-Layered and Single-Layered Rotavirus-Like Particles Expressed from Baculovirus Recombinants. *J Virol* **70**:2736–42.

CHAPTER 6

Summary

In this study, sequences encoding VP2 and VP6 rotavirus structural proteins from rotavirus strain RVA/Human-wt/ZAF/GR10924/1999/G9P[6] were used in construction of wide-range yeast expression vectors containing the ORFs encoding rotavirus structural proteins VP2 and VP6. VP2 and VP6 sequences were codon-optimized for expression in yeast strains *K. lactis, A. adeninivorans* and *Pichia pastoris/Hansenula polymorpha*. Wide-range yeast expression vectors containing either VP6 or VP2 yeast optimized ORFs were constructed for expression of single proteins in different yeast strains. Dual vectors containing both VP6 and VP2 yeast optimized ORFs were constructed to allow simultaneous expression of proteins in different yeast strains and to enable the formation of double-layered rotavirus-like particles.

A total of eight yeast strains namely: *Kluyveromyces marxianus, Kluyveromyces lactis, Debaryomyces hansenii, Yarrowia lipolytica, Hansenula polymorpha, Pichia pastoris Candida deformans* and *Arxula adeninivorans* were selected for screening. *Saccharomyces cerevisiae* was included as a positive control as triple-layered rotavirus virus-like particles (tlRLPs) have been successfully produced in this yeast before. All nine yeast strains were successfully transformed with pKM177 vectors containing yeast codon-optimized VP6 ORFs. However, only six yeast strains indicated positive VP6 ORF integration. The other three yeast strains indicated no VP6 integration but hygromycin B integration suggesting that the vectors successfully integrated in these yeast strains but VP6 ORF was lost. Integration of both rotavirus VP2/6 ORFs was evident in all yeast strains transformed with the dual expression vectors.

A control to monitor rotavirus VP6 protein expression in yeast was successfully prepared in bacterial cells. The positive control indicated specific reaction when polyclonal rotavirus antibody raised against Nebraska calf diarrhoea rotavirus strain. Rotavirus VP6 expression in yeast strains shown no reactions for *K. lactis* and *A. adeninivorans* codon-optimized VP6 ORF in all six yeast strains that tested positive for intergration of the VP6 ORF. All yeast strains except *K. marxianus* showed possible reaction for VP6 expression for *P. pastoris/H. polymorpha* codon-optimized VP6 ORF although the identity of VP6 should be confirmed.

Keywords: Rotavirus, Vaccine, Double-layer particles, Western blot analysis, Yeast, Expression, Vector, Open reading Frame, rotavirus VP2, rotavirus VP6.

CHAPTER 7

Opsomming

In hierdie studie, basispaaropeenvolgings wat kodeer vir VP2 en VP6 rotavirus strukturele proteïene van rotavirus stam RVA/Human-wt/ZAF/GR10924/1999/G9P [6] is gebruik in die konstruksie van 'n wye verskeidenheid gis uitdrukkingsvektore met rotavirus strukturele proteïene VP2 en VP6 oopleesrame. VP2 en VP6 basispaaropeenvolgings is kodon geoptimaliseer vir uitdrukking in gis stamme *K. lactis*, *A. adeninivorans* en *P. pastoris/H. polymorpha*. Wye verskeidenheid gis uitdrukkingsvektore wat geoptimaliseerde oopleesrame van VP6 of VP2 bevat, is gebou vir uitdrukking van enkel proteïene in verskillende gisstamme. Vektore wat beide VP6 en VP2 geoptimaliseerde oopleesrame bevat is gebou vir gelyktydige uitdrukking van proteïene in verskillende gisstamme en om die vorming van dubbellaag rotavirus-agtige partikels toe te laat.

'n Totaal van agt gisstamme naamlik: *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Pichia pastoris*, *Candida deformans* en *Arxula adeninivorans* is gekies vir sifting. *Saccharomyces cerevisiae* is ingesluit as 'n positiewe kontrole aangesien trippellaag rotavirus-agtige partikels (tlRLPs) voorheen suksesvol in hierdie gis geproduseer is. Al nege gisstamme is suksesvol getransformeer met pKM177 vektore wat gis geoptimaliseerde kodon VP6 oopleesrame bevat. Daar was egter slegs ses gisstamme waarin positiewe VP6 ORF integrasie gevind is. Die oorblywende drie gisstamme het higromisien B integrasie getoon, wat daarop dui dat die vektore suksesvol geïntegreer het in hierdie gisstamme, maar dat VP6 oopleesraam verloor is. Integrasie van beide rotavirus VP2/6 oopleesrame was suksesvol vir al die gis stamme wat met die dubbele uitdrukkingsvektore getransformeer is.

Rotavirus VP6 proteïenuitdrukking is suksesvol uitgevoer in bakteriële selle. Die positiewe kontrole het 'n spesifieke reaksie getoon met poliklonale rotavirus teenliggaampies wat teen Nebraska calf diarrhoea rotavirus opgewek is, maar geen reaksie is waargeneem wanneer groep-spesifieke muis monoklonale teenliggaampies gebruik is nie. Rotavirus VP6 uitdrukking in gisstamme het geen reaksies getoon vir *K. lactis* en *A. adeninivorans* geoptimiseerde VP6 oopleesrame nie in al ses gisstamme wat positief getoets het vir integrasie van die VP6 oopleesrame. Alle gis stamme behalwe *K. marxianus* het moontlike reaksie vir VP6 uitdrukking vir *P.pastoris / H. polymorpha* geoptimiseerde VP6 oopleesraam, alhoewel die identiteit van VP6 nog bevestig moet word.

APPENDIX A ClustalX Multiple alignments of sequences obtained using Sanger sequencing with *in silico* clones

```
Sequenced     ATCCCCCCCTTTCTTTTTCTCCTGAAATTGGTTGTCAACTCACACCCGAAATGCTCGAGC
GeneScript    --------------------------------------------------------CTCGAGC
                                                                    *******

Sequenced     ATGGCTTACAGAAAGAGAGGAGCAAGAAGAGAAGCAAACTTAAATAACAACGATAGAATG
GeneScript    ATGGCTTACAGAAAGAGAGGAGCAAGAAGAGAAGCAAACTTAAATAACAACGATAGAATG
              ************************************************************

Sequenced     CAAGAAAAGATTGATGAAAAACAAGATAGTAATAAGATCCAATTGAGTGATAAGGTTTTG
GeneScript    CAAGAAAAGATTGATGAAAAACAAGATAGTAATAAGATCCAATTGAGTGATAAGGTTTTG
              ************************************************************

Sequenced     TCCAAAAAGGAAGAAATCGTGACTGATTCCCATGAAGAAGTTAAAGTGACTGATGAATTG
GeneScript    TCCAAAAAGGAAGAAATCGTGACTGATTCCCATGAAGAAGTTAAAGTGACTGATGAATTG
              ************************************************************

Sequenced     AAAAAGTCTACAAAGGAAGAATCAAAGCAATTGTTGGAAGTTTTGAAGACAAAGGAAGAA
GeneScript    AAAAAGTCTACAAAGGAAGAATCAAAGCAATTGTTGGAAGTTTTGAAGACAAAGGAAGAA
              ************************************************************

Sequenced     CATCAAAAGGAAATCCAATACGAAATCTTGCAAAAGACTATACCAACCTTCGAACCTAAG
GeneScript    CATCAAAAGGAAATCCAATACGAAATCTTGCAAAAGACTATACCAACCTTCGAACCTAAG
              ************************************************************

Sequenced     GAAACAATCTTGAGAAAGTTGGAAGATATTCAACCAGAATTGGCTAAGAAACAAACCAAG
GeneScript    GAAACAATCTTGAGAAAGTTGGAAGATATTCAACCAGAATTGGCTAAGAAACAAACCAAG
              ************************************************************

Sequenced     TTATTCAGAATCTTCGAACCAAAGCAATTGCCTATCTACAGAGCTAATGGTGAAAGAGAA
GeneScript    TTATTCAGAATCTTCGAACCAAAGCAATTGCCTATCTACAGAGCTAATGGTGAAAGAGAA
              ************************************************************

Sequenced     TTGAGAAACAGATGGTACTGGAAGTTGAAAAAGGATACCTTACCAGATGGAGATTATGAT
GeneScript    TTGAGAAACAGATGGTACTGGAAGTTGAAAAAGGATACCTTACCAGATGGAGATTATGAT
              ************************************************************

Sequenced     GTCAGAGAATACTTTTTGAATTTGTACGATCAAGTATTGACAGAAATGCCTGATTACTTG
GeneScript    GTCAGAGAATACTTTTTGAATTTGTACGATCAAGTATTGACAGAAATGCCTGATTACTTG
              ************************************************************

Sequenced     TTGTTGAAGGATATGGCAGTTGAAAACAAGAATTCTAGAGATGCTGGTAAAGTTGTGGAT
GeneScript    TTGTTGAAGGATATGGCAGTTGAAAACAAGAATTCTAGAGATGCTGGTAAAGTTGTGGAT
              ************************************************************

Sequenced     TCTGAAACTGCATCAATTTGTGATGCTATCTTCCAAGATGAAGAAACAGAAGGAGCAGTT
GeneScript    TCTGAAACTGCATCAATTTGTGATGCTATCTTCCAAGATGAAGAAACAGAAGGAGCAGTT
              ************************************************************

Sequenced     AGAAGATTCATTGCTGAAATGAGACAAAGAGTGCAAGCAGATAGAAACGTCGTAAACTAC
GeneScript    AGAAGATTCATTGCTGAAATGAGACAAAGAGTGCAAGCAGATAGAAACGTCGTAAACTAC
              ************************************************************

Sequenced     CCATCAATCTTGCATCCTATCGATTACGCTTTCAATGAATATTTCTTGCAACATCAATTG
GeneScript    CCATCAATCTTGCATCCTATCGATTACGCTTTCAATGAATATTTCTTGCAACATCAATTG
              ************************************************************

Sequenced     GTTGAACCATTGAACAACGATATAATCTTCAACTACATCCCTGAAAGAATCAGAAACGAT
GeneScript    GTTGAACCATTGAACAACGATATAATCTTCAACTACATCCCTGAAAGAATCAGAAACGAT
              ************************************************************

Sequenced     GTGAACTACATATTGAATATGGATAGAAACTTACCAAGTACTGCAAGATACATCAGACCT
```

```
GeneScript     GTGAACTACATATTGAATATGGATAGAAACTTACCAAGTACTGCAAGATACATCAGACCT
               ************************************************************

Sequenced      AATTTGTTGCAAGATAGATTGAATTTGCATGATAACTTCGAATCATTGTGGGATACAATC
GeneScript     AATTTGTTGCAAGATAGATTGAATTTGCATGATAACTTCGAATCATTGTGGGATACAATC
               ************************************************************

Sequenced      ACTACAAGTAACTACATATTGGCTAGATCCGTTGTGCCAGATTTGAAGGAATTAGTTAGT
GeneScript     ACTACAAGTAACTACATATTGGCTAGATCCGTTGTGCCAGATTTGAAGGAATTAGTTAGT
               ************************************************************

Sequenced      ACTGAAGCACAAATACAAAAGATGTCCCAAGATTTGCAATTGGAAGCTTTGACAATCCAA
GeneScript     ACTGAAGCACAAATACAAAAGATGTCCCAAGATTTGCAATTGGAAGCTTTGACAATCCAA
               ************************************************************

Sequenced      AGTGAAACCCAATTCTTAACTGGTATTAATTCCCAAGCTGCAAACGATTGTTTCAAGACA
GeneScript     AGTGAAACCCAATTCTTAACTGGTATTAATTCCCAAGCTGCAAACGATTGTTTCAAGACA
               ************************************************************

Sequenced      TTGATAGCTGCAATGTTAAGTCAAAGAACCATGTCCTTGGATTTCGTTACCACTAACTAC
GeneScript     TTGATAGCTGCAATGTTAAGTCAAAGAACCATGTCCTTGGATTTCGTTACCACTAACTAC
               ************************************************************

Sequenced      ATGTCTTTGATCTCAGGAATGTGGTTATTGACTGTCGTACCAAACGATATGTTCATCAGA
GeneScript     ATGTCTTTGATCTCAGGAATGTGGTTATTGACTGTCGTACCAAACGATATGTTCATCAGA
               ************************************************************

Sequenced      GAATCTTTGGTTGCTTGTCAATTGGCAATCGTGAACACTATCATCTATCCAGCTTTCGGT
GeneScript     GAATCTTTGGTTGCTTGTCAATTGGCAATCGTGAACACTATCATCTATCCAGCTTTCGGT
               ************************************************************

Sequenced      ATGCAAAGAATGCATTACAGAAACGGAGATCCACAAACACCTTTCCAAATAGCAGAACAA
GeneScript     ATGCAAAGAATGCATTACAGAAACGGAGATCCACAAACACCTTTCCAAATAGCAGAACAA
               ************************************************************

Sequenced      CAAATCCAAAAACTTCCAAGTCGCTAACTGGTTGCATTTCGTAAACAACAACCAATTCAGA
GeneScript     CAAATCCAAAAACTTCCAAGTCGCTAACTGGTTGCATTTCGTAAACAACAACCAATTCAGA
               ************************************************************

Sequenced      CAAGCAGTTATTGATGGTGTGTTGAACCAAGTCTTGAACGATAACATCAGAAACGGACAT
GeneScript     CAAGCAGTTATTGATGGTGTGTTGAACCAAGTCTTGAACGATAACATCAGAAACGGACAT
               ************************************************************

Sequenced      GTTATAAACCAATTGATGGAAGCTTTGATGCAATTATCTAGACAACAATTCCCAACTATG
GeneScript     GTTATAAACCAATTGATGGAAGCTTTGATGCAATTATCTAGACAACAATTCCCAACTATG
               ************************************************************

Sequenced      CCTATCGATTACAAGAGATCTATCCAAAGAGGTATCTTGTTGTTGTCAAACAGATTGGGA
GeneScript     CCTATCGATTACAAGAGATCTATCCAAAGAGGTATCTTGTTGTTGTCAAACAGATTGGGA
               ************************************************************

Sequenced      CAATTGGTTGATTTGACTAGATTGTTGGCATACAACTACGAAACATTGATGGCTTGTATT
GeneScript     CAATTGGTTGATTTGACTAGATTGTTGGCATACAACTACGAAACATTGATGGCTTGTATT
               ************************************************************

Sequenced      ACCATGAACATGCAACATGTGCAAACATTGACAACCGAAAAGTTGCAATTGACTAGTGTC
GeneScript     ACCATGAACATGCAACATGTGCAAACATTGACAACCGAAAAGTTGCAATTGACTAGTGTC
               ************************************************************

Sequenced      ACATCCTTATGTATGTTGATTGGTAATGCTACCGTAATCCCATCTCCTCAAACTTTGTTC
GeneScript     ACATCCTTATGTATGTTGATTGGTAATGCTACCGTAATCCCATCTCCTCAAACTTTGTTC
               ************************************************************

Sequenced      CATTACTACAACGTCAACGTAAACTTCCATTCAAATTACAACGAAAGAATTAACGATGCT
GeneScript     CATTACTACAACGTCAACGTAAACTTCCATTCAAATTACAACGAAAGAATTAACGATGCT
               ************************************************************

Sequenced      GTCGCAATCATCACTGCTGCAAACAGATTGAACTTATACCAAAAGAAAATGAAGGCTATC
```

```
GeneScript     GTCGCAATCATCACTGCTGCAAACAGATTGAACTTATACCAAAAGAAAATGAAGGCTATC
               ************************************************************

Sequenced      GTTGAAGATTTCTTAAAGAGATTGTACATCTTTGATGTTTCTAGAGTGCCAGATGATCAA
GeneScript     GTTGAAGATTTCTTAAAGAGATTGTACATCTTTGATGTTTCTAGAGTGCCAGATGATCAA
               ************************************************************

Sequenced      ATGTACAGATTGAGAGATAGATTGAGATTGTTGCCTGTTGAAATCAGAAGATTGGATATA
GeneScript     ATGTACAGATTGAGAGATAGATTGAGATTGTTGCCTGTTGAAATCAGAAGATTGGATATA
               ************************************************************

Sequenced      TTCAATTTGATCTTGATGAACATGGATCAAATCGAAAGAGCATCAGATAAGATAGCTCAA
GeneScript     TTCAATTTGATCTTGATGAACATGGATCAAATCGAAAGAGCATCAGATAAGATAGCTCAA
               ************************************************************

Sequenced      GGTGTTATTATCGCATACAGAGATATGCATTTGGAAAGAGATGAAATGTATGGTTACGTG
GeneScript     GGTGTTATTATCGCATACAGAGATATGCATTTGGAAAGAGATGAAATGTATGGTTACGTG
               ************************************************************

Sequenced      AATATTGCTAGAAACTTGGAAGGATTCCAACAAATTAATTTGGAAGAATTGATGAGATCT
GeneScript     AATATTGCTAGAAACTTGGAAGGATTCCAACAAATTAATTTGGAAGAATTGATGAGATCT
               ************************************************************

Sequenced      GGAGATTACGCTCAAATCACCAACATGTTGTTGAACAACCAACCAGTCGCTTTGGTAGGA
GeneScript     GGAGATTACGCTCAAATCACCAACATGTTGTTGAACAACCAACCAGTCGCTTTGGTAGGA
               ************************************************************

Sequenced      GCATTACCTTTCATCACTGATTCTTCAGTTATATCATTGATCGCTAAGTTAGATGCAACC
GeneScript     GCATTACCTTTCATCACTGATTCTTCAGTTATATCATTGATCGCTAAGTTAGATGCAACC
               ************************************************************

Sequenced      GTGTTTGCTCAAATAGTCAAGTTGAGAAAGGTAGATACTTTGAAGCCAATCTTGTACAAG
GeneScript     GTGTTTGCTCAAATAGTCAAGTTGAGAAAGGTAGATACTTTGAAGCCAATCTTGTACAAG
               ************************************************************

Sequenced      ATAAATAGTGATTCCAACGATTTCTACTTGGTCGCTAATTATGATTGGGTACCAACTTCT
GeneScript     ATAAATAGTGATTCCAACGATTTCTACTTGGTCGCTAATTATGATTGGGTACCAACTTCT
               ************************************************************

Sequenced      ACTACAAAAGTATACAAGCAAGTTCCTCAACAATTCGATTTCAGAAACTCTATGCATATG
GeneScript     ACTACAAAAGTATACAAGCAAGTTCCTCAACAATTCGATTTCAGAAACTCTATGCATATG
               ************************************************************

Sequenced      TTGACATCAAACTTAACCTTCACTGTCTATTCTGATTTGTTAGCATTTGTTTCAGCTGAT
GeneScript     TTGACATCAAACTTAACCTTCACTGTCTATTCTGATTTGTTAGCATTTGTTTCAGCTGAT
               ************************************************************

Sequenced      ACCGTGGAACCAATTAATGCTGTAGCATTCGATAACATGAGAATTATGAACGAATTGTAA
GeneScript     ACCGTGGAACCAATTAATGCTGTAGCATTCGATAACATGAGAATTATGAACGAATTGTAA
               ************************************************************

Sequenced      CCTAGGTGATCTGATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAAAAAAAA
GeneScript     CCTAGG------------------------------------------------------
               ******
```

Figure 1: CLUSTAXL 2.1 multiple sequence alignment comparing sequenced *K. lactis* codon-optimized VP2 ORF to *in silico* sequences from GenScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *XhoI* and *XmaJI* restriction enzymes recognition site = underlined.

```
Sequenced   TTTCCCTCTTTCTTTTTTCTCTCNTCTCCTTGTCAACTCACACCCGAAATGCTCGACCAT
GeneScript  -----------------------------------------------GTCGACCAT
                                                           *******

Sequenced   GGCCTACCGAAAGCGAGGTGCCCGACGAGAGGCTAACTTGAATAATAATGACCGAATGCA
GeneScript  GGCCTACCGAAAGCGAGGTGCCCGACGAGAGGCTAACTTGAATAATAATGACCGAATGCA
            ************************************************************

Sequenced   GGAGAAGATTGACGAGAAGCAGGATTCTAACAAGATTCAGCTCTCCGATAAGGTTTTGTC
GeneScript  GGAGAAGATTGACGAGAAGCAGGATTCTAACAAGATTCAGCTCTCCGATAAGGTTTTGTC
            ************************************************************

Sequenced   GAAGAAGGAGGAGATCGTGACTGATTCCCACGAGGAGGTTAAGGTTACTGACGAGCTGAA
GeneScript  GAAGAAGGAGGAGATCGTGACTGATTCCCACGAGGAGGTTAAGGTTACTGACGAGCTGAA
            ************************************************************

Sequenced   GAAGTCTACCAAGGAGGAGTCCAAGCAGCTGCTTGAGGTGCTTAAGACCAAGGAGGAGCA
GeneScript  GAAGTCTACCAAGGAGGAGTCCAAGCAGCTGCTTGAGGTGCTTAAGACCAAGGAGGAGCA
            ************************************************************

Sequenced   TCAGAAGGAGATTCAGTACGAGATCCTCCAGAAGACTATTCCTACCTTCGAGCCCAAGGA
GeneScript  TCAGAAGGAGATTCAGTACGAGATCCTCCAGAAGACTATTCCTACCTTCGAGCCCAAGGA
            ************************************************************

Sequenced   GACTATTCTGCGAAAGCTTGAGGACATCCAGCCTGAGCTGGCCAAGAAGCAGACCAAGCT
GeneScript  GACTATTCTGCGAAAGCTTGAGGACATCCAGCCTGAGCTGGCCAAGAAGCAGACCAAGCT
            ************************************************************

Sequenced   TTTCCGAATTTTTGAGCCTAAGCAGCTCCCCATCTATCGAGCTAACGGAGAGCGAGAGTT
GeneScript  TTTCCGAATTTTTGAGCCTAAGCAGCTCCCCATCTATCGAGCTAACGGAGAGCGAGAGTT
            ************************************************************

Sequenced   GCGAAATCGATGGTACTGGAAGCTGAAGAAGGATACTCTTCCTGACGGAGACTACGACGT
GeneScript  GCGAAATCGATGGTACTGGAAGCTGAAGAAGGATACTCTTCCTGACGGAGACTACGACGT
            ************************************************************

Sequenced   CCGAGAGTATTTTCTGAACCTTTACGACCAGGTTCTGACCGAGATGCCCGATTATCTCTT
GeneScript  CCGAGAGTATTTTCTGAACCTTTACGACCAGGTTCTGACCGAGATGCCCGATTATCTCTT
            ************************************************************

Sequenced   GCTGAAGGACATGGCCGTCGAGAATAAGAACTCTCGAGATGCTGGAAAGGTCGTGGACTC
GeneScript  GCTGAAGGACATGGCCGTCGAGAATAAGAACTCTCGAGATGCTGGAAAGGTCGTGGACTC
            ************************************************************

Sequenced   GGAGACTGCCTCTATTTGCGATGCTATCTTCCAGGACGAGGAGACCGAGGGTGCCGTGCG
GeneScript  GGAGACTGCCTCTATTTGCGATGCTATCTTCCAGGACGAGGAGACCGAGGGTGCCGTGCG
            ************************************************************

Sequenced   ACGATTCATTGCTGAGATGCGACAGCGAGTCCAGGCCGATCGAAACGTTGTGAATTACCC
GeneScript  ACGATTCATTGCTGAGATGCGACAGCGAGTCCAGGCCGATCGAAACGTTGTGAATTACCC
            ************************************************************

Sequenced   TTCCATTCTGCACCCCATCGACTATGCTTTCAACGAGTACTTTTTGCAGCATCAGCTGGT
GeneScript  TTCCATTCTGCACCCCATCGACTATGCTTTCAACGAGTACTTTTTGCAGCATCAGCTGGT
            ************************************************************

Sequenced   TGAGCCTCTTAACAACGATATCATCTTCAACTACATTCCCGAGCGAATCCGAAACGATGT
GeneScript  TGAGCCTCTTAACAACGATATCATCTTCAACTACATTCCCGAGCGAATCCGAAACGATGT
            ************************************************************

Sequenced   GAATTACATTCTCAACATGGACCGAAATTTGCCTTCTACTGCCCGATATATCCGACCCAA
GeneScript  GAATTACATTCTCAACATGGACCGAAATTTGCCTTCTACTGCCCGATATATCCGACCCAA
            ************************************************************

Sequenced   CCTTCTCCAGGATCGACTCAACTTGCACGACAATTTTGAGTCCCTGTGGGATACCATTAC
GeneScript  CCTTCTCCAGGATCGACTCAACTTGCACGACAATTTTGAGTCCCTGTGGGATACCATTAC
            ************************************************************
```

101

```
Sequenced    TACCTCTAACTACATCCTTGCTCGATCCGTCGTTCCTGACCTCAAGGAGTTGGTTTCCAC
GeneScript   TACCTCTAACTACATCCTTGCTCGATCCGTCGTTCCTGACCTCAAGGAGTTGGTTTCCAC
             ************************************************************

Sequenced    TGAGGCCCAGATTCAGAAGATGTCGCAGGACCTGCAGCTTGAGGCTCTCACCATTCAGTC
GeneScript   TGAGGCCCAGATTCAGAAGATGTCGCAGGACCTGCAGCTTGAGGCTCTCACCATTCAGTC
             ************************************************************

Sequenced    GGAGACTCAGTTCTTGACCGGAATCAACTCTCAGGCTGCCAATGATTGCTTTAAGACTCT
GeneScript   GGAGACTCAGTTCTTGACCGGAATCAACTCTCAGGCTGCCAATGATTGCTTTAAGACTCT
             ************************************************************

Sequenced    CATTGCTGCCATGTTGTCGCAGCGAACCATGTCTCTGGACTTCGTGACTACCAACTACAT
GeneScript   CATTGCTGCCATGTTGTCGCAGCGAACCATGTCTCTGGACTTCGTGACTACCAACTACAT
             ************************************************************

Sequenced    GTCCCTTATTTCGGGAATGTGGTTGCTGACTGTGGTCCCTAACGATATGTTCATTCGAGA
GeneScript   GTCCCTTATTTCGGGAATGTGGTTGCTGACTGTGGTCCCTAACGATATGTTCATTCGAGA
             ************************************************************

Sequenced    GTCTCTGGTCGCTTGTCAGCTTGCCATCGTTAACACTATTATCTATCCTGCTTTTGGAAT
GeneScript   GTCTCTGGTCGCTTGTCAGCTTGCCATCGTTAACACTATTATCTATCCTGCTTTTGGAAT
             ************************************************************

Sequenced    GCAGCGAATGCACTACCGAAACGGTGACCCTCAGACCCCCTTCCAGATTGCCGAGCAGCA
GeneScript   GCAGCGAATGCACTACCGAAACGGTGACCCTCAGACCCCCTTCCAGATTGCCGAGCAGCA
             ************************************************************

Sequenced    GATCCAGAACTTTCAGGTGGCTAATTGGCTTCATTTCGTCAACAATAACCAGTTTCGACA
GeneScript   GATCCAGAACTTTCAGGTGGCTAATTGGCTTCATTTCGTCAACAATAACCAGTTTCGACA
             ************************************************************

Sequenced    GGCCGTCATTGATGGAGTTCTCAACCAGGTGTTGAATGACAACATTCGAAACGGTCACGT
GeneScript   GGCCGTCATTGATGGAGTTCTCAACCAGGTGTTGAATGACAACATTCGAAACGGTCACGT
             ************************************************************

Sequenced    GATCAATCAGCTGATGGAGGCTCTGATGCAGCTTTCTCGACAGCAGTTTCCTACTATGCC
GeneScript   GATCAATCAGCTGATGGAGGCTCTGATGCAGCTTTCTCGACAGCAGTTTCCTACTATGCC
             ************************************************************

Sequenced    CATTGATTACAAGCGATCTATTCAGCGAGGAATCCTTCTCTTGTCCAACCGACTCGGACA
GeneScript   CATTGATTACAAGCGATCTATTCAGCGAGGAATCCTTCTCTTGTCCAACCGACTCGGACA
             ************************************************************

Sequenced    GTTGGTGGACCTGACTCGACTGCTTGCCTACAACTATGAGACTCTCATGGCTTGCATTAC
GeneScript   GTTGGTGGACCTGACTCGACTGCTTGCCTACAACTATGAGACTCTCATGGCTTGCATTAC
             ************************************************************

Sequenced    CATGAATATGCAGCATGTCCAGACCTTGACTACCGAGAAGCTCCAGTTGACTTCTGTTAC
GeneScript   CATGAATATGCAGCATGTCCAGACCTTGACTACCGAGAAGCTCCAGTTGACTTCTGTTAC
             ************************************************************

Sequenced    CTCCCTTTGTATGCTCATTGGAAACGCCACTGTGATCCCTTCCCCCCAGACCCTGTTCCA
GeneScript   CTCCCTTTGTATGCTCATTGGAAACGCCACTGTGATCCCTTCCCCCCAGACCCTGTTCCA
             ************************************************************

Sequenced    CTACTACAACGTCAACGTTAACTTCCATTCGAACTACAACGAGCGAATTAACGATGCTGT
GeneScript   CTACTACAACGTCAACGTTAACTTCCATTCGAACTACAACGAGCGAATTAACGATGCTGT
             ************************************************************

Sequenced    CGCCATTATCACTGCTGCCAACCGACTGAATCTTTACCAGAAGAAGATGAAGGCTATTGT
GeneScript   CGCCATTATCACTGCTGCCAACCGACTGAATCTTTACCAGAAGAAGATGAAGGCTATTGT
             ************************************************************

Sequenced    TGAGGATTTCTTGAAGCGACTGTATATCTTTGACGTGTCTCGAGTCCCTGACGATCAGAT
GeneScript   TGAGGATTTCTTGAAGCGACTGTATATCTTTGACGTGTCTCGAGTCCCTGACGATCAGAT
             ************************************************************
```

102

```
Sequenced    GTACCGACTGCGAGATCGACTTCGACTCTTGCCCGTCGAGATTCGACGACTGGACATTTT
GeneScript   GTACCGACTGCGAGATCGACTTCGACTCTTGCCCGTCGAGATTCGACGACTGGACATTTT
             ************************************************************

Sequenced    CAACCTCATCTTGATGAATATGGATCAGATTGAGCGAGCCTCCGACAAGATCGCTCAGGG
GeneScript   CAACCTCATCTTGATGAATATGGATCAGATTGAGCGAGCCTCCGACAAGATCGCTCAGGG
             ************************************************************

Sequenced    AGTCATTATCGCCTATCGAGATATGCACTTGGAGCGAGACGAGATGTACGGATATGTTAA
GeneScript   AGTCATTATCGCCTATCGAGATATGCACTTGGAGCGAGACGAGATGTACGGATATGTTAA
             ************************************************************

Sequenced    CATTGCTCGAAATCTTGAGGGTTTTTCAGCAGATCAACCTTGAGGAGCTCATGCGATCTGG
GeneScript   CATTGCTCGAAATCTTGAGGGTTTTTCAGCAGATCAACCTTGAGGAGCTCATGCGATCTGG
             ************************************************************

Sequenced    AGATTACGCCCAGATTACTAACATGCTGCTTAATAACCAGCCTGTTGCTCTCGTGGGTGC
GeneScript   AGATTACGCCCAGATTACTAACATGCTGCTTAATAACCAGCCTGTTGCTCTCGTGGGTGC
             ************************************************************

Sequenced    CTTGCCCTTCATTACCGATTCTTCCGTTATTTCCCTGATCGCTAAGCTTGACGCCACTGT
GeneScript   CTTGCCCTTCATTACCGATTCTTCCGTTATTTCCCTGATCGCTAAGCTTGACGCCACTGT
             ************************************************************

Sequenced    GTTTGCTCAGATTGTCAAGCTCCGAAAGGTTGACACCCTCAAGCCTATTTTGTACAAGAT
GeneScript   GTTTGCTCAGATTGTCAAGCTCCGAAAGGTTGACACCCTCAAGCCTATTTTGTACAAGAT
             ************************************************************

Sequenced    CAACTCGGATTCTAATGACTTCTATCTGGTTGCTAACTACGATTGGGTGCCTACTTCTAC
GeneScript   CAACTCGGATTCTAATGACTTCTATCTGGTTGCTAACTACGATTGGGTGCCTACTTCTAC
             ************************************************************

Sequenced    TACCAAGGTCTATAAGCAGGTTCCCCAGCAGTTCGACTTTCGAAACTCCATGCATATGCT
GeneScript   TACCAAGGTCTATAAGCAGGTTCCCCAGCAGTTCGACTTTCGAAACTCCATGCATATGCT
             ************************************************************

Sequenced    GACTTCGAATCTTACTTTCACCGTGTACTCGGATCTCTTGGCCTTTGTGTCTGCTGACAC
GeneScript   GACTTCGAATCTTACTTTCACCGTGTACTCGGATCTCTTGGCCTTTGTGTCTGCTGACAC
             ************************************************************

Sequenced    CGTCGAGCCTATTAACGCTGTCGCCTTTGACAACATGCGAATCATGAATGAGCTGTAACC
GeneScript   CGTCGAGCCTATTAACGCTGTCGCCTTTGACAACATGCGAATCATGAATGAGCTGTAACC
             ************************************************************

Sequenced    TAGGTGATCTGATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAAAAAAAATA
GeneScript   TAGG--------------------------------------------------------
             ****
```

Figure 2: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *A. adeninivorans* codon-optimized VP2 ORF to *in silico* sequences from GeneScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *SalI* and *XmaJI* restriction enzymes recognition site = underlined.

```
Sequenced    TCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGATGGCTTACAGAA
GeneArt      TCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGATGGCTTACAGAA
             ************************************************************

Sequenced    AGAGAGGTGCTAGAAGAGAAGCTAACTTGAACAACAACGACAGAATGCAAGAAAAGATCG
GeneArt      AGAGAGGTGCTAGAAGAGAAGCTAACTTGAACAACAACGACAGAATGCAAGAAAAGATCG
             ************************************************************

Sequenced    ACGAAAAGCAAGACTCTAACAAGATCCAATTGTCTGACAAGGTTTTGTCTAAGAAGGAAG
GeneArt      ACGAAAAGCAAGACTCTAACAAGATCCAATTGTCTGACAAGGTTTTGTCTAAGAAGGAAG
             ************************************************************

Sequenced    AGATCGTTACCGACTCTCACGAAGAAGTTAAGGTTACCGACGAATTGAAGAAGTCTACCA
GeneArt      AGATCGTTACCGACTCTCACGAAGAAGTTAAGGTTACCGACGAATTGAAGAAGTCTACCA
             ************************************************************

Sequenced    AGGAAGAATCTAAGCAATTGTTGGAAGTTTTGAAGACCAAGGAAGAACACCAAAAGGAAA
GeneArt      AGGAAGAATCTAAGCAATTGTTGGAAGTTTTGAAGACCAAGGAAGAACACCAAAAGGAAA
             ************************************************************

Sequenced    TCCAATACGAAATCTTGCAAAAGACCATCCCAACCTTCGAACCTAAGGAAACCATCTTGA
GeneArt      TCCAATACGAAATCTTGCAAAAGACCATCCCAACCTTCGAACCTAAGGAAACCATCTTGA
             ************************************************************

Sequenced    GAAAGTTGGAAGACATCCAACCAGAATTGGCTAAGAAGCAAACCAAGTTGTTCAGAATCT
GeneArt      GAAAGTTGGAAGACATCCAACCAGAATTGGCTAAGAAGCAAACCAAGTTGTTCAGAATCT
             ************************************************************

Sequenced    TCGAACCAAAGCAATTGCCAATCTACAGAGCTAACGGTGAAAGAGAATTGAGAAACAGAT
GeneArt      TCGAACCAAAGCAATTGCCAATCTACAGAGCTAACGGTGAAAGAGAATTGAGAAACAGAT
             ************************************************************

Sequenced    GGTACTGGAAGTTGAAGAAGGACACCTTGCCAGACGGTGACTACGACGTTAGAGAGTACT
GeneArt      GGTACTGGAAGTTGAAGAAGGACACCTTGCCAGACGGTGACTACGACGTTAGAGAGTACT
             ************************************************************

Sequenced    TCTTGAACCTCTACGACCAAGTCTTGACCGAAATGCCAGACTACTTGTTGTTGAAGGACA
GeneArt      TCTTGAACCTCTACGACCAAGTCTTGACCGAAATGCCAGACTACTTGTTGTTGAAGGACA
             ************************************************************

Sequenced    TGGCTGTTGAAAACAAGAACTCCAGAGACGCTGGTAAGGTTGTTGACTCTGAAACCGCTT
GeneArt      TGGCTGTTGAAAACAAGAACTCCAGAGACGCTGGTAAGGTTGTTGACTCTGAAACCGCTT
             ************************************************************

Sequenced    CTATCTGTGACGCTATCTTCCAAGACGAAGAAACCGAAGGTGCTGTTAGAAGATTCATCG
GeneArt      CTATCTGTGACGCTATCTTCCAAGACGAAGAAACCGAAGGTGCTGTTAGAAGATTCATCG
             ************************************************************

Sequenced    CTGAAATGAGACAAAGAGTTCAAGCTGACAGAAACGTTGTTAACTACCCATCTATCTTGC
GeneArt      CTGAAATGAGACAAAGAGTTCAAGCTGACAGAAACGTTGTTAACTACCCATCTATCTTGC
             ************************************************************

Sequenced    ACCCAATCGATTACGCTTTCAACGAATACTTCTTGCAACACCAACTTGTCGAACCATTGA
GeneArt      ACCCAATCGATTACGCTTTCAACGAATACTTCTTGCAACACCAACTTGTCGAACCATTGA
             ************************************************************

Sequenced    ACAACGACATCATCTTCAACTACATCCCAGAAAGAATCAGAAACGACGTTAACTATATCT
GeneArt      ACAACGACATCATCTTCAACTACATCCCAGAAAGAATCAGAAACGACGTTAACTATATCT
             ************************************************************

Sequenced    TGAACATGGACAGAAACTTGCCATCTACCGCTAGATACATCAGACCAAACTTGTTGCAAG
GeneArt      TGAACATGGACAGAAACTTGCCATCTACCGCTAGATACATCAGACCAAACTTGTTGCAAG
             ************************************************************

Sequenced    ACAGATTGAACTTGCACGACAACTTCGAATCTTTGTGGGACACCATCACCACCTCTAACT
GeneArt      ACAGATTGAACTTGCACGACAACTTCGAATCTTTGTGGGACACCATCACCACCTCTAACT
             ************************************************************
```

104

```
Sequenced    ACATCTTGGCTAGATCCGTTGTTCCAGACTTGAAGGAATTGGTTTCTACCGAAGCTCAAA
GeneArt      ACATCTTGGCTAGATCCGTTGTTCCAGACTTGAAGGAATTGGTTTCTACCGAAGCTCAAA
             ************************************************************

Sequenced    TCCAAAAGATGTCTCAAGACTTGCAATTGGAAGCCTTGACCATCCAATCTGAAACCCAAT
GeneArt      TCCAAAAGATGTCTCAAGACTTGCAATTGGAAGCCTTGACCATCCAATCTGAAACCCAAT
             ************************************************************

Sequenced    TCTTGACCGGTATCAACTCTCAAGCTGCTAACGACTGTTTCAAGACCTTGATCGCTGCTA
GeneArt      TCTTGACCGGTATCAACTCTCAAGCTGCTAACGACTGTTTCAAGACCTTGATCGCTGCTA
             ************************************************************

Sequenced    TGTTGTCTCAAAGAACCATGTCTTTGGACTTCGTTACCACCAACTACATGTCTTTGATCT
GeneArt      TGTTGTCTCAAAGAACCATGTCTTTGGACTTCGTTACCACCAACTACATGTCTTTGATCT
             ************************************************************

Sequenced    CTGGTATGTGGTTGTTGACCGTCGTTCCAAACGACATGTTCAT<mark>CAGAGAATCTTTGGTTG</mark>
GeneArt      CTGGTATGTGGTTGTTGACCGTCGTTCCAAACGACATGTTCAT<mark>CAGAGAATCTTTGGTTG</mark>
             ************************************************************

Sequenced    <mark>CTTG</mark>TCAATTGGCTATCGTTAACACCATCATCTACCCAGCTTTCGGTATGCAAAGAATGC
GeneArt      <mark>CTTG</mark>TCAATTGGCTATCGTTAACACCATCATCTACCCAGCTTTCGGTATGCAAAGAATGC
             ************************************************************

Sequenced    ACTACAGAAACGGTGACCCACAAACCCCATTCCAAATCGCTGAACAACAAATCCAAAACT
GeneArt      ACTACAGAAACGGTGACCCACAAACCCCATTCCAAATCGCTGAACAACAAATCCAAAACT
             ************************************************************

Sequenced    TCCAAGTTGCTAACTGGTTGCACTTCGTTAACAACAACCAATTCAGACAAGCTGTTATCG
GeneArt      TCCAAGTTGCTAACTGGTTGCACTTCGTTAACAACAACCAATTCAGACAAGCTGTTATCG
             ************************************************************

Sequenced    ACGGTGTTTTGAACCAAGTTTTGAA<mark>CGACAACATCAGAAACGGTC</mark>ACGTTATCAACCAAT
GeneArt      ACGGTGTTTTGAACCAAGTTTTGAA<mark>CGACAACATCAGAAACGGTC</mark>ACGTTATCAACCAAT
             ************************************************************

Sequenced    TGATGGAAGCCTTGATGCAATTGTCCAGACAACAATTCCCAACCATGCCAATCGACTACA
GeneArt      TGATGGAAGCCTTGATGCAATTGTCCAGACAACAATTCCCAACCATGCCAATCGACTACA
             ************************************************************

Sequenced    AGAGATCCATCCAAAGAGGTATCTTGTTGTTGTCTAACAGATTGGGTCAATTGGTTGACT
GeneArt      AGAGATCCATCCAAAGAGGTATCTTGTTGTTGTCTAACAGATTGGGTCAATTGGTTGACT
             ************************************************************

Sequenced    TGACCAGATTGTTGGCTTACAACTACGAAACCTTGATGGCTTGTATCACCATGAACATGC
GeneArt      TGACCAGATTGTTGGCTTACAACTACGAAACCTTGATGGCTTGTATCACCATGAACATGC
             ************************************************************

Sequenced    AACACGTTCAAACCTTGACCACCGAAAAGTTGCAATTGACCTCTGTTACCTCTTTGTGTA
GeneArt      AACACGTTCAAACCTTGACCACCGAAAAGTTGCAATTGACCTCTGTTACCTCTTTGTGTA
             ************************************************************

Sequenced    TGTTGATCGGTAACGCTACCGTTATCCCATCTCCACAAACCTTGTTCCACTACTACAACG
GeneArt      TGTTGATCGGTAACGCTACCGTTATCCCATCTCCACAAACCTTGTTCCACTACTACAACG
             ************************************************************

Sequenced    TTAACGTTAACTTCCACTCTAACTACAACGAAAGAATCAACGACGCTGTTGCTATCATCA
GeneArt      TTAACGTTAACTTCCACTCTAACTACAACGAAAGAATCAACGACGCTGTTGCTATCATCA
             ************************************************************

Sequenced    CCGCTGCTAACAGATTGAACTTGTACCAAAAGAAGATGAAGGCTATCGTTGAAGACTTCT
GeneArt      CCGCTGCTAACAGATTGAACTTGTACCAAAAGAAGATGAAGGCTATCGTTGAAGACTTCT
             ************************************************************

Sequenced    TGAAGAGATTGTACATCTTC<mark>GACGTTTCCAGAGTTCCAGAC</mark>GACCAAATGTACAGATTGA
GeneArt      TGAAGAGATTGTACATCTTC<mark>GACGTTTCCAGAGTTCCAGAC</mark>GACCAAATGTACAGATTGA
             ************************************************************
```

105

```
Sequenced   GAGACAGATTGAGATTGTTGCCAGTTGAAATCAGAAGATTGGACATCTTCAACTTGATCT
GeneArt     GAGACAGATTGAGATTGTTGCCAGTTGAAATCAGAAGATTGGACATCTTCAACTTGATCT
            ************************************************************

Sequenced   TGATGAACATGGACCAAATCGAAAGAGCTTCTGACAAGATCGCTCAAGGTGTTATCATCG
GeneArt     TGATGAACATGGACCAAATCGAAAGAGCTTCTGACAAGATCGCTCAAGGTGTTATCATCG
            ************************************************************

Sequenced   CTTACAGAGACATGCACTTGGAAAGAGACGAAATGTACGGTTACGTTAACATCGCTAGAA
GeneArt     CTTACAGAGACATGCACTTGGAAAGAGACGAAATGTACGGTTACGTTAACATCGCTAGAA
            ************************************************************

Sequenced   ACTTGGAAGGTTTCCAACAAATCAACTTGGAAGAATTGATGAGATCCGGTGACTACGCTC
GeneArt     ACTTGGAAGGTTTCCAACAAATCAACTTGGAAGAATTGATGAGATCCGGTGACTACGCTC
            ************************************************************

Sequenced   AAATCACCAACATGTTGTTGAACAACCAACCAGTTGCTTTGGTTGGTGCTTTGCCATTCA
GeneArt     AAATCACCAACATGTTGTTGAACAACCAACCAGTTGCTTTGGTTGGTGCTTTGCCATTCA
            ************************************************************

Sequenced   TCACCGACTCTTCTGTTATCTCTTTGATCGCTAAGTTGGACGCTACCGTTTTCGCTCAAA
GeneArt     TCACCGACTCTTCTGTTATCTCTTTGATCGCTAAGTTGGACGCTACCGTTTTCGCTCAAA
            ************************************************************

Sequenced   TCGTTAAGTTGAGAAAGGTTGACACCTTGAAGCCAATCTTGTACAAGATCAACTCTGACT
GeneArt     TCGTTAAGTTGAGAAAGGTTGACACCTTGAAGCCAATCTTGTACAAGATCAACTCTGACT
            ************************************************************

Sequenced   CTAACGACTTCTACTTGGTTGCTAACTACGACTGGGTTCCAACCTCTACCACCAAGGTTT
GeneArt     CTAACGACTTCTACTTGGTTGCTAACTACGACTGGGTTCCAACCTCTACCACCAAGGTTT
            ************************************************************

Sequenced   ACAAGCAAGTTCCACAACAATTCGACTTCAGAAACTCTATGCACATGTTGACCTCTAACT
GeneArt     ACAAGCAAGTTCCACAACAATTCGACTTCAGAAACTCTATGCACATGTTGACCTCTAACT
            ************************************************************

Sequenced   TGACCTTCACCGTTTACTCTGACTTGTTGGCTTTCGTTTCTGCTGACACCGTTGAACCAA
GeneArt     TGACCTTCACCGTTTACTCTGACTTGTTGGCTTTCGTTTCTGCTGACACCGTTGAACCAA
            ************************************************************

Sequenced   TCAACGCTGTTGCTTTCGACAACATGAGAATCATGAACGAATTGTAACCTAGGTGATCTG
GeneArt     TCAACGCTGTTGCTTTCGACAACATGAGAATCATGAACGAATTGTAACCTAGGTGATCTG
            ************************************************************

Sequenced   ATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAAAAAAAATATTACTACTATT
GeneArt     ATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAA-------------------
            *****************************************
```

Figure 3: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *P. pastoris/H. polymorpha* codon-optimized VP2 ORF to *in silico* sequences from GeneArt. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *XhoI* and *XmaJI* restriction enzymes recognition site = underlined.

```
Sequenced    TCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGCATGGATGTATTGTATTCATTGAGT
GeneScript   ----------------------------CTCGAGCATGGATGTATTGTATTCATTGAGT
                                         ****************************

Sequenced    AAAACTTTGAAGGATGCTAGAGATAAAATCGTGGAAGGAACTTTGTATAGTAACGTGTCT
GeneScript   AAAACTTTGAAGGATGCTAGAGATAAAATCGTGGAAGGAACTTTGTATAGTAACGTGTCT
             ************************************************************

Sequenced    GATTTGATACAACAATTCAACCAAATGATAATCACCATGAATGGAAACGAATTTCAAACA
GeneScript   GATTTGATACAACAATTCAACCAAATGATAATCACCATGAATGGAAACGAATTTCAAACA
             ************************************************************

Sequenced    GGTGGAATCGGTAATTTGCCAATCAGAAACTGGAACTTCGATTTCGGTTTGTTGGGAACT
GeneScript   GGTGGAATCGGTAATTTGCCAATCAGAAACTGGAACTTCGATTTCGGTTTGTTGGGAACT
             ************************************************************

Sequenced    ACATTGTTGAATTTGGATGCTAACTACGTGGAAACAGCAAGAAACACCATAGATTACTTC
GeneScript   ACATTGTTGAATTTGGATGCTAACTACGTGGAAACAGCAAGAAACACCATAGATTACTTC
             ************************************************************

Sequenced    GTCGATTTCGTAGATAACGTTTGTATGGATGAAATGGTTAGAGAATCACAAAGAAATGGT
GeneScript   GTCGATTTCGTAGATAACGTTTGTATGGATGAAATGGTTAGAGAATCACAAAGAAATGGT
             ************************************************************

Sequenced    ATCGCTCCACAAAGTGATTCCTTGAGAAAATTATCTGGTATTAAATTCAAGAGAATCAAT
GeneScript   ATCGCTCCACAAAGTGATTCCTTGAGAAAATTATCTGGTATTAAATTCAAGAGAATCAAT
             ************************************************************

Sequenced    TTCGATAACTCTTCAGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAACC
GeneScript   TTCGATAACTCTTCAGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAACC
             ************************************************************

Sequenced    GGTTTCACTTTTCATAAGCCAAATATCTTCCCTTACTCTGCATCTTTCACTTTAAATAGA
GeneScript   GGTTTCACTTTTCATAAGCCAAATATCTTCCCTTACTCTGCATCTTTCACTTTAAATAGA
             ************************************************************

Sequenced    TCTCAACCAGCTCATGATAACTTGATGGGTACAATGTGGTTAAATGCTGGATCTGAAATC
GeneScript   TCTCAACCAGCTCATGATAACTTGATGGGTACAATGTGGTTAAATGCTGGATCTGAAATC
             ************************************************************

Sequenced    CAAGTCGCAGGTTTTGATTATTCATGTGCTATAAATGCTCCTGCAAACACTCAACAATTC
GeneScript   CAAGTCGCAGGTTTTGATTATTCATGTGCTATAAATGCTCCTGCAAACACTCAACAATTC
             ************************************************************

Sequenced    GAACATATCGTACAATTGAGAAGAGTTTTGACCACTGCAACTATCACATTGTTACCAGAT
GeneScript   GAACATATCGTACAATTGAGAAGAGTTTTGACCACTGCAACTATCACATTGTTACCAGAT
             ************************************************************

Sequenced    GCTGAAAGATTCTCTTTTCCTAGAGTGATAAACTCAGCTGATGGTGCAACAACCTGGTAT
GeneScript   GCTGAAAGATTCTCTTTTCCTAGAGTGATAAACTCAGCTGATGGTGCAACAACCTGGTAT
             ************************************************************

Sequenced    TTCAATCCAGTCATTTTGAGACCTAACAACGTTGAAGTGGAATTCTTGTTGAACGGACAA
GeneScript   TTCAATCCAGTCATTTTGAGACCTAACAACGTTGAAGTGGAATTCTTGTTGAACGGACAA
             ************************************************************

Sequenced    ATAATTAACACTTACCAAGCAAGATTCGGTACAATCGTTGCTAGAAACTTCGATACTATC
GeneScript   ATAATTAACACTTACCAAGCAAGATTCGGTACAATCGTTGCTAGAAACTTCGATACTATC
             ************************************************************

Sequenced    AGATTGAGTTTCCAATTGATGAGACCACCTAACATGACACCATCAGTAGCTGCATTGTTC
GeneScript   AGATTGAGTTTCCAATTGATGAGACCACCTAACATGACACCATCAGTAGCTGCATTGTTC
             ************************************************************

Sequenced    CCAAATGCACAACCTTTTGAACATCATGCTACTGTTGGTTTGACATTGAAGATCGAAAGT
GeneScript   CCAAATGCACAACCTTTTGAACATCATGCTACTGTTGGTTTGACATTGAAGATCGAAAGT
             ************************************************************
```

107

```
Sequenced    GCAGTATGTGAATCCGTTTTGGCTGATGCAAGTGAAACAATGTTAGCAAACGTGACCTCC
GeneScript   GCAGTATGTGAATCCGTTTTGGCTGATGCAAGTGAAACAATGTTAGCAAACGTGACCTCC
             ************************************************************

Sequenced    GTCAGACAAGAATATGCTATTCCAGTAGGTCCTGTTTTTCCACCTGGAATGAATTGGACC
GeneScript   GTCAGACAAGAATATGCTATTCCAGTAGGTCCTGTTTTTCCACCTGGAATGAATTGGACC
             ************************************************************

Sequenced    GATTTGATCACTAACTACAGTCCATCCAGAGAAGATAATTTGCAAAGAGTGTTCACTGTC
GeneScript   GATTTGATCACTAACTACAGTCCATCCAGAGAAGATAATTTGCAAAGAGTGTTCACTGTC
             ************************************************************

Sequenced    GCTTCTATCAGATCAATGTTGGTTAAATAAAGCGCTATTAATCCTAGGTGATCTGATCTG
GeneScript   GCTTCTATCAGATCAATGTTGGTTAAATAAAGCGCT------------------------
             ***********************************
```

Figure 4: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *K. lactis* codon-optimized VP6 ORF to *in silico* sequences from GeneScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *XhoI* and *Eco47III* restriction enzyme recognition site = underlined.

```
Sequenced    TCTCTCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGCATGGATGTTTTGTATTCGCT
GeneScript   --------------------------------CTCGAGCATGGATGTTTTGTATTCGCT
                                             ************************

Sequenced    TTCTAAGACTCTGAAGGATGCCCGAGATAAGATTGTTGAGGGAACTTTGTATTCTAATGT
GeneScript   TTCTAAGACTCTGAAGGATGCCCGAGATAAGATTGTTGAGGGAACTTTGTATTCTAATGT
             ************************************************************

Sequenced    TTCTGATCTGATTCAGCAGTTCAACCAGATGATTATCACTATGAACGGTAATGAGTTTCA
GeneScript   TTCTGATCTGATTCAGCAGTTCAACCAGATGATTATCACTATGAACGGTAATGAGTTTCA
             ************************************************************

Sequenced    GACCGGAGGTATTGGAAACCTCCCTATCCGAAACTGGAATTTCGACTTTGGACTGCTTGG
GeneScript   GACCGGAGGTATTGGAAACCTCCCTATCCGAAACTGGAATTTCGACTTTGGACTGCTTGG
             ************************************************************

Sequenced    TACTACCCTCTTGAACTTGGATGCTAATTACGTGGAGACTGCCCGAAATACCATTGACTA
GeneScript   TACTACCCTCTTGAACTTGGATGCTAATTACGTGGAGACTGCCCGAAATACCATTGACTA
             ************************************************************

Sequenced    TTTCGTTGACTTTGTGGATAACGTCTGCATGGATGAGATGGTCCGAGAGTCTCAGCGAAA
GeneScript   TTTCGTTGACTTTGTGGATAACGTCTGCATGGATGAGATGGTCCGAGAGTCTCAGCGAAA
             ************************************************************

Sequenced    CGGTATTGCTCCTCAGTCGGACTCTCTCCGAAAGTTGTCCGGAATTAAGTTCAAGCGAAT
GeneScript   CGGTATTGCTCCTCAGTCGGACTCTCTCCGAAAGTTGTCCGGAATTAAGTTCAAGCGAAT
             ************************************************************

Sequenced    CAACTTCGATAATTCTTCCGAGTACATTGAGAACTGGAATCTGCAGAACCGACGACAGCG
GeneScript   CAACTTCGATAATTCTTCCGAGTACATTGAGAACTGGAATCTGCAGAACCGACGACAGCG
             ************************************************************

Sequenced    AACTGGATTCACCTTTCACAAGCCTAACATCTTCCCCTATTCCGCCTCGTTTACTCTTAA
GeneScript   AACTGGATTCACCTTTCACAAGCCTAACATCTTCCCCTATTCCGCCTCGTTTACTCTTAA
             ************************************************************

Sequenced    CCGATCTCAGCCTGCTCATGACAATCTGATGGGAACCATGTGGCTTAACGCTGGTTCCGA
GeneScript   CCGATCTCAGCCTGCTCATGACAATCTGATGGGAACCATGTGGCTTAACGCTGGTTCCGA
             ************************************************************

Sequenced    GATTCAGGTTGCCGGATTTGATTACTCGTGTGCTATCAACGCTCCCGCCAATACTCAGCA
GeneScript   GATTCAGGTTGCCGGATTTGATTACTCGTGTGCTATCAACGCTCCCGCCAATACTCAGCA
             ************************************************************

Sequenced    GTTCGAGCACATTGTGCAGCTCCGACGAGTCTTGACTACCGCCACTATCACCCTGCTTCC
GeneScript   GTTCGAGCACATTGTGCAGCTCCGACGAGTCTTGACTACCGCCACTATCACCCTGCTTCC
             ************************************************************

Sequenced    TGACGCTGAGCGATTCTCTTTTCCCCGAGTTATTAATTCCGCTGATGGTGCCACTACCTG
GeneScript   TGACGCTGAGCGATTCTCTTTTCCCCGAGTTATTAATTCCGCTGATGGTGCCACTACCTG
             ************************************************************

Sequenced    GTACTTCAACCCTGTGATCCTCCGACCCAACAATGTGGAGGTCGAGTTTCTCTTGAACGG
GeneScript   GTACTTCAACCCTGTGATCCTCCGACCCAACAATGTGGAGGTCGAGTTTCTCTTGAACGG
             ************************************************************

Sequenced    TCAGATTATCAATACTTATCAGGCCCGATTCGGAACCATTGTGGCTCGAAACTTCGACAC
GeneScript   TCAGATTATCAATACTTATCAGGCCCGATTCGGAACCATTGTGGCTCGAAACTTCGACAC
             ************************************************************

Sequenced    TATCCGACTGTCTTTTCAGCTTATGCGACCTCCCAATATGACCCCTTCCGTCGCTGCCCT
GeneScript   TATCCGACTGTCTTTTCAGCTTATGCGACCTCCCAATATGACCCCTTCCGTCGCTGCCCT
             ************************************************************

Sequenced    TTTCCCTAACGCCCAGCCCTTTGAGCACCATGCTACTGTTGGACTGACCCTTAAGATTGA
GeneScript   TTTCCCTAACGCCCAGCCCTTTGAGCACCATGCTACTGTTGGACTGACCCTTAAGATTGA
             ************************************************************
```

```
Sequenced    GTCCGCCGTTTGCGAGTCGGTGCTCGCTGATGCCTCGGAGACTATGTTGGCCAACGTTAC
GeneScript   GTCCGCCGTTTGCGAGTCGGTGCTCGCTGATGCCTCGGAGACTATGTTGGCCAACGTTAC
             ************************************************************

Sequenced    CTCTGTGCGACAGGAGTACGCTATTCCTGTCGGACCCGTTTTTCCTCCCGGTATGAACTG
GeneScript   CTCTGTGCGACAGGAGTACGCTATTCCTGTCGGACCCGTTTTTCCTCCCGGTATGAACTG
             ************************************************************

Sequenced    GACTGACCTGATCACCAATTATTCTCCTTCCCGAGAGGATAACCTTCAGCGAGTGTTCAC
GeneScript   GACTGACCTGATCACCAATTATTCTCCTTCCCGAGAGGATAACCTTCAGCGAGTGTTCAC
             ************************************************************

Sequenced    TGTCGCTTCGATTCGATCTATGCTGGTCAAGTAGAGCGCTATTAATCCTAGGTGATCTGA
GeneScript   TGTCGCTTCGATTCGATCTATGCTGGTCAAGTAGAGCGCTATTAA---------------
             ********************************************
```

Figure 5: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *A. adeninivorans* codon-optimized VP6 ORF to *in silico* sequences from GeneScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *XhoI* and *Eco47III* restriction enzymes recognition site = underlined.

```
Sequenced   TTTCTCTCTCTCCTTGTCAACTCACACCCGNAAATGCTCGAGATGGACGTTTTGTACTCT
GeneArt     ----------------------------------CTCGAGATGGACGTTTTGTACTCT
                                              ********************

Sequenced   TTGTCTAAGACCTTGAAGGACGCTAGAGACAAGATCGTTGAAGGCACCTTGTACTCTAAC
GeneArt     TTGTCTAAGACCTTGAAGGACGCTAGAGACAAGATCGTTGAAGGCACCTTGTACTCTAAC
            ************************************************************

Sequenced   GTTTCTGACTTGATCCAACAATTCAACCAAATGATCATCACCATGAACGGTAACGAGTTC
GeneArt     GTTTCTGACTTGATCCAACAATTCAACCAAATGATCATCACCATGAACGGTAACGAGTTC
            ************************************************************

Sequenced   CAAACCGGTGGTATCGGTAACTTGCCAATCAGAAACTGGAATTTCGACTTCGGTTTGTTG
GeneArt     CAAACCGGTGGTATCGGTAACTTGCCAATCAGAAACTGGAATTTCGACTTCGGTTTGTTG
            ************************************************************

Sequenced   GGTACTACCTTGTTGAACTTGGACGCTAACTACGTTGAAACCGCTAGAAACACCATCGAC
GeneArt     GGTACTACCTTGTTGAACTTGGACGCTAACTACGTTGAAACCGCTAGAAACACCATCGAC
            ************************************************************

Sequenced   TACTTCGTTGACTTCGTTGACAACGTTTGTATGGACGAAATGGTTAGAGAATCTCAAAGA
GeneArt     TACTTCGTTGACTTCGTTGACAACGTTTGTATGGACGAAATGGTTAGAGAATCTCAAAGA
            ************************************************************

Sequenced   AACGGTATCGCTCCACAATCTGACTCTTTGAGAAAGTTGTCTGGTATCAAGTTCAAGAGA
GeneArt     AACGGTATCGCTCCACAATCTGACTCTTTGAGAAAGTTGTCTGGTATCAAGTTCAAGAGA
            ************************************************************

Sequenced   ATCAACTTCGACAACTCTTCTGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAA
GeneArt     ATCAACTTCGACAACTCTTCTGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAA
            ************************************************************

Sequenced   AGAACCGGTTTCACCTTCCACAAGCCAAACATCTTCCCATACTCTGCTTCTTTCACCTTG
GeneArt     AGAACCGGTTTCACCTTCCACAAGCCAAACATCTTCCCATACTCTGCTTCTTTCACCTTG
            ************************************************************

Sequenced   AACAGATCCCAACCAGCTCACGACAACTTGATGGGTACTATGTGGTTGAACGCTGGTTCT
GeneArt     AACAGATCCCAACCAGCTCACGACAACTTGATGGGTACTATGTGGTTGAACGCTGGTTCT
            ************************************************************

Sequenced   GAAATCCAAGTTGCTGGTTTCGACTACTCTTGTGCTATCAACGCTCCAGCTAACACCCAA
GeneArt     GAAATCCAAGTTGCTGGTTTCGACTACTCTTGTGCTATCAACGCTCCAGCTAACACCCAA
            ************************************************************

Sequenced   CAATTCGAACACATCGTTCAATTGAGAAGAGTTTTTGACCACCGCTACCATCACCTTGTTG
GeneArt     CAATTCGAACACATCGTTCAATTGAGAAGAGTTTTTGACCACCGCTACCATCACCTTGTTG
            ************************************************************

Sequenced   CCAGACGCTGAAAGATTCTCTTTCCCAAGAGTTATCAACTCTGCTGACGGTGCTACCACC
GeneArt     CCAGACGCTGAAAGATTCTCTTTCCCAAGAGTTATCAACTCTGCTGACGGTGCTACCACC
            ************************************************************

Sequenced   TGGTACTTCAACCCAGTTATCTTGAGACCAAACAACGTTGAAGTTGAGTTCTTGTTGAAC
GeneArt     TGGTACTTCAACCCAGTTATCTTGAGACCAAACAACGTTGAAGTTGAGTTCTTGTTGAAC
            ************************************************************

Sequenced   GGTCAAATCATCAACACCTACCAAGCTAGATTCGGCACCATCGTTGCTAGAAACTTCGAC
GeneArt     GGTCAAATCATCAACACCTACCAAGCTAGATTCGGCACCATCGTTGCTAGAAACTTCGAC
            ************************************************************

Sequenced   ACCATCAGATTGTCTTTCCAATTGATGAGACCACCAAACATGACCCCATCTGTTGCTGCT
GeneArt     ACCATCAGATTGTCTTTCCAATTGATGAGACCACCAAACATGACCCCATCTGTTGCTGCT
            ************************************************************

Sequenced   TTGTTCCCAAACGCTCAACCATTCGAACACCACGCTACCGTTGGTTTGACCTTGAAGATC
GeneArt     TTGTTCCCAAACGCTCAACCATTCGAACACCACGCTACCGTTGGTTTGACCTTGAAGATC
            ************************************************************
```

```
Sequenced    GAATCTGCTGTTTGTGAATCTGTTTTGGCTGACGCTTCTGAAACCATGTTGGCTAACGTT
GeneArt      GAATCTGCTGTTTGTGAATCTGTTTTGGCTGACGCTTCTGAAACCATGTTGGCTAACGTT
             ************************************************************

Sequenced    ACCTCTGTTAGACAAGAATACGCTATCCCAGTTGGTCCAGTTTTTCCCACCAGGTATGAAC
GeneArt      ACCTCTGTTAGACAAGAATACGCTATCCCAGTTGGTCCAGTTTTTCCCACCAGGTATGAAC
             ************************************************************

Sequenced    TGGACCGACTTGATCACCAACTACTCTCCATCCAGAGAAGACAACTTGCAAAGAGTTTTC
GeneArt      TGGACCGACTTGATCACCAACTACTCTCCATCCAGAGAAGACAACTTGCAAAGAGTTTTC
             ************************************************************

Sequenced    ACCGTTGCTTCTATCAGATCCATGTTGGTTAAGTAAAGCGCTTGCATCCTAGGTGATCTG
GeneArt      ACCGTTGCTTCTATCAGATCCATGTTGGTTAAGTAAAGCGCTTGC---------------
             *********************************************
```

Figure 6: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *P. pastoris/H. polymorpha* codon-optimized VP6 ORF to *in silico* sequences from GeneArt. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *XhoI* and *Eco47III* restriction enzymes recognition site = underlined.

```
Sequenced    TCGCTGGCCGGGTGACCCGGCGGGGACGAGGCAAGCTAAACAGATCTAGGGATAACAGGG
GeneScript   TCGCTGGCCGGGTGACCCGGCGGGGACGAGGCAAGCTAAACAGATCTAGGGATAACAGGG
             ************************************************************


Sequenced    TAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAACAGCCCCAATTGC
GeneScript   TAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAACAGCCCCAATTGC
             ************************************************************


Sequenced    CCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGAGCCCCCTTCACC
GeneScript   CCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGAGCCCCCTTCACC
             ************************************************************


Sequenced    CCACATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTAGGGTACTGCAGT
GeneScript   CCACATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTAGGGTACTGCAGT
             ************************************************************


Sequenced    CTGGAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCCATCCGGGTAACC
GeneScript   CTGGAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCCATCCGGGTAACC
             ************************************************************


Sequenced    CATGCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTTGGGACTTTAGCC
GeneScript   CATGCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTTGGGACTTTAGCC
             ************************************************************


Sequenced    AAGGGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTG
GeneScript   AAGGGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTG
             ************************************************************


Sequenced    TCAACTCACACCCGAAATGCTCGAGCATGGATGTATTGTATTCATTGAGTAAAACTTTGA
GeneScript   TCAACTCACACCCGAAATGCTCGAGCATGGATGTATTGTATTCATTGAGTAAAACTTTGA
             ************************************************************


Sequenced    AGGATGCTAGAGATAAAATCGTGGAAGGAACTTTGTATAGTAACGTGTCTGATTTGATAC
GeneScript   AGGATGCTAGAGATAAAATCGTGGAAGGAACTTTGTATAGTAACGTGTCTGATTTGATAC
             ************************************************************


Sequenced    AACAATTCAACCAAATGATAATCACCATGAATGGAAACGAATTTCAAACAGGTGGAATCG
GeneScript   AACAATTCAACCAAATGATAATCACCATGAATGGAAACGAATTTCAAACAGGTGGAATCG
             ************************************************************


Sequenced    GTAATTTGCCAATCAGAAACTGGAACTTCGATTTCGGTTTGTTGGGAACTACATTGTTGA
GeneScript   GTAATTTGCCAATCAGAAACTGGAACTTCGATTTCGGTTTGTTGGGAACTACATTGTTGA
             ************************************************************


Sequenced    ATTTGGATGCTAACTACGTGGAAACAGCAAGAAACACCATAGATTACTTCGTCGATTTCG
GeneScript   ATTTGGATGCTAACTACGTGGAAACAGCAAGAAACACCATAGATTACTTCGTCGATTTCG
             ************************************************************


Sequenced    TAGATAACGTTTGTATGGATGAAATGGTTAGAGAATCACAAAGAAATGGTATCGCTCCAC
GeneScript   TAGATAACGTTTGTATGGATGAAATGGTTAGAGAATCACAAAGAAATGGTATCGCTCCAC
             ************************************************************


Sequenced    AAAGTGATTCCTTGAGAAAATTATCTGGTATTAAATTCAAGAGAATCAATTTCGATAACT
GeneScript   AAAGTGATTCCTTGAGAAAATTATCTGGTATTAAATTCAAGAGAATCAATTTCGATAACT
             ************************************************************


Sequenced    CTTCAGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAACCGGTTTCACTT
GeneScript   CTTCAGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAACCGGTTTCACTT
             ************************************************************


Sequenced    TTCATAAGCCAAATATCTTCCCTTACTCTGCATCTTTCACTTTAAATAGATCTCAACCAG
GeneScript   TTCATAAGCCAAATATCTTCCCTTACTCTGCATCTTTCACTTTAAATAGATCTCAACCAG
             ************************************************************


Sequenced    CTCATGATAACTTGATGGGTACAATGTGGTTAAATGCTGGATCTGAAATCCAAGTCGCAG
GeneScript   CTCATGATAACTTGATGGGTACAATGTGGTTAAATGCTGGATCTGAAATCCAAGTCGCAG
             ************************************************************
```

```
Sequenced    GTTTTGATTATTCATGTGCTATAAATGCTCCTGCAAACACTCAACAATTCGAACATATCG
GeneScript   GTTTTGATTATTCATGTGCTATAAATGCTCCTGCAAACACTCAACAATTCGAACATATCG
             ************************************************************


Sequenced    TACAATTGAGAAGAGTTTTGACCACTGCAACTATCACATTGTTACCAGATGCTGAAAGAT
GeneScript   TACAATTGAGAAGAGTTTTGACCACTGCAACTATCACATTGTTACCAGATGCTGAAAGAT
             ************************************************************


Sequenced    TCTCTTTTCCTAGAGTGATAAACTCAGCTGATGGTGCAACAACCTGGTATTTCAATCCAG
GeneScript   TCTCTTTTCCTAGAGTGATAAACTCAGCTGATGGTGCAACAACCTGGTATTTCAATCCAG
             ************************************************************


Sequenced    TCATTTTGAGACCTAACAACGTTGAAGTGGAATTCTTGTTGAACGGACAAATAATTAACA
GeneScript   TCATTTTGAGACCTAACAACGTTGAAGTGGAATTCTTGTTGAACGGACAAATAATTAACA
             ************************************************************


Sequenced    CTTA==CCAAGCAAGATTCGGTAC==AATCGTTGCTAGAAACTTCGATACTATCAGATTGAGTT
GeneScript   CTTA==CCAAGCAAGATTCGGTAC==AATCGTTGCTAGAAACTTCGATACTATCAGATTGAGTT
             ************************************************************


Sequenced    TCCAATTGATGAGACCACCTAACATGACACCATCAGTAGCTGCATTGTTCCCAAATGCAC
GeneScript   TCCAATTGATGAGACCACCTAACATGACACCATCAGTAGCTGCATTGTTCCCAAATGCAC
             ************************************************************


Sequenced    AACCTTTTGAACATCATGCTACTGTTGGTTTGACATTGAAGATCGAAAGTGCAGTATGTG
GeneScript   AACCTTTTGAACATCATGCTACTGTTGGTTTGACATTGAAGATCGAAAGTGCAGTATGTG
             ************************************************************


Sequenced    AATCCGTTTTGGCTGATGCAAGTGAAACAATGTTAGCAAACGTGACCTCCGTCAGACAAG
GeneScript   AATCCGTTTTGGCTGATGCAAGTGAAACAATGTTAGCAAACGTGACCTCCGTCAGACAAG
             ************************************************************


Sequenced    AATATGCTATTCCAGTAGGTCCTGTTTTTCCACCTGGAATGAATTGGACCGATTTGATCA
GeneScript   AATATGCTATTCCAGTAGGTCCTGTTTTTCCACCTGGAATGAATTGGACCGATTTGATCA
             ************************************************************


Sequenced    CTAACTACAGTCCATCCAGAGAAGATAATTTGCAAA==GAGTGTTCACTGTCGCTTC==TATCA
GeneScript   CTAACTACAGTCCATCCAGAGAAGATAATTTGCAAA==GAGTGTTCACTGTCGCTTC==TATCA
             ************************************************************


Sequenced    GATCAATGTTGGTTAAA==TAA==AGCGCTATTAATCCTAGGTGATCTGATCTGCTTACTTTAC
GeneScript   GATCAATGTTGGTTAAA==TAA==AGCGCTATTAATCCTAGGTGATCTGATCTGCTTACTTTAC
             ************************************************************


Sequenced    TTAACGACCAAAGAAAAACGACAAAAAAAAAAATATTACTACTATTAAAATAAATTAGTAT
GeneScript   TTAACGACCAAAGAAAAACGACAAAAAAAAAAATATTACTACTATTAAAATAAATTAGTAT
             ************************************************************


Sequenced    TTTTCTCTTCTTACGATATGATATGATGCTATGAAATCATCATCTTCTTAACTTTCTTGT
GeneScript   TTTTCTCTTCTTACGATATGATATGATGCTATGAAATCATCATCTTCTTAACTTTCTTGT
             ************************************************************


Sequenced    CTCTTACACGTCACTTACTCTATATACCCGTTTATATAAGTGTACGTATTTTCTTTTTTT
GeneScript   CTCTTACACGTCACTTACTCTATATACCCGTTTATATAAGTGTACGTATTTTCTTTTTTT
             ************************************************************


Sequenced    TAAAAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCAGTTCCAACGCACTCTAG
GeneScript   TAAAAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCAGTTCCAACGCACTCTAG
             ************************************************************


Sequenced    CTGTTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCACCCCAAAAATAACTTTC
GeneScript   CTGTTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCACCCCAAAAATAACTTTC
             ************************************************************


Sequenced    TTCCCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCATTTTCGCATCTGAAACTA
GeneScript   TTCCCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCATTTTCGCATCTGAAACTA
             ************************************************************
```

```
Sequenced    ATTCTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCATCACTAGAAATGAGAAAA
GeneScript   ATTCTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCATCACTAGAAATGAGAAAA
             ************************************************************

Sequenced    AGATCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCGGATCCTAGGGATAACAG
GeneScript   AGATCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCGGATCCTAGGGATAACAG
             ************************************************************

Sequenced    GGTAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAACAGCCCCAATT
GeneScript   GGTAATGG----------------------------------------------------
             ********
```

Figure 7: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *K. lactis* codon-optimized VP6 cassette ORF to *in silico* sequences from GeneScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *I-SceI* restriction enzyme recognition site = underlined, *Yarrowia lypolytica* TEF promoter = blue text and *Kluyveromyces marxianus* inulase terminator = red text.

```
Sequenced    CTGGCCGGGTGACCCGGCGGGGACGAGGCAAGCTAAACAGATCTAGGGATAACAGGGTAA
GeneScript   ------------CCCGGCGGGGACGAGGCAAGCTAAACAGATCTAGGGATAACAGGGTAA
             ************************************************

Sequenced    TGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAACAGCCCCAATTGCCCC
GeneScript   TGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAACAGCCCCAATTGCCCC
             ************************************************************

Sequenced    AATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGAGCCCCCTTCACCCCA
GeneScript   AATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGAGCCCCCTTCACCCCA
             ************************************************************

Sequenced    CATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTAGGGTACTGCAGTCTG
GeneScript   CATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTAGGGTACTGCAGTCTG
             ************************************************************

Sequenced    GAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCCATCCGGGTAACCCAT
GeneScript   GAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCCATCCGGGTAACCCAT
             ************************************************************

Sequenced    GCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTTGGGACTTTAGCCAAG
GeneScript   GCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTTGGGACTTTAGCCAAG
             ************************************************************

Sequenced    GGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCA
GeneScript   GGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCA
             ************************************************************

Sequenced    ACTCACACCCGAAATGCTCGAGCATGGATGTTTTGTATTCGCTTTCTAAGACTCTGAAGG
GeneScript   ACTCACACCCGAAATGCTCGAGCATGGATGTTTTGTATTCGCTTTCTAAGACTCTGAAGG
             ************************************************************

Sequenced    ATGCCCGAGATAAGATTGTTGAGGGAACTTTGTATTCTAATGTTTCTGATCTGATTCAGC
GeneScript   ATGCCCGAGATAAGATTGTTGAGGGAACTTTGTATTCTAATGTTTCTGATCTGATTCAGC
             ************************************************************

Sequenced    AGTTCAACCAGATGATTATCACTATGAACGGTAATGAGTTTCAGACCGGAGGTATTGGAA
GeneScript   AGTTCAACCAGATGATTATCACTATGAACGGTAATGAGTTTCAGACCGGAGGTATTGGAA
             ************************************************************

Sequenced    ACCTCCCTATCCGAAACTGGAATTTCGACTTTGGACTGCTTGGTACTACCCTCTTGAACT
GeneScript   ACCTCCCTATCCGAAACTGGAATTTCGACTTTGGACTGCTTGGTACTACCCTCTTGAACT
             ************************************************************

Sequenced    TGGATGCTAATTACGTGGAGACTGCCCGAAATACCATTGACTATTTCGTTGACTTTGTGG
GeneScript   TGGATGCTAATTACGTGGAGACTGCCCGAAATACCATTGACTATTTCGTTGACTTTGTGG
             ************************************************************

Sequenced    ATAACGTCTGCATGGATGAGATGGTCCGAGAGTCTCAGCGAAACGGTATTGCTCCTCAGT
GeneScript   ATAACGTCTGCATGGATGAGATGGTCCGAGAGTCTCAGCGAAACGGTATTGCTCCTCAGT
             ************************************************************

Sequenced    CGGACTCTCTCCGAAAGTTGTCCGGAATTAAGTTCAAGCGAATCAACTTCGATAATTCTT
GeneScript   CGGACTCTCTCCGAAAGTTGTCCGGAATTAAGTTCAAGCGAATCAACTTCGATAATTCTT
             ************************************************************

Sequenced    CCGAGTACATTGAGAACTGGAATCTGCAGAACCGACGACAGCGAACTGGATTCACCTTTC
GeneScript   CCGAGTACATTGAGAACTGGAATCTGCAGAACCGACGACAGCGAACTGGATTCACCTTTC
             ************************************************************

Sequenced    ACAAGCCTAACATCTTCCCCTATTCCGCCTCGTTTACTCTTAACCGATCTCAGCCTGCTC
GeneScript   ACAAGCCTAACATCTTCCCCTATTCCGCCTCGTTTACTCTTAACCGATCTCAGCCTGCTC
             ************************************************************

Sequenced    ATGACAATCTGATGGGAACCATGTGGCTTAACGCTGGTTCCGAGATTCAGGTTGCCGGAT
GeneScript   ATGACAATCTGATGGGAACCATGTGGCTTAACGCTGGTTCCGAGATTCAGGTTGCCGGAT
             ************************************************************
```

```
Sequenced    TTGATTACTCGTGTGCTATCAACGCTCCCGCCAATACTCAGCAGTTCGAGCACATTGTGC
GeneScript   TTGATTACTCGTGTGCTATCAACGCTCCCGCCAATACTCAGCAGTTCGAGCACATTGTGC
             ************************************************************


Sequenced    AGCTCCGACGAGTCTTGACTACCGCCACTATCACCCTGCTTCCTGACGCTGAGCGATTCT
GeneScript   AGCTCCGACGAGTCTTGACTACCGCCACTATCACCCTGCTTCCTGACGCTGAGCGATTCT
             ************************************************************


Sequenced    CTTTTCCCCGAGTTATTAATTCCGCTGATGGTGCCACTACCTGGTACTTCAACCCTGTGA
GeneScript   CTTTTCCCCGAGTTATTAATTCCGCTGATGGTGCCACTACCTGGTACTTCAACCCTGTGA
             ************************************************************


Sequenced    TCCTCCGACCCAACAATGTGGAGGTCGAGTTTCTCTTGAACGGTCAGATTATCAAT<mark>ACTT</mark>
GeneScript   TCCTCCGACCCAACAATGTGGAGGTCGAGTTTCTCTTGAACGGTCAGATTATCAAT<mark>ACTT</mark>
             ************************************************************


Sequenced    <mark>ATCAGGCCCGATTCG</mark>GAACCATTGTGGCTCGAAACTTCGACACTATCCGACTGTCTTTTC
GeneScript   <mark>ATCAGGCCCGATTCG</mark>GAACCATTGTGGCTCGAAACTTCGACACTATCCGACTGTCTTTTC
             ************************************************************


Sequenced    AGCTTATGCGACCTCCCAATATGACCCCTTCCGTCGCTGCCCTTTTCCCTAACGCCCAGC
GeneScript   AGCTTATGCGACCTCCCAATATGACCCCTTCCGTCGCTGCCCTTTTCCCTAACGCCCAGC
             ************************************************************


Sequenced    CCTTTGAGCACCATGCTACTGTTGGACTGACCCTTAAGATTGAGTCCGCCGTTTGCGAGT
GeneScript   CCTTTGAGCACCATGCTACTGTTGGACTGACCCTTAAGATTGAGTCCGCCGTTTGCGAGT
             ************************************************************


Sequenced    CGGTGCTCGCTGATGCCTCGGAGACTATGTTGGCCAACGTTACCTCTGTGCGACAGGAGT
GeneScript   CGGTGCTCGCTGATGCCTCGGAGACTATGTTGGCCAACGTTACCTCTGTGCGACAGGAGT
             ************************************************************


Sequenced    ACGCTATTCCTGTCGGACCCGTTTTTCCTCCCGGTATGAACTGGACTGACCTGATCACCA
GeneScript   ACGCTATTCCTGTCGGACCCGTTTTTCCTCCCGGTATGAACTGGACTGACCTGATCACCA
             ************************************************************


Sequenced    ATTATTCTCCTTCCCGAGAGGATAACCTTCAGCGAGTGTTCACTGTCGCTTCGATTCGA<mark>T</mark>
GeneScript   ATTATTCTCCTTCCCGAGAGGATAACCTTCAGCGAGTGTTCACTGTCGCTTCGATTCGA<mark>T</mark>
             ************************************************************


Sequenced    <mark>CTATGCTGGTCAAG</mark><mark style="background:red">TAG</mark><mark>AGC</mark>GCTATTAA<span style="color:red">TCCTAGGTGATCTGATCTGCTTACTTTACTTA</span>
GeneScript   <mark>CTATGCTGGTCAAG</mark><mark style="background:red">TAG</mark><mark>AGC</mark>GCTATTAA<span style="color:red">TCCTAGGTGATCTGATCTGCTTACTTTACTTA</span>
             *****************************************************


Sequenced    <span style="color:red">ACGACCAAAGAAAAACGACAAAAAAAAAAATATTACTACTATTAAAATAAATTAGTATTTT</span>
GeneScript   <span style="color:red">ACGACCAAAGAAAAACGACAAAAAAAAAAATATTACTACTATTAAAATAAATTAGTATTTT</span>
             ************************************************************


Sequenced    <span style="color:red">TCTCTTCTTACGATATGATATGATGCTATGAAATCATCATCTTCTTAACTTTCTTGTCTC</span>
GeneScript   <span style="color:red">TCTCTTCTTACGATATGATATGATGCTATGAAATCATCATCTTCTTAACTTTCTTGTCTC</span>
             ************************************************************


Sequenced    <span style="color:red">TTACACGTCACTTACTCTATATACCCGTTTATATAAGTGTACGTATTTTCTTTTTTTTAA</span>
GeneScript   <span style="color:red">TTACACGTCACTTACTCTATATACCCGTTTATATAAGTGTACGTATTTTCTTTTTTTTAA</span>
             ************************************************************


Sequenced    <span style="color:red">AAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCAGTTCCAACGCACTCTAGCTG</span>
GeneScript   <span style="color:red">AAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCAGTTCCAACGCACTCTAGCTG</span>
             ************************************************************


Sequenced    <span style="color:red">TTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCACCCCAAAAATAACTTTCTTC</span>
GeneScript   <span style="color:red">TTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCACCCCAAAAATAACTTTCTTC</span>
             ************************************************************


Sequenced    <span style="color:red">CCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCATTTTCGCATCTGAAACTAATT</span>
GeneScript   <span style="color:red">CCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCATTTTCGCATCTGAAACTAATT</span>
             ************************************************************
```

```
Sequenced    CTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCATCACTAGAAATGAGAAAAAGA
GeneScript   CTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCATCACTAGAAATGAGAAAAAGA
             ************************************************************

Sequenced    TCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCGGATCCTAGGGATAACAGGGT
GeneScript   TCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCGGATCCTAGGGATAACAGGGT
             ************************************************************

Sequenced    AATGGTACCAGAGACCGGGTTGCGCGTATTTGTGTCCCAAAAAACAGCCCAATTGCCCAAT
GeneScript   AATGG-------------------------------------------------------
             *****
```

Figure 8: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *A. adeninivorans* codon-optimized VP6 cassette ORF to *in silico* sequences from GeneScript. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *I-SceI* restriction enzyme recognition site = underlined, *Yarrowia lypolytica* TEF promoter = blue text and *Kluyveromyces marxianus* inulase terminator = red text.

```
Sequenced   TGGGCCTCCATGTCGCTGGCCGGGTGACCCGGCGGGGACGAGGCAAGCTAAACAGATCTA
GeneArt     -----------------GCCGGGTGACCCGGCGGGGACGAGGCAAGCTAAACAGATCTA
                             ****************************************


Sequenced   GGGATAACAGGGTAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAAC
GeneArt     GGGATAACAGGGTAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCCAAAAAAC
            ************************************************************


Sequenced   AGCCCCAATTGCCCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGA
GeneArt     AGCCCCAATTGCCCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCCGGCGAGA
            ************************************************************


Sequenced   GCCCCCTTCACCCCACATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTA
GeneArt     GCCCCCTTCACCCCACATATCAAACCTCCCCCGGTTCCCACACTTGCCGTTAAGGGCGTA
            ************************************************************


Sequenced   GGGTACTGCAGTCTGGAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCC
GeneArt     GGGTACTGCAGTCTGGAATCTACGCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCC
            ************************************************************


Sequenced   ATCCGGGTAACCCATGCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTT
GeneArt     ATCCGGGTAACCCATGCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTTTGTGGTT
            ************************************************************


Sequenced   GGGACTTTAGCCAAGGGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTC
GeneArt     GGGACTTTAGCCAAGGGTATAAAAGACCACCGTCCCCGAATTACCTTTCCTCTTCTTTTC
            ************************************************************


Sequenced   TCTCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGATGGACGTTTTGTACTCTTTGTC
GeneArt     TCTCTCTCCTTGTCAACTCACACCCGAAATGCTCGAGATGGACGTTTTGTACTCTTTGTC
            ************************************************************


Sequenced   TAAGACCTTGAAGGACGCTAGAGACAAGATCGTTGAAGGCACCTTGTACTCTAACGTTTC
GeneArt     TAAGACCTTGAAGGACGCTAGAGACAAGATCGTTGAAGGCACCTTGTACTCTAACGTTTC
            ************************************************************


Sequenced   TGACTTGATCCAACAATTCAACCAAATGATCATCACCATGAACGGTAACGAGTTCCAAAC
GeneArt     TGACTTGATCCAACAATTCAACCAAATGATCATCACCATGAACGGTAACGAGTTCCAAAC
            ************************************************************


Sequenced   CGGTGGTATCGGTAACTTGCCAATCAGAAACTGGAATTTCGACTTCGGTTTGTTGGGTAC
GeneArt     CGGTGGTATCGGTAACTTGCCAATCAGAAACTGGAATTTCGACTTCGGTTTGTTGGGTAC
            ************************************************************


Sequenced   TACCTTGTTGAACTTGGACGCTAACTACGTTGAAACCGCTAGAAACACCATCGACTACTT
GeneArt     TACCTTGTTGAACTTGGACGCTAACTACGTTGAAACCGCTAGAAACACCATCGACTACTT
            ************************************************************


Sequenced   CGTTGACTTCGTTGACAACGTTTGTATGGACGAAATGGTTAGAGAATCTCAAAGAAACGG
GeneArt     CGTTGACTTCGTTGACAACGTTTGTATGGACGAAATGGTTAGAGAATCTCAAAGAAACGG
            ************************************************************


Sequenced   TATCGCTCCACAATCTGACTCTTTGAGAAAGTTGTCTGGTATCAAGTTCAAGAGAATCAA
GeneArt     TATCGCTCCACAATCTGACTCTTTGAGAAAGTTGTCTGGTATCAAGTTCAAGAGAATCAA
            ************************************************************


Sequenced   CTTCGACAACTCTTCTGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAAC
GeneArt     CTTCGACAACTCTTCTGAATACATCGAAAACTGGAACTTGCAAAACAGAAGACAAAGAAC
            ************************************************************


Sequenced   CGGTTTCACCTTCCACAAGCCAAACATCTTCCCATACTCTGCTTCTTTCACCTTGAACAG
GeneArt     CGGTTTCACCTTCCACAAGCCAAACATCTTCCCATACTCTGCTTCTTTCACCTTGAACAG
            ************************************************************


Sequenced   ATCCCAACCAGCTCACGACAACTTGATGGGTACTATGTGGTTGAACGCTGGTTCTGAAAT
GeneArt     ATCCCAACCAGCTCACGACAACTTGATGGGTACTATGTGGTTGAACGCTGGTTCTGAAAT
            ************************************************************
```

119

```
Sequenced    CCAAGTTGCTGGTTTCGACTACTCTTGTGCTATCAACGCTCCAGCTAACACCCAACAATT
GeneArt      CCAAGTTGCTGGTTTCGACTACTCTTGTGCTATCAACGCTCCAGCTAACACCCAACAATT
             ************************************************************


Sequenced    CGAACACATCGTTCAATTGAGAAGAGTTTTGACCACCGCTACCATCACCTTGTTGCCAGA
GeneArt      CGAACACATCGTTCAATTGAGAAGAGTTTTGACCACCGCTACCATCACCTTGTTGCCAGA
             ************************************************************


Sequenced    CGCTGAAAGATTCTCTTTCCCAAGAGTTATCAACTCTGCTGACGGTGCTACCACCTGGTA
GeneArt      CGCTGAAAGATTCTCTTTCCCAAGAGTTATCAACTCTGCTGACGGTGCTACCACCTGGTA
             ************************************************************


Sequenced    CTTCAACCCAGTTATCTTGAGACCAAACAACGTTGAAGTTGAGTTCTTGTTGAACGGTCA
GeneArt      CTTCAACCCAGTTATCTTGAGACCAAACAACGTTGAAGTTGAGTTCTTGTTGAACGGTCA
             ************************************************************


Sequenced    AATCATCAACACCTACCAAGCTAGATTCGGCACCATCGTTGCTAGAAACTTCGACACCAT
GeneArt      AATCATCAACACCTACCAAGCTAGATTCGGCACCATCGTTGCTAGAAACTTCGACACCAT
             ************************************************************


Sequenced    CAGATTGTCTTTCCAATTGATGAGACCACCAAACATGACCCCATCTGTTGCTGCTTTGTT
GeneArt      CAGATTGTCTTTCCAATTGATGAGACCACCAAACATGACCCCATCTGTTGCTGCTTTGTT
             ************************************************************


Sequenced    CCCAAACGCTCAACCATTCGAACACCACGCTACCGTTGGTTTGACCTTGAAGATCGAATC
GeneArt      CCCAAACGCTCAACCATTCGAACACCACGCTACCGTTGGTTTGACCTTGAAGATCGAATC
             ************************************************************


Sequenced    TGCTGTTTGTGAATCTGTTTTGGCTGACGCTTCTGAAACCATGTTGGCTAACGTTACCTC
GeneArt      TGCTGTTTGTGAATCTGTTTTGGCTGACGCTTCTGAAACCATGTTGGCTAACGTTACCTC
             ************************************************************


Sequenced    TGTTAGACAAGAATACGCTATCCCAGTTGGTCCAGTTTTCCCACCAGGTATGAACTGGAC
GeneArt      TGTTAGACAAGAATACGCTATCCCAGTTGGTCCAGTTTTCCCACCAGGTATGAACTGGAC
             ************************************************************


Sequenced    CGACTTGATCACCAACTACTCTCCATCCAGAGAAGACAACTTGCAAAGAGTTTTCACCGT
GeneArt      CGACTTGATCACCAACTACTCTCCATCCAGAGAAGACAACTTGCAAAGAGTTTTCACCGT
             ************************************************************


Sequenced    TGCTTCTATCAGATCCATGTTGGTTAAGTAAAGCGCTTGCGCTATTAATCCTAGGTGATC
GeneArt      TGCTTCTATCAGATCCATGTTGGTTAAGTAAAGCGCTTGCGCTATTAATCCTAGGTGATC
             ************************************************************


Sequenced    TGATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAAAAAAAATATTACTACTA
GeneArt      TGATCTGCTTACTTTACTTAACGACCAAAGAAAAACGACAAAAAAAAAATATTACTACTA
             ************************************************************


Sequenced    TTAAAATAAATTAGTATTTTTCTCTTCTTACGATATGATATGATGCTATGAAATCATCAT
GeneArt      TTAAAATAAATTAGTATTTTTCTCTTCTTACGATATGATATGATGCTATGAAATCATCAT
             ************************************************************


Sequenced    CTTCTTAACTTTCTTGTCTCTTACACGTCACTTACTCTATATACCCGTTTATATAAGTGT
GeneArt      CTTCTTAACTTTCTTGTCTCTTACACGTCACTTACTCTATATACCCGTTTATATAAGTGT
             ************************************************************


Sequenced    ACGTATTTTCTTTTTTTTAAAAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCA
GeneArt      ACGTATTTTCTTTTTTTTAAAAAATTTCTATTCTATCCTTAGAAAAGTGCCCTTACATCA
             ************************************************************


Sequenced    GTTCCAACGCACTCTAGCTGTTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCA
GeneArt      GTTCCAACGCACTCTAGCTGTTCTGGCACTGTATCTTCATCATGTGCCGGTCGTTTTCCA
             ************************************************************


Sequenced    CCCCAAAAATAACTTTCTTCCCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCAT
GeneArt      CCCCAAAAATAACTTTCTTCCCTTTTCCTTTCAATTAATGGCCTGGAATTCCGAACCCAT
             ************************************************************
```

120

```
Sequenced    TTTCGCATCTGAAACTAATTCTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCAT
GeneArt      TTTCGCATCTGAAACTAATTCTCGAAACCTTTAATATCAAACAATTGAAAAGATCATCAT
             ************************************************************

Sequenced    CACTAGAAATGAGAAAAAGATCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCG
GeneArt      CACTAGAAATGAGAAAAAGATCAACAGCACTTAATAACAGTACGAAAGAAAGATCGCTCG
             ************************************************************

Sequenced    GATCCTAGGGATAACAGGGTAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCC
GeneArt      GATCCTAGGGATAACAGGGTAATGGTACCAGAGACCGGGTTGGCGGCGTATTTGTGTCCC
             ************************************************************

Sequenced    AAAAAACAGCCCCAATTGCCCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCCC
GeneArt      AAAAAACAGC--------------------------------------------------
             **********
```

Figure 9: CLUSTALX 2.1 multiple sequence alignment comparing sequenced *P. pastoris/H. polymorpha* codon-optimized VP6 cassette to *in silico* sequences from GeneArt. Primer binding site: Forward primers = yellow and Reverse primers = dark yellow, Open reading frame stop codon = red, Open reading frame start codon = green and *I-SceI* restriction enzyme recognition site = underlined, *Yarrowia lipolytica* TEF promoter = blue text and *Kluyveromyces marxianus* inulase terminator = red text.

APPENDIX B ClustalX Multiple sequence alignment of optimized and wild-type VP2 and VP6 rotavirus ORFs nucleotide sequences

```
KO VP2          ATGGCTTACAGAAAGAGAGGAGCAAGAAGAGAAGCAAACTTAAATAACAACGATAGAATG
PO VP2          ATGGCTTACAGAAAGAGAGGTGCTAGAAGAGAAGCTAACTTGAACAACAACGACAGAATG
Wild-type VP2   ATGGCGTACAGGAAACGTGGAGCGCGCCGTGAGGCGAACTTAAATAATAATGACCGAATG
AO VP2          ATGGCCTACCGAAAGCGAGGTGCCCGACGAGAGGCTAACTTGAATAATAATGACCGAATG
                ***** *** * **   * ** **   *  * ** ** ***** ** ** ** **  *****

KO VP2          CAAGAAAAGATTGATGAAAAACAAGATAGTAATAAGATCCAATTGAGTGATAAGGTTTTG
PO VP2          CAAGAAAAGATCGACGAAAAGCAAGACTCTAACAAGATCCAATTGTCTGACAAGGTTTTG
Wild-type VP2   CAGGAGAAAATTGATGAAAAACAAGATTCAAATAAAATACAATTATCCGATAAGGTACTT
AO VP2          CAGGAGAAGATTGACGAGAAGCAGGATTCTAACAAGATTCAGCTCTCCGATAAGGTTTTG
                ** ** ** ** ** ** ** ** ** **      ** ** ** **   ** ***** *

KO VP2          TCCAAAAAGGAAGAAATCGTGACTGATTCCCATGAAGAAGTTAAAGTGACTGATGAATTG
PO VP2          TCTAAGAAGGAAGAGATCGTTACCGACTCTCACGAAGAAGTTAAGGTTACCGACGAATTG
Wild-type VP2   TCGAAGAAAGAAGAAATTGTAACGGATAGTCATGAGGAAGTTAAAGTTACTGATGAGTTA
AO VP2          TCGAAGAAGGAGGAGATCGTGACTGATTCCCACGAGGAGGTTAAGGTTACTGACGAGCTG
                ** ** ** ** ** ** ** ** ** **      ** ** ** ***** ** ** ** **   *

KO VP2          AAAAAGTCTACAAAGGAAGAATCAAAGCAATTGTTGGAAGTTTTGAAGACAAAGGAAGAA
PO VP2          AAGAAGTCTACCAAGGAAGAATCTAAGCAATTGTTGGAAGTTTTGAAGACCAAGGAAGAA
Wild-type VP2   AAAAAAATCAACGAAAGAAGAATCAAAAACAATTGCTTGAAGTGTTGAAAACAAAGGAAGAA
AO VP2          AAGAAGTCTACCAAGGAGGAGTCCAAGCAGCTGCTTGAGGTGCTTAAGACCAAGGAGGAG
                ** ** ** ** ** ** ** ** ** **  ** * ** **   * ** ** ***** **

KO VP2          CATCAAAAGGAAATCCAATACGAAATCTTGCAAAAGACTATACCAACCTTCGAACCTAAG
PO VP2          CACCAAAAGGAAATCCAATACGAAATCTTGCAAAAGACCATCCCAACCTTCGAACCTAAG
Wild type VP2   CATCAGAAAGAAATACAGTATGAAATATTACAGAAAACTATACCAACATTCGAACCTAAA
AO VP2          CATCAGAAGGAGATTCAGTACGAGATCCTCCAGAAGACTATTCCTACCTTCGAGCCCAAG
                ** ** ** ** ** ** ** ** ** **  * ** ** ** ** ** ** ** ***** ** **

KO VP2          GAAACAATCTTGAGAAAGTTGGAAGATATTCAACCAGAATTGGCTAAGAAACAAACCAAG
PO VP2          GAAACCATCTTGAGAAAGTTGGAAGACATCCAACCAGAATTGGCTAAGAAGCAAACCAAG
Wild-type VP2   GAGACGATATTGAGAAAATTAGAGGATATTCAACCAGAACTAGCGAAAAAACAGACTAAG
AO VP2          GAGACTATTCTGCGAAAGCTTGAGGACATCCAGCCTGAGCTGGCCAAGAAGCAGACCAAG
                ** ** ** **   ** ****  * ** ** ** ** ** **   * ** ** ** ** ** ***

KO VP2          TTATTCAGAATCTTCGAACCAAAGCAATTGCCTATCTACAGAGCTAATGGTGAAAGAGAA
PO VP2          TTGTTCAGAATCTTCGAACCAAAGCAATTGCCAATCTACAGAGCTAACGGTGAAAGAGAA
Wild-type VP2   TTATTTAGAATATTTGAACCGAAACAATTACCGATTTATAGAGCAAATGGAGAGAGAGAA
AO VP2          CTTTTCCGAATTTTTGAGCCTAAGCAGCTCCCCATCTATCGAGCTAACGGAGAGAGCGAGAG
                 * **  ****  ** ** ** ** ** ** ** ** * ** ** ** ** ** **   ****

KO VP2          TTGAGAAACAGATGGTACTGGAAGTTGAAAAAGGATACCTTACCAGATGGAGATTATGAT
PO VP2          TTGAGAAACAGATGGTACTGGAAGTTGAAGAAGGACACCTTGCCAGACGGTGACTACGAC
Wild-type VP2   TTGCGTAATAGATGGTATTGGAAATTAAAAAAAAGATACACTACCAGACGGAGACTATGAT
AO VP2          TTGCGAAATCGATGGTACTGGAAGCTGAAGAAGGATACTCTTCCTGACGGAGACTACGAC
                *** * ** ******* ***** * ** ** ** ** * ** ** ** ** ** ** **

KO VP2          GTCAGAGAATACTTTTTGAATTTGTACGATCAAGTATTGACAGAAATGCCTGATTACTTG
PO VP2          GTTAGAGAGTACTTCTTGAACCTCTACGACCAAGTCTTGACCGAAATGCCAGACTACTTG
Wild-type VP2   GTGAGAGAGTATTTTCTGAATTTGTATGATCAAGTGCTTACTGAAATGCCAGACTACTTA
AO VP2          GTCCGAGAGTATTTTCTGAACCTTTACGACCAGGTTCTGACCGAGATGCCCGATTATCTC
                **  **** ** **  ****  * ** ** ** **   * ** ** ***** ** **   *

KO VP2          TTGTTGAAGGATATGGCAGTTGAAAACAAGAATTCTAGAGATGCTGGTAAAGTTGTGGAT
PO VP2          TTGTTGAAGGACATGGCTGTTGAAAACAAGAACTCCAGAGACGCTGGTAAGGTTGTTGAC
Wild-type VP2   TTATTGAAAGATATGGCAGTAGAAAAATAAGAACTCTAGGGATGCAGGTAAAGTTGTTGAC
AO VP2          TTGCTGAAGGACATGGCCGTCGAGAATAAGAACTCTCGAGATGCTGGTAAAGGTGGTCGAC
                ** **** ** ***** ** ** ** ** ***** ** * ** ** ** ** ** ** **

KO VP2          TCTGAAACTGCATCAATTTGTGATGCTATCTTCCAAGATGAAGAAACAGAAGGAGCAGTT
```

```
PO VP2          TCTGAAACCGCTTCTATCTGTGACGCTATCTTCCAAGACGAAGAAACCGAAGGTGCTGTT
Wild-type VP2   TCAGAAACGGCTAGTATATGCGATGCCATATTTCAAGATGAAGAAACGGAAGGTGCCGTT
AO VP2          TCGGAGACTGCCTCTATTTGCGATGCTATCTTCCAGGACGAGGAGACCGAGGGTGCCGTG
                ** ** ** **       ** ** ** ** ** ** ** ** ** ** ** ** ** ** **


KO VP2          AGAAGATTCATTGCTGAAATGAGACAAAGAGTGCAAGCAGATAGAAACGTCGTAAACTAC
PO VP2          AGAAGATTCATCGCTGAAATGAGACAAAGAGTTCAAGCTGACAGAAACGTTGTTAACTAC
Wild-type VP2   AGAAGATTCATTGCAGAAATGAGACAACGTGTGCAAGCTGATAGAAATGTTGTCAATTAT
AO VP2          CGACGATTCATTGCTGAGATGCGACAGCGAGTCCAGGCCGATCGAAACGTTGTGAATTAC
                 ** ******* ** ** *** ****  * ** ** ** ** **   **** ** ** ** **


KO VP2          CCATCAATCTTGCATCCTATCGATTACGCTTTCAATGAATATTTCTTGCAACATCAATTG
PO VP2          CCATCTATCTTGCACCCAATCGATTACGCTTTCAACGAATACTTCTTGCAACACCAACTT
Wild-type VP2   CCATCAATATTACATCCAATAGATTATGCATTTAATGAATACTTTTTACAACATCAATTG
AO VP2          CCTTCCATTCTGCACCCCATCGACTATGCTTTCAACGAGTACTTTTTGCAGCATCAGCTG
                ** ** **   * ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **  *


KO VP2          GTTGAACCATTGAACAACGATATAATCTTCAACTACATCCCTGAAAGAATCAGAAACGAT
PO VP2          GTCGAACCATTGAACAACGACATCATCTTCAACTACATCCCAGAAAGAATCAGAAACGAC
Wild-type VP2   GTTGAACCATTGAATAATGATATAATATTTAATTATATACCAGAAAGGATAAGAAATGAT
AO VP2          GTTGAGCCTCTTAACAACGATATCATCTTCAACTACATTCCCGAGCGAATCCGAAACGAT
                ** ** **   * ** ** ** ** ** ** ** ** ** ** **   * **   **** **


KO VP2          GTGAACTACATATTGAATATGGATAGAAACTTACCAAGTACTGCAAGATACATCAGACCT
PO VP2          GTTAACTATATCTTGAACATGGACAGAAACTTGCCATCTACCGCTAGATACATCAGACCA
Wild-type VP2   GTTAATTATATTCTCAATATGGACAGAAATTTACCATCAACTGCCAGATATATAAGACCT
AO VP2          GTGAATTACATTCTCAACATGGACCGAAATTTGCCTTCTACTGCCCGATATATCCGACCC
                ** ** ** **  * ** *****  **** ** **   ** **  **** **   ****


KO VP2          AATTTGTTGCAAGATAGATTGAATTTGCATGATAACTTCGAATCATTGTGGGATACAATC
PO VP2          AACTTGTTGCAAGACAGATTGAACTTGCACGACAACTTCGAATCTTTGTGGGACACCATC
Wild-type VP2   AATTTACTTCAAGATAGATTAAATTTGCACGATAATTTTGAATCACTATGGGATACAATA
AO VP2          AACCTTCTCCAGGATCGACTCAACTTGCACGACAATTTTGAGTCCCTGTGGGATACCATT
                **   *   * ** **   ** * ** ***** ** ** ** ** **   * ***** ** **


KO VP2          ACTACAAGTAACTACATATTGGCTAGATCCGTTGTGCCAGATTTGAAGGAATTAGTTAGT
PO VP2          ACCACCTCTAACTACATCTTGGCTAGATCCGTTGTTCCAGACTTGAAGGAATTGGTTTCT
Wild-type VP2   ACTACATCAAATTATATTTTGGCGAGATCGGTAGTACCAGATTTAAAGGAATTAGTGTCA
AO VP2          ACTACCTCTAACTACATCCTTGCTCGATCCGTCGTTCCTGACCTCAAGGAGTTGGTTTCC
                ** **        ** ** **  * **   **** ** ** ** **   * ***** ** **


KO VP2          ACTGAAGCACAAATACAAAAGATGTCCCAAGATTTGCAATTGGAAGCTTTGACAATCCAA
PO VP2          ACCGAAGCTCAAATCCAAAAGATGTCTCAAGACTTGCAATTGGAAGCCTTGACCATCCAA
Wild-type VP2   ACGGAAGCACAAATTCAGAAAATGTCACAAGATTTGCAATTAGAAGCATTAACAATTCAG
AO VP2          ACTGAGGCCCAGATTCAGAAGATGTCGCAGGACCTGCAGCTTGAGGCTCTCACCATTCAG
                ** ** ** ** ** ** ** ** ***** ** **   **   * ** **   * ** ** **


KO VP2          AGTGAAACCCAATTCTTAACTGGTATTAATTCCCAAGCTGCAAACGATTGTTTCAAGACA
PO VP2          TCTGAAACCCAATTCTTGACCGGTATCAACTCTCAAGCTGCTAACGACTGTTTCAAGACC
Wild-type VP2   TCAGAAACACAATTTCTAACAGGTATAAATTCACAAGCAGCTAACGATTGTTTTAAAACC
AO VP2          TCGGAGACTCAGTTCTTGACCGGAATCAACTCTCAGGCTGCCAATGATTGCTTTAAGACT
                 ** ** ** **   * ** ** ** ** ** ** ** ** ** ** ** ** ** ** **


KO VP2          TTGATAGCTGCAATGTTAAGTCAAAGAACCATGTCCTTGGATTTCGTTACCACTAACTAC
PO VP2          TTGATCGCTGCTATGTTGTCTCAAAGAACCATGTCTTTGGACTTCGTTACCACCAACTAC
Wild-type VP2   TTAATTGCAGCAATGTTAAGTCAACGTACTATGTCATTAGATTTTGTAACTACTAATTAT
AO VP2          CTCATTGCTGCCATGTTGTCGCAGCGAACCATGTCTCTGGACTTCGTGACTACCAACTAC
                 * ** ** ** *****       ** * ** ***** * ** ** ** ** ** ** **


KO VP2          ATGTCTTTGATCTCAGGAATGTGGTTATTGACTGTCGTACCAAACGATATGTTCATCAGA
PO VP2          ATGTCTTTGATCTCTGGTATGTGGTTGTTGACCGTCGTTCCAAACGACATGTTCATCAGA
Wild-type VP2   ATGTCATTGATTTCAGGTATGTGGCTATTGACGGTTGTGCCAAATGATATGTTTATAAGG
AO VP2          ATGTCCCTTATTTCGGGAATGTGGTTGCTGACTGTGGTCCCTAACGATATGTTCATTCGA
                *****   * ** ** ** ****** *   **** ** ** ** ** ** ***** **  *


KO VP2          GAATCTTTGGTTGCTTGTCAATTGGCAATCGTGAACACTATCATCTATCCAGCTTTCGGT
PO VP2          GAATCTTTGGTTGCTTGTCAATTGGCTATCGTTAACACCATCATCTACCCAGCTTTCGGT
```

123

```
Wild-type VP2    GAATCGTTAGTCGCGTGTCAACTAGCTATAGTAAATACAATAATCTATCCAGCATTTGGA
AO VP2           GAGTCTCTGGTCGCTTGTCAGCTTGCCATCGTTAACACTATTATCTATCCTGCTTTTGGA
                 ** **   * ** **  *****   * ** ** ** ** ** ** ** ***** ** ** ** **


KO VP2           ATGCAAAGAATGCATTACAGAAACGGAGATCCACAAACACCTTTCCAAATAGCAGAACAA
PO VP2           ATGCAAAGAATGCACTACAGAAACGGTGACCCACAAACCCCATTCCAAATCGCTGAACAA
Wild-type VP2    ATGCAACGAATGCATTATAGAAACGGGGATCCACAAACACCGTTTCAGATAGCAGAACAG
AO VP2           ATGCAGCGAATGCACTACCGAAACGGTGACCCTCAGACCCCCTTCCAGATTGCCGAGCAG
                 *****   ******* **   ******* ** ** ** ** ** ** ** ** ** ** **


KO VP2           CAAATCCAAAACTTCCAAGTCGCTAACTGGTTGCATTTCGTAAACAACAACCAATTCAGA
PO VP2           CAAATCCAAAACTTCCAAGTTGCTAACTGGTTGCACTTCGTTAACAACAACCAATTCAGA
Wild-type VP2    CAGATTCAAAATTTCCAAGTCGCAAATTGGTTACATTTTGTTAATAATAATCAATTTAGA
AO VP2           CAGATCCAGAACTTTCAGGTGGCTAATTGGCTTCATTTCGTCAACAATAACCAGTTTCGA
                 ** ** ** ** ** ** ** ** ** ** *** * ** ** ** ** ** ** ** ** **


KO VP2           CAAGCAGTTATTGATGGTGTGTTGAACCAAGTCTTGAACGATAACATCAGAAACGGACAT
PO VP2           CAAGCTGTTATCGACGGTGTTTTGAACCAAGTTTTGAACGACAACATCAGAAACGGTCAC
Wild-type VP2    CAGGCAGTTATTGATGGTGTATTGAATCAGGTACTGAATGACAATATTAGAAATGGTCAT
AO VP2           CAGGCCGTCATTGATGGAGTTCTCAACCAGGTGTTGAATGACAACATTCGAAACGGTCAC
                 ** ** ** ** ** ** **   * ** ** **   **** ** ** **   **** ** **


KO VP2           GTTATAAACCAATTGATGGAAGCTTTGATGCAATTATCTAGACAACAATTCCCAACTATG
PO VP2           GTTATCAACCAATTGATGGAAGCCTTGATGCAATTGTCCAGACAACAATTCCCAACCATG
Wild-type VP2    GTTATTAACCAACTGATGGAGGCTCTAATGCAGCTGTCGCGACAACAATTTCCAACCATG
AO VP2           GTGATCAATCAGCTGATGGAGGCTCTGATGCAGCTTTCTCGACAGCAGTTTCCTACTATG
                 ** ** ** **   ******* **   * ** ***   * ** **   * ** ** ** ** ***


KO VP2           CCTATCGATTACAAGAGATCTATCCAAAGAGGTATCTTGTTGTTGTCAAACAGATTGGGA
PO VP2           CCAATCGACTACAAGAGATCCATCCAAAGAGGTATCTTGTTGTTGTCTAACAGATTGGGT
Wild-type VP2    CCAATTGATTATAAGAGATCAATTCAACGTGGAATATTACTGTTATCTAACAGACTTGGT
AO VP2           CCCATTGATTACAAGCGATCTATTCAGCGAGGAATCCTTCTCTTGTCCAACCGACTCGGA
                 ** ** ** ** ** ***  ****  ** **   * ** **   * * ** ** *** ** * **


KO VP2           CAATTGGTTGATTTGACTAGATTGTTGGCATACAACTACGAAACATTGATGGCTTGTATT
PO VP2           CAATTGGTTGACTTGACCAGATTGTTGGCTTACAACTACGAAACCTTGATGGCTTGTATC
Wild-type VP2    CAGTTAGTTGATTTAACTAGATTATTAGCTTACAATTATGAGACATTAATGGCATGCATT
AO VP2           CAGTTGGTGGACCTGACTCGACTGCTTGCCTACAACTATGAGACTCTCATGGCTTGCATT
                 ** ** ** **   * ** ** * *   * ** ***** ** ** **   * ***** ** **


KO VP2           ACCATGAACATGCAACATGTGCAAACATTGACAACCGAAAAGTTGCAATTGACTAGTGTC
PO VP2           ACCATGAACATGCAACACGTTCAAACCTTGACCACCGAAAAGTTGCAATTGACCTCTGTT
Wild-type VP2    ACAATGAACATGCAACATGTTCAAACCTTAACAACAGAAAAATTACAATTAACGTCAGTT
AO VP2           ACCATGAATATGCAGCATGTCCAGACCTTGACTACCGAGAAGCTCCAGTTGACTTCTGTT
                 ** ***** ***** ** ** ** ** ** ** ** **   * ** ** **     **


KO VP2           ACATCCTTATGTATGTTGATTGGTAATGCTACCGTAATCCCATCTCCTCAAACTTTGTTC
PO VP2           ACCTCTTTGTGTATGTTGATCGGTAACGCTACCGTTATCCCATCTCCACAAACCTTGTTC
Wild-type VP2    ACATCATTATGTATGCTTATTGGAAATGCGACTGTTATACCAAGTCCACAAACATTATTT
AO VP2           ACCTCCCTTTGTATGCTCATTGGAAACGCCACTGTGATCCCTTCCCCCCAGACCCTGTTC
                 ** **   * ****** * ** ** ** ** ** **     ** ** **   * **


KO VP2           CATTACTACAACGTCAACGTAAACTTCCATTCAAATTACAACGAAAGAATTAACGATGCT
PO VP2           CACTACTACAACGTTAACGTTAACTTCCACTCTAACTACAACGAAAGAATCAACGACGCT
Wild-type VP2    CATTATTATAACGTTAACGTTAATTTTCATTCAAATTACAATGAGAGAATTAATGATGCA
AO VP2           CACTACTACAACGTCAACGTTAACTTCCATTCGAACTACAACGAGCGAATTAACGATGCT
                 ** ** ** ***** ***** ** ** ** ** ** ** ***** **   **** ** ** **


KO VP2           GTCGCAATCATCACTGCTGCAAACAGATTGAACTTATACCAAAAGAAAATGAAGGCTATC
PO VP2           GTTGCTATCATCACCGCTGCTAACAGATTGAACTTGTACCAAAAGAAGATGAAGGCTATC
Wild-type VP2    GTAGCTATAATAACTGCTGCTAACAGACTGAATCTATATCAGAAAAAAATGAAGGCTATT
AO VP2           GTCGCCATTATCACTGCTGCCAACCGACTGAATCTTTACCAGAAGAAGATGAAGGCTATT
                 ** ** ** ** ** ** ***** *** ** ****   * ** ** ** ** **********


KO VP2           GTTGAAGATTTCTTAAAGAGATTGTACATCTTTGATGTTTCTAGAGTGCCAGATGATCAA
PO VP2           GTTGAAGACTTCTTGAAGAGATTGTACATCTTCGACGTTTCCAGAGTTCCAGACGACCAA
Wild-type VP2    GTTGAGGATTTCTTAAAAAGATTATACATTTTTGATGTATCTAGAGTTCCGGACGACCAA
```

124

```
AO VP2          GTTGAGGATTTCTTGAAGCGACTGTATATCTTTGACGTGTCTCGAGTCCCTGACGATCAG
                ***** ** ***** **   ** * ** ** ** ** ** ** **   **** ** ** ** **


KO VP2          ATGTACAGATTGAGAGATAGATTGAGATTGTTGCCTGTTGAAATCAGAAGATTGGATATA
PO VP2          ATGTACAGATTGAGAGACAGATTGAGATTGTTGCCAGTTGAAATCAGAAGATTGGACATC
Wild-type VP2   ATGTATAGATTAAGGGATAGATTACGCTTATTGCCAGTAGAAATCAGAAGATTAGATATC
AO VP2          ATGTACCGACTGCGAGATCGACTTCGACTCTTGCCCGTCGAGATTCGACGACTGGACATT
                *****   ** *   * **  ** *   *  * ***** ** ** **   ** ** * ** **


KO VP2          TTCAATTTGATCTTGATGAACATGGATCAAATCGAAAGAGCATCAGATAAGATAGCTCAA
PO VP2          TTCAACTTGATCTTGATGAACATGGACCAAATCGAAAGAGCTTCTGACAAGATCGCTCAA
Wild-type VP2   TTCAATCTAATACTAATGAACATGGATCAAATTGAACGTGCCTCAGATAAAATTGCTCAA
AO VP2          TTCAACCTCATCTTGATGAATATGGATCAGATTGAGCGAGCCTCCGACAAGATCGCTCAG
                *****   * **   * ***** ***** ** ** **   * ** ** ** ** ** *****


KO VP2          GGTGTTATTATCGCATACAGAGATATGCATTTGGAAAGAGATGAAATGTATGGTTACGTG
PO VP2          GGTGTTATCATCGCTTACAGAGACATGCACTTGGAAAGAGACGAAATGTACGGTTACGTT
Wild-type VP2   GGTGTAATCATTGCTTATCGTGACATGCATCTTGAAAGAGATGAGATGTACGGATATGTA
AO VP2          GGAGTCATTATCGCCTATCGAGATATGCACTTGGAGCGAGACGAGATGTACGGATATGTT
                ** ** ** ** ** ** **   * ** *****   * **  **** ** ***** ** ** **


KO VP2          AATATTGCTAGAAACTTGGAAGGATTCCAACAAATTAATTTGGAAGAATTGATGAGATCT
PO VP2          AACATCGCTAGAAACTTGGAAGGTTTCCAACAAATCAACTTGGAAGAATTGATGAGATCC
Wild-type VP2   AATATAGCTAGAAATTTAGAGGGATTTCAACAGATAAATTTAGAGGAGCTGATGAGATCA
AO VP2          AACATTGCTCGAAATCTTGAGGGTTTTCAGCAGATCAACCTTGAGGAGCTCATGCGATCT
                ** ** *** ****  * ** ** ** ** ** ** **   * ** **  * *** ****


KO VP2          GGAGATTACGCTCAAATCACCAACATGTTGTTGAACAACCAACCAGTCGCTTTGGTAGGA
PO VP2          GGTGACTACGCTCAAATCACCAACATGTTGTTGAACAACCAACCAGTTGCTTTGGTTGGT
Wild-type VP2   GGTGACTATGCGCAAATAACTAACATGCTTTTGAATAATCAACCAGTAGCATTGGTTGGA
AO VP2          GGAGATTACGCCCAGATTACTAACATGCTGCTTAATAACCAGCCTGTTGCTCTCGTGGGT
                ** ** ** ** ** ** ** ** ****** *   * ** ** ** ** ** **   * ** **


KO VP2          GCATTACCTTTCATCACTGATTCTTCAGTTATATCATTGATCGCTAAGTTAGATGCAACC
PO VP2          GCTTTGCCATTCATCACCGACTCTTCTGTTATCTCTTTGATCGCTAAGTTGGACGCTACC
Wild-type VP2   GCACTTCCATTTATTACTGATTCATCAGTTATATCGCTAATAGCAAAACTTGACGCTACA
AO VP2          GCCTTGCCCTTCATTACCGATTCTTCCGTTATTTCCCTGATCGCTAAGCTTGACGCCACT
                **   * ** ** ** ** ** ** ** ** ***** **   * ** ** **   * ** ** **


KO VP2          GTGTTTGCTCAAATAGTCAAGTTGAGAAAGGTAGATACTTTGAAGCCAATCTTGTACAAG
PO VP2          GTTTTCGCTCAAATCGTTAAGTTGAGAAAGGTTGACACCTTGAAGCCAATCTTGTACAAG
Wild-type VP2   GTGTTCGCTCAAATAGTTAAATTACGAAAAGTTGATACTTTAAAACCAATATTATACAAG
AO VP2          GTGTTTGCTCAGATTGTCAAGCTCCGAAAGGTTGACACCCTCAAGCCTATTTTGTACAAG
                ** ** ***** ** ** **   * **** ** ** **   * ** ** ** ** ******


KO VP2          ATAAATAGTGATTCCAACGATTTCTACTTGGTCGCTAATTATGATTGGGTACCAACTTCT
PO VP2          ATCAACTCTGACTCTAACGACTTCTACTTGGTTGCTAACTACGACTGGGTTCCAACCTCT
Wild-type VP2   ATAAAATTCAGACTCAAATGACTTTTATTTAGTAGCTAATTACGATTGGGTGCCAACTTCG
AO VP2          ATCAACTCGGATTCTAATGACTTCTATCTGGTTGCTAACTACGATTGGGTGCCTACTTCT
                ** **      ** ** ** ** ** **   * ** ***** ** ** ***** ** ** **


KO VP2          ACTACAAAAGTATACAAGCAAGTTCCTCAACAATTCGATTTCAGAAACTCTATGCATATG
PO VP2          ACCACCAAGGTTTACAAGCAAGTTCCACAACAATTCGACTTCAGAAACTCTATGCACATG
Wild-type VP2   ACTACAAAAGTATACAAACAGGTTCCGCAACAATTTGATTTTAGAAATTCAATGCATATG
AO VP2          ACTACCAAGGTCTATAAGCAGGTTCCCCAGCAGTTCGACTTTCGAAACTCCATGCATATG
                ** ** ** ** ** ** ** ***** ** ** ** **   **** ** ***** ***


KO VP2          TTGACATCAAACTTAACCTTCACTGTCTATTCTGATTTGTTAGCATTTGTTTCAGCTGAT
PO VP2          TTGACCTCTAACTTGACCTTCACCGTTTACTCTGACTTGTTGGCTTTTCGTTTCTGCTGAC
Wild-type VP2   TTAACTTCGAATCTTACTTTTACGGTTTATTCAGATCTTCTCGCGTTCGTATCAGCTGAC
AO VP2          CTGACTTCGAATCTTACTTTCACCGTGTACTCGGATCTCTTGGCCTTTGTGTCTGCTGAC
                 * ** ** **   * ** ** ** ** ** **   * ** **   * ** ** *****


KO VP2          ACCGTGGAACCAATTAATGCTGTAGCATTCGATAACATGAGAATTATGAACGAATTGTAA
PO VP2          ACCGTTGAACCAATCAACGCTGTTGCTTTCGACAACATGAGAATCATGAACGAATTGTAA
Wild-type VP2   ACAGTAGAACCTATAAATGCAGTTGCATTTGATAATATGCGCATCATGAACGAATTGTAG
AO VP2          ACCGTCGAGCCTATTAACGCTGTCGCCTTTGACAACATGCGAATCATGAATGAGCTGTAA
```

125

```
 **  **  **  **  **  **  **  **  **  **  **  **  ***  *  **  *****  **    ****
```

Figure 10: CLUSTAL 2.1 multiple sequence alignment of optimized and wild-type rotavirus VP2 ORFs nucleotide sequences, KO = *K. lactis* **optimized, PO =** *P. pastoris/H. polymorpha* **optimized and AO =** *A. adeninivorans* **optimized**

```
 **  **  **  **  **  **  **  **  **  **  **  **  ***  *  **  *****  **    ****
```

```
KO VP6          ATGGATGTATTGTATTCATTGAGTAAAACTTTGAAGGATGCTAGAGATAAAATCGTGGAA
PO VP6          ATGGACGTTTTGTACTCTTTGTCTAAGACCTTGAAGGACGCTAGAGACAAGATCGTTGAA
Wild-type VP6   ATGGATGTCCTGTACTCCTTATCAAAAACTCTTAAAGATGCTAGAGACAAAATTGTCGAA
AO VP6          ATGGATGTTTTGTATTCGCTTTCTAAGACTCTGAAGGATGCCCGAGATAAGATTGTTGAG
                ***** **  **** **   *     ** **   * ** ** **   **** ** ** ** **


KO VP6          GGAACTTTGTATAGTAACGTGTCTGATTTGATACAACAATTCAACCAAATGATAATCACC
PO VP6          GGCACCTTGTACTCTAACGTTTCTGACTTGATCCAACAATTCAACCAAATGATCATCACC
Wild-type VP6   GGCACATTATACTCTAATGTGAGTGATCTAATTCAACAATTTAACCAAATGATAATTACT
AO VP6          GGAACTTTGTATTCTAATGTTTCTGATCTGATTCAGCAGTTCAACCAGATGATTATCACT
                ** ** ** **    *** **    ***   * ** ** ** ** ***** ***** ** **


KO VP6          ATGAATGGAAACGAATTTCAAACAGGTGGAATCGGTAATTTGCCAATCAGAAACTGGAAC
PO VP6          ATGAACGGTAACGAGTTCCAAACCGGTGGTATCGGTAACTTGCCAATCAGAAACTGGAAT
Wild-type VP6   ATGAATGGAAATGAGTTCCAAACTGGAGGAATTGGTAATCTACCAATTAGAAATTGGAAT
AO VP6          ATGAACGGTAATGAGTTTCAGACCGGAGGTATTGGAAACCTCCCTATCCGAAACTGGAAT
                ***** ** ** ** ** ** ** ** ** ** ** **   * ** **   **** *****


KO VP6          TTCGATTTCGGTTTGTTGGGAACTACATTGTTGAATTTGGATGCTAACTACGTGGAAACA
PO VP6          TTCGACTTCGGTTTGTTGGGTACTACCTTGTTGAACTTGGACGCTAACTACGTTGAAACC
Wild-type VP6   TTTGATTTTGGATTACTTGGAACAACTCTACTAAATTTAGACGCTAACTACGTCGAAACA
AO VP6          TTCGACTTTGGACTGCTTGGTACTACCCTCTTGAACTTGGATGCTAATTACGTGGAGACT
                ** ** ** **   *   * ** ** **   *   * ** ** ** ***** ***** ** **


KO VP6          GCAAGAAACACCATAGATTACTTCGTCGATTTCGTAGATAACGTTTGTATGGATGAAATG
PO VP6          GCTAGAAACACCATCGACTACTTCGTTGACTTCGTTGACAACGTTTGTATGGACGAAATG
Wild-type VP6   GCCCGTAACACAATTGATTATTTTGTAGATTTTGTAGATAACGTATGTATGGATGAAATG
AO VP6          GCCCGAAATACCATTGACTATTTCGTTGACTTTGTGGATAACGTCTGCATGGATGAGATG
                **   * ** ** ** ** ** ** ** ** ** ** ** ***** ** ***** ** ***


KO VP6          GTTAGAGAATCACAAAGAAATGGTATCGCTCCACAAAGTGATTCCTTGAGAAAATTATCT
PO VP6          GTTAGAGAATCTCAAAGAAACGGTATCGCTCCACAATCTGACTCTTTGAGAAAGTTGTCT
Wild-type VP6   GTTAGAGAATCACAAAGAAATGGAATTGCACCACAATCAGACTCACTTAGAAAATTGTCA
AO VP6          GTCCGAGAGTCTCAGCGAAACGGTATTGCTCCTCAGTCGGACTCTCTCCGAAAGTTGTCC
                **   **** ** **   **** ** ** ** ** **      ** **   * **** ** **


KO VP6          GGTATTAAATTCAAGAGAATCAATTTCGATAACTCTTCAGAATACATCGAAAACTGGAAC
PO VP6          GGTATCAAGTTCAAGAGAATCAACTTCGACAACTCTTCTGAATACATCGAAAACTGGAAC
Wild-type VP6   GGCATTAAGTTCAAAAGGATAAATTTTGATAATTCATCGGAATATATAGAGAACTGGAAT
AO VP6          GGAATTAAGTTCAAGCGAATCAACTTCGATAATTCTTCCGAGTACATTGAGAACTGGAAT
                ** ** ** *****   * ** ** ** ** ** ** ** ** ** ** ** ********


KO VP6          TTGCAAAACAGAAGACAAAGAACCGGTTTCACTTTTCATAAGCCAAATATCTTCCCTTAC
PO VP6          TTGCAAAACAGAAGACAAAGAACCGGTTTCACCTTCCACAAGCCAAACATCTTCCCCATAC
Wild-type VP6   CTACAAAACAGAAGACAACGAACAGGTTTTACATTTCATAAACCAAATATTTTCCCTTAT
AO VP6          CTGCAGAACCGACGACAGCGAACTGGATTCACCTTTCACAAGCCTAACATCTTCCCCTAT
                 * ** *** ** **** **** ** ** ** ** ** ** ** ** ** ***** **


KO VP6          TCTGCATCTTTCACTTTAAATAGATCTCAACCAGCTCATGATAACTTGATGGGTACAATG
PO VP6          TCTGCTTCTTTCACCTTGAACAGATCCCAACCAGCTCACGACAACTTGATGGGTACTATG
Wild-type VP6   TCAGCGTCATTCACACTGAATAGATCACAACCAGCTCATGATAACTTGATGGGTACAATG
AO VP6          TCCGCCTCGTTTACTCTTAACCGATCTCAGCCTGCTCATGACAATCTGATGGGAACCATG
                ** ** ** ** ** **   * **  **** ** ** ***** ** **  ******* ** ***


KO VP6          TGGTTAAATGCTGGATCTGAAATCCAAGTCGCAGGTTTTGATTATTCATGTGCTATAAAT
PO VP6          TGGTTGAACGCTGGTTCTGAAATCCAAGTTGCTGGTTTCGACTACTCTTGTGCTATCAAC
Wild-type VP6   TGGCTGAACGCAGGATCAGAAATTCAGGTCGCTGGATTCGACTATTCGTGTGCAATTAAT
AO VP6          TGGCTTAACGCTGGTTCCGAGATTCAGGTTGCCGGATTTGATTACTCGTGTGCTATCAAC
                *** * ** ** ** ** ** ** ** ** ** ** ** ** ** ***** ** **


KO VP6          GCTCCTGCAAACACTCAACAATTCGAACATATCGTACAATTGAGAAGAGTTTTGACCACT
PO VP6          GCTCCAGCTAACACCCAACAATTCGAACACATCGTTCAATTGAGAAGAGTTTTGACCACC
Wild-type VP6   GCGCCAGCTAATACACAACAATTTGAACATATTGTACAGCTCCGAAGAGTTTTAACTACA
AO VP6          GCTCCCGCCAATACTCAGCAGTTCGAGCACATTGTGCAGCTCCGACGAGTCTTGACTACC
                ** ** ** ** ** ** ** ** ** **   *   ** **** ** ** ** ** **
```

```
KO VP6          GCAACTATCACATTGTTACCAGATGCTGAAAGATTCTCTTTTCCTAGAGTGATAAACTCA
PO VP6          GCTACCATCACCTTGTTGCCAGACGCTGAAAGATTCTCTTTCCCAAGAGTTATCAACTCT
Wild-type VP6   GCTACAATAACACTTTTACCGGATGCAGAAAGATTCAGTTTTCCAAGAGTGATTAATTCA
AO VP6          GCCACTATCACCCTGCTTCCTGACGCTGAGCGATTCTCTTTTCCCCGAGTTATTAATTCC
                ** ** ** **    *  * ** ** ** **  *****  *** **  **** ** ** **

KO VP6          GCTGATGGTGCAACAACCTGGTATTTCAATCCAGTCATTTTTGAGACCTAACAACGTTGAA
PO VP6          GCTGACGGTGCTACCACCTGGTACTTCAACCCAGTTATCTTGAGACCAAACAACGTTGAA
Wild-type VP6   GCTGACGGAGCAACTACATGGTATTTTAATCCAGTAATTCTTAGACCAAATAACGTTGAA
AO VP6          GCTGATGGTGCCACTACCTGGTACTTCAACCCTGTGATCCTCCGACCCAACAATGTGGAG
                *****  ** ** ** ** ***** ** ** ** ** **  *   **** ** ** ** **

KO VP6          GTGGAATTCTTGTTGAACGGACAAATAATTAACACTTACCAAGCAAGATTCGGTACAATC
PO VP6          GTTGAGTTCTTGTTGAACGGTCAAATCATCAACACCTACCAAGCTAGATTCGGCACCATC
Wild-type VP6   GTGGAGTTTCTACTAAACGGGCAGATAATAAACACTTACCAGGCTAGATTTGGAACGATC
AO VP6          GTCGAGTTTCTCTTGAACGGTCAGATTATCAATACTTATCAGGCCCGATTCGGAACCATT
                ** ** **   *  * ***** ** ** ** ** ** ** ** **  **** ** ** **

KO VP6          GTTGCTAGAAACTTCGATACTATCAGATTGAGTTTCCAATTGATGAGACCACCTAACATG
PO VP6          GTTGCTAGAAACTTCGACACCATCAGATTGTCTTTCCAATTGATGAGACCACCAAACATG
Wild-type VP6   GTAGCTAGAAATTTTGATACAATCAGATTGTCGTTTCAGTTGATGAGACCACCAAATATG
AO VP6          GTGGCTCGAAACTTCGACACTATCCGACTGTCTTTTCAGCTTATGCGACCTCCCAATATG
                ** *** **** ** ** ** ***  ** ** **  ** **  * *** **** ** ** ***

KO VP6          ACACCATCAGTAGCTGCATTGTTCCCAAATGCACAACCTTTTGAACATCATGCTACTGTT
PO VP6          ACCCCATCTGTTGCTGCTTTGTTCCCAAACGCTCAACCATTCGAACACCACGCTACCGTT
Wild-type VP6   ACACCATCGGTAGCAGCATTATTTCCAAATGCGCAACCATTTGAACATCATGCTACAGTA
AO VP6          ACCCCTTCCGTCGCTGCCCTTTTCCCTAACGCCCAGCCCTTTGAGCACCATGCTACTGTT
                ** ** ** ** ** **   * ** ** ** ** ** ** ** ** ** ** ***** **

KO VP6          GGTTTGACATTGAAGATCGAAAGTGCAGTATGTGAATCCGTTTTGGCTGATGCAAGTGAA
PO VP6          GGTTTGACCTTGAAGATCGAATCTGCTGTTTGTGAATCTGTTTTGGCTGACGCTTCTGAA
Wild-type VP6   GGACTTACATTGAAAATTGAATCTGCAGTTTGTGAATCTGTACTTGCTGACGCAAGCGAG
AO VP6          GGACTGACCCTTAAGATTGAGTCCGCCGTTTGCGAGTCGGTGCTCGCTGATGCCTCGGAG
                **   *  *   *  ** ** **    ** ** ** ** ** ** **  * ***** **       **

KO VP6          ACAATGTTAGCAAACGTGACCTCCGTCAGACAAGAATATGCTATTCCAGTAGGTCCTGTT
PO VP6          ACCATGTTGGCTAACGTTACCTCTGTTAGACAAGAATACGCTATCCCAGTTGGTCCAGTT
Wild-type VP6   ACAATGCTAGCAAATGTGACATCTGTTAGACAAGAATACGCGATACCAGTTGGACCAGTC
AO VP6          ACTATGTTGGCCAACGTTACCTCTGTGCGACAGGAGTACGCTATTCCTGTCGGACCCGTT
                ** *** * ** ** ** ** ** **   **** ** ** ** ** ** ** ** ** **

KO VP6          TTTCCACCTGGAATGAATTGGACCGATTTGATCACTAACTACAGTCCATCCAGAGAAGAT
PO VP6          TTCCCACCAGGTATGAACTGGACCGACTTGATCACCAACTACTCTCCATCCAGAGAAGAC
Wild-type VP6   TTTCCACCAGGTATGAATTGGACTGATTTGATCACTAACTATTCACCATCTAGAGAGGAT
AO VP6          TTTCCTCCCGGTATGAACTGGACTGACCTGATCACCAATTATTCTCCTTCCCGAGAGGAT
                ** ** ** ** ***** ***** **   ******* ** **    ** ** **** **

KO VP6          AATTTGCAAAGAGTGTTCACTGTCGCTTCTATCAGATCAATGTTGGTTAAATAA
PO VP6          AACTTGCAAAGAGTTTTCACCGTTGCTTCTATCAGATCCATGTTGGTTAAGTAA
Wild-type VP6   AACTTGCAGCGTGTATTTACAGTGGCTTCCATTAGAAGCATGCTTGTCAAATAA
AO VP6          AACCTTCAGCGAGTGTTCACTGTCGCTTCGATTCGATCTATGCTGGTCAAGTAG
                **   * **  * ** ** ** ** ** ** ***** **   *** * ** ** **
```

Figure 11: CLUSTAL 2.1 multiple sequence alignment of optimized and wild-type rotavirus VP6 ORFs nucleotide sequences, KO = *K. lactis* optimized, PO = *P. pastoris/H. polymorpha* optimized and AO = *A. adeninivorans* optimized

APPENDIX C ClustalX Multiple sequence alignment of optimized and wild-type VP2 and VP6 rotavirus ORFs amino acid sequences

```
Wild-type VP2    MAYRKRGARREANLNNNDRMQEKIDEKQDSNKIQLSDKVLSKKEEIVTDSHEEVKVTDEL
AO VP2           MAYRKRGARREANLNNNDRMQEKIDEKQDSNKIQLSDKVLSKKEEIVTDSHEEVKVTDEL
PO VP2           MAYRKRGARREANLNNNDRMQEKIDEKQDSNKIQLSDKVLSKKEEIVTDSHEEVKVTDEL
KO VP2           MAYRKRGARREANLNNNDRMQEKIDEKQDSNKIQLSDKVLSKKEEIVTDSHEEVKVTDEL
                 ***********************************************************

Wild-type VP2    KKSTKEESKQLLEVLKTKEEHQKEIQYEILQKTIPTFEPKETILRKLEDIQPELAKKQTK
AO VP2           KKSTKEESKQLLEVLKTKEEHQKEIQYEILQKTIPTFEPKETILRKLEDIQPELAKKQTK
PO VP2           KKSTKEESKQLLEVLKTKEEHQKEIQYEILQKTIPTFEPKETILRKLEDIQPELAKKQTK
KO VP2           KKSTKEESKQLLEVLKTKEEHQKEIQYEILQKTIPTFEPKETILRKLEDIQPELAKKQTK
                 ***********************************************************

Wild-type VP2    LFRIFEPKQLPIYRANGERELRNRWYWKLKKDTLPDGDYDVREYFLNLYDQVLTEMPDYL
AO VP2           LFRIFEPKQLPIYRANGERELRNRWYWKLKKDTLPDGDYDVREYFLNLYDQVLTEMPDYL
PO VP2           LFRIFEPKQLPIYRANGERELRNRWYWKLKKDTLPDGDYDVREYFLNLYDQVLTEMPDYL
KO VP2           LFRIFEPKQLPIYRANGERELRNRWYWKLKKDTLPDGDYDVREYFLNLYDQVLTEMPDYL
                 ***********************************************************

Wild-type VP2    LLKDMAVENKNSRDAGKVVDSETASICDAIFQDEETEGAVRRFIAEMRQRVQADRNVVNY
AO VP2           LLKDMAVENKNSRDAGKVVDSETASICDAIFQDEETEGAVRRFIAEMRQRVQADRNVVNY
PO VP2           LLKDMAVENKNSRDAGKVVDSETASICDAIFQDEETEGAVRRFIAEMRQRVQADRNVVNY
KO VP2           LLKDMAVENKNSRDAGKVVDSETASICDAIFQDEETEGAVRRFIAEMRQRVQADRNVVNY
                 ***********************************************************

Wild-type VP2    PSILHPIDYAFNEYFLQHQLVEPLNNDIIFNYIPERIRNDVNYILNMDRNLPSTARYIRP
AO VP2           PSILHPIDYAFNEYFLQHQLVEPLNNDIIFNYIPERIRNDVNYILNMDRNLPSTARYIRP
PO VP2           PSILHPIDYAFNEYFLQHQLVEPLNNDIIFNYIPERIRNDVNYILNMDRNLPSTARYIRP
KO VP2           PSILHPIDYAFNEYFLQHQLVEPLNNDIIFNYIPERIRNDVNYILNMDRNLPSTARYIRP
                 ***********************************************************

Wild-type VP2    NLLQDRLNLHDNFESLWDTITTSNYILARSVVPDLKELVSTEAQIQKMSQDLQLEALTIQ
AO VP2           NLLQDRLNLHDNFESLWDTITTSNYILARSVVPDLKELVSTEAQIQKMSQDLQLEALTIQ
PO VP2           NLLQDRLNLHDNFESLWDTITTSNYILARSVVPDLKELVSTEAQIQKMSQDLQLEALTIQ
KO VP2           NLLQDRLNLHDNFESLWDTITTSNYILARSVVPDLKELVSTEAQIQKMSQDLQLEALTIQ
                 ***********************************************************

Wild-type VP2    SETQFLTGINSQAANDCFKTLIAAMLSQRTMSLDFVTTNYMSLISGMWLLTVVPNDMFIR
AO VP2           SETQFLTGINSQAANDCFKTLIAAMLSQRTMSLDFVTTNYMSLISGMWLLTVVPNDMFIR
PO VP2           SETQFLTGINSQAANDCFKTLIAAMLSQRTMSLDFVTTNYMSLISGMWLLTVVPNDMFIR
KO VP2           SETQFLTGINSQAANDCFKTLIAAMLSQRTMSLDFVTTNYMSLISGMWLLTVVPNDMFIR
                 ***********************************************************

Wild-type VP2    ESLVACQLAIVNTIIYPAFGMQRMHYRNGDPQTPFQIAEQQIQNFQVANWLHFVNNNQFR
AO VP2           ESLVACQLAIVNTIIYPAFGMQRMHYRNGDPQTPFQIAEQQIQNFQVANWLHFVNNNQFR
PO VP2           ESLVACQLAIVNTIIYPAFGMQRMHYRNGDPQTPFQIAEQQIQNFQVANWLHFVNNNQFR
KO VP2           ESLVACQLAIVNTIIYPAFGMQRMHYRNGDPQTPFQIAEQQIQNFQVANWLHFVNNNQFR
                 ***********************************************************

Wild-type VP2    QAVIDGVLNQVLNDNIRNGHVINQLMEALMQLSRQQFPTMPIDYKRSIQRGILLLSNRLG
AO VP2           QAVIDGVLNQVLNDNIRNGHVINQLMEALMQLSRQQFPTMPIDYKRSIQRGILLLSNRLG
PO VP2           QAVIDGVLNQVLNDNIRNGHVINQLMEALMQLSRQQFPTMPIDYKRSIQRGILLLSNRLG
KO VP2           QAVIDGVLNQVLNDNIRNGHVINQLMEALMQLSRQQFPTMPIDYKRSIQRGILLLSNRLG
                 ***********************************************************

Wild-type VP2    QLVDLTRLLAYNYETLMACITMNMQHVQTLTTEKLQLTSVTSLCMLIGNATVIPSPQTLF
AO VP2           QLVDLTRLLAYNYETLMACITMNMQHVQTLTTEKLQLTSVTSLCMLIGNATVIPSPQTLF
PO VP2           QLVDLTRLLAYNYETLMACITMNMQHVQTLTTEKLQLTSVTSLCMLIGNATVIPSPQTLF
KO VP2           QLVDLTRLLAYNYETLMACITMNMQHVQTLTTEKLQLTSVTSLCMLIGNATVIPSPQTLF
                 ***********************************************************
```

```
Wild-type VP2    HYYNVNVNFHSNYNERINDAVAIITAANRLNLYQKKMKAIVEDFLKRLYIFDVSRVPDDQ
AO VP2           HYYNVNVNFHSNYNERINDAVAIITAANRLNLYQKKMKAIVEDFLKRLYIFDVSRVPDDQ
PO VP2           HYYNVNVNFHSNYNERINDAVAIITAANRLNLYQKKMKAIVEDFLKRLYIFDVSRVPDDQ
KO VP2           HYYNVNVNFHSNYNERINDAVAIITAANRLNLYQKKMKAIVEDFLKRLYIFDVSRVPDDQ
                 ************************************************************

Wild-type VP2    MYRLRDRLRLLPVEIRRLDIFNLILMNMDQIERASDKIAQGVIIAYRDMHLERDEMYGYV
AO VP2           MYRLRDRLRLLPVEIRRLDIFNLILMNMDQIERASDKIAQGVIIAYRDMHLERDEMYGYV
PO VP2           MYRLRDRLRLLPVEIRRLDIFNLILMNMDQIERASDKIAQGVIIAYRDMHLERDEMYGYV
KO VP2           MYRLRDRLRLLPVEIRRLDIFNLILMNMDQIERASDKIAQGVIIAYRDMHLERDEMYGYV
                 ************************************************************

Wild-type VP2    NIARNLEGFQQINLEELMRSGDYAQITNMLLNNQPVALVGALPFITDSSVISLIAKLDAT
AO VP2           NIARNLEGFQQINLEELMRSGDYAQITNMLLNNQPVALVGALPFITDSSVISLIAKLDAT
PO VP2           NIARNLEGFQQINLEELMRSGDYAQITNMLLNNQPVALVGALPFITDSSVISLIAKLDAT
KO VP2           NIARNLEGFQQINLEELMRSGDYAQITNMLLNNQPVALVGALPFITDSSVISLIAKLDAT
                 ************************************************************

Wild-type VP2    VFAQIVKLRKVDTLKPILYKINSDSNDFYLVANYDWVPTSTTKVYKQVPQQFDFRNSMHM
AO VP2           VFAQIVKLRKVDTLKPILYKINSDSNDFYLVANYDWVPTSTTKVYKQVPQQFDFRNSMHM
PO VP2           VFAQIVKLRKVDTLKPILYKINSDSNDFYLVANYDWVPTSTTKVYKQVPQQFDFRNSMHM
KO VP2           VFAQIVKLRKVDTLKPILYKINSDSNDFYLVANYDWVPTSTTKVYKQVPQQFDFRNSMHM
                 ************************************************************

Wild-type VP2    LTSNLTFTVYSDLLAFVSADTVEPINAVAFDNMRIMNEL
AO VP2           LTSNLTFTVYSDLLAFVSADTVEPINAVAFDNMRIMNEL
PO VP2           LTSNLTFTVYSDLLAFVSADTVEPINAVAFDNMRIMNEL
KO VP2           LTSNLTFTVYSDLLAFVSADTVEPINAVAFDNMRIMNEL
                 ***************************************
```

Figure 12: CLUSTAL 2.1 multiple sequence alignment of optimized and wild-type rotavirus VP6 ORFs amino acid sequences, KO = *K. lactis* optimized, PO = *P. pastoris*/*H. polymorpha* optimized and AO = *A. adeninivorans* optimized

```
Wild-type VP6    MDVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIITMNGNEFQTGGIGNLPIRNWN
AO VP6           MDVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIITMNGNEFQTGGIGNLPIRNWN
PO VP6           MDVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIITMNGNEFQTGGIGNLPIRNWN
KO VP6           MDVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIITMNGNEFQTGGIGNLPIRNWN
                 ************************************************************

Wild-type VP6    FDFGLLGTTLLNLDANYVETARNTIDYFVDFVDNVCMDEMVRESQRNGIAPQSDSLRKLS
AO VP6           FDFGLLGTTLLNLDANYVETARNTIDYFVDFVDNVCMDEMVRESQRNGIAPQSDSLRKLS
PO VP6           FDFGLLGTTLLNLDANYVETARNTIDYFVDFVDNVCMDEMVRESQRNGIAPQSDSLRKLS
KO VP6           FDFGLLGTTLLNLDANYVETARNTIDYFVDFVDNVCMDEMVRESQRNGIAPQSDSLRKLS
                 ************************************************************

Wild-type VP6    GIKFKRINFDNSSEYIENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPAHDNLMGTM
AO VP6           GIKFKRINFDNSSEYIENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPAHDNLMGTM
PO VP6           GIKFKRINFDNSSEYIENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPAHDNLMGTM
KO VP6           GIKFKRINFDNSSEYIENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPAHDNLMGTM
                 ************************************************************

Wild-type VP6    WLNAGSEIQVAGFDYSCAINAPANTQQFEHIVQLRRVLTTATITLLPDAERFSFPRVINS
AO VP6           WLNAGSEIQVAGFDYSCAINAPANTQQFEHIVQLRRVLTTATITLLPDAERFSFPRVINS
PO VP6           WLNAGSEIQVAGFDYSCAINAPANTQQFEHIVQLRRVLTTATITLLPDAERFSFPRVINS
KO VP6           WLNAGSEIQVAGFDYSCAINAPANTQQFEHIVQLRRVLTTATITLLPDAERFSFPRVINS
                 ************************************************************

Wild-type VP6    ADGATTWYFNPVILRPNNVEVEFLLNGQIINTYQARFGTIVARNFDTIRLSFQLMRPPNM
AO VP6           ADGATTWYFNPVILRPNNVEVEFLLNGQIINTYQARFGTIVARNFDTIRLSFQLMRPPNM
PO VP6           ADGATTWYFNPVILRPNNVEVEFLLNGQIINTYQARFGTIVARNFDTIRLSFQLMRPPNM
KO VP6           ADGATTWYFNPVILRPNNVEVEFLLNGQIINTYQARFGTIVARNFDTIRLSFQLMRPPNM
                 ************************************************************

Wild-type VP6    TPSVAALFPNAQPFEHHATVGLTLKIESAVCESVLADASETMLANVTSVRQEYAIPVGPV
AO VP6           TPSVAALFPNAQPFEHHATVGLTLKIESAVCESVLADASETMLANVTSVRQEYAIPVGPV
PO VP6           TPSVAALFPNAQPFEHHATVGLTLKIESAVCESVLADASETMLANVTSVRQEYAIPVGPV
KO VP6           TPSVAALFPNAQPFEHHATVGLTLKIESAVCESVLADASETMLANVTSVRQEYAIPVGPV
                 ************************************************************

Wild-type VP6    FPPGMNWTDLITNYSPSREDNLQRVFTVASIRSMLVK
AO VP6           FPPGMNWTDLITNYSPSREDNLQRVFTVASIRSMLVK
PO VP6           FPPGMNWTDLITNYSPSREDNLQRVFTVASIRSMLVK
KO VP6           FPPGMNWTDLITNYSPSREDNLQRVFTVASIRSMLVK
                 *************************************
```

Figure 13: CLUSTAL 2.1 multiple sequence alignment of optimized and wild-type rotavirus VP6 ORFs amino acid sequences, KO = *K. lactis* **optimized, PO =** *P. pastoris/H. polymorpha* **optimized and AO =** *A. adeninivorans* **optimized**
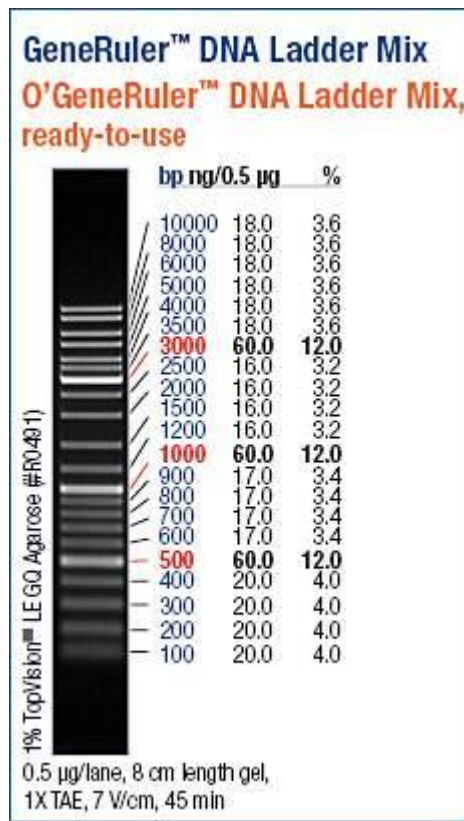
APPENDIX D GeneRuler DNA Ladder Mix ladder

Figure 14: GeneRuler DNA Ladder Mix ladder used in this study for size of DNA fragments of genomic, plasmid, restriction enzyme digests and PCR DNA (obtained from https://www.fishersci.com/shop/products/thermo-scientific-o-generuler-dna-ladder-mix-ready-to-use-100-10-000-bp/p-4073123).